Computer Assisted Classification and Identification of Actinomycetes

Jongsik Chun

(B.Sc. Microbiology, Seoul National University, Seoul, Korea)

Thesis submitted in accordance with the requirements
of the University of Newcastle upon Tyne
for the Degree of Doctor of Philosophy

Department of Microbiology
The Medical School
Newcastle upon Tyne
England-UK

July 1995

ABSTRACT

Three computer software packages were written in the C++ language for the analysis of numerical phenetic, 16S rRNA sequence and pyrolysis mass spectrometric data. The ***X*** program, which provides routines for editing binary data, for calculating test error, for estimating cluster overlap and for selecting diagnostic and selective tests, was evaluated using phenotypic data held on streptomycetes. The ***AL16S*** program has routines for editing 16S rRNA sequences, for determining secondary structure, for finding signature nucleotides and for comparative sequence analysis; it was used to analyse 16S rRNA sequences of mycolic acid-containing actinomycetes. The ***ANN*** program was used to generate backpropagation-artificial neural networks using pyrolysis mass spectra as input data.

Almost complete 16S rDNA sequences of the type strains of all of the validly described species of the genera *Nocardia* and *Tsukamurella* were determined following isolation and cloning of the amplified genes. The resultant nucleotide sequences were aligned with those of representatives of the genera *Corynebacterium*, *Gordona*, *Mycobacterium*, *Rhodococcus* and *Turicella* and phylogenetic trees inferred by using the neighbor-joining, least squares, maximum likelihood and maximum parsimony methods. The mycolic acid-containing actinomycetes formed a monophyletic line within the evolutionary radiation encompassing actinomycetes.

The "mycolic acid" lineage was divided into two clades which were equated with the families *Corynebacteriaceae* and *Mycobacteriaceae*. The family *Corynebacteriaceae* contained the genera *Corynebacterium*, *Dietzia* and *Turicella* and the family *Mycobacteriaceae* the genera *Gordona*, *Mycobacterium*, *Nocardia*, *Rhodococcus* and *Tsukamurella*. It was clear from the 16S rDNA sequence data that *Nocardia pinensis* was misclassified in the genus *Nocardia* and that *Tsukamurella wratislaviensis* belonged to the genus *Rhodococcus*. The genus *Nocardia* formed a distinct clade that was clearly associated with the genus *Rhodococcus*. Two sublines were recognised within the *Nocardia* clade; one consisted of *Nocardia asteroides* and related taxa and the other of *Nocardia otitidiscaviarum* and allied species. The two sublines are distinguished by nucleotide differences in helix 37-1. The type strains of all of the *Nocardia* species contained

hexahydrogenated menaquinones with eight isoprene units in which the two end units were cyclised.

Actinomycetes selectively isolated from an activated sludge plant showing extensive foaming were the subject of a polyphasic taxonomic study. The sludge isolates, which clearly belong to the genus *Tsukamurella* on the basis of 16S rRNA data, contained highly unsaturated long chain mycolic acids and unsaturated menaquinones with nine isoprene units, properties consistent with their classification in the genus *Tsukamurella*. Six representative isolates and marker strains of *Tsukamurella paurometabola* were the subject of a numerical phenetic taxonomic study. The test strains were assigned to four groups in the simple matching coefficient, unweighted pair group method with arithmetic averages analysis. The sludge isolates formed a homogeneous cluster with the three remaining clusters composed of *Tsukamurella paurometabola* strains. Excellent congruence was found between these numerical taxonomic data and results derived from corresponding studies based on Curie point pyrolysis mass spectrometric and whole-organism protein electrophoretic analyses. The combined data suggest that the sludge isolates form the nucleus of a new species of the genus *Tsukamurella* and that *Tsukamurella paurometabola* is a heterogeneous taxon.

Representatives of three putatively novel streptomycete species isolated from soil were used to develop and evaluate an identification system based on Curie point pyrolysis mass spectromety and artificial neural network analysis. The test strains consisted of sixteen target *Streptomyces* strains and one hundred and thirty-eight actinomycetes belonging to the genera *Actinomadura*, *Mycobacterium*, *Nocardia*, *Nocardiopsis*, *Saccharomonospora* and *Streptosporangium*. It was clear from the identification results that artificial neural network analysis was superior to conventional procedure based on principal component and canonical variate statistics. The problem of misidentification of some of the non-target strains was solved by the development of a neural network topology which contained an output neuron designed to detect non-target pyrolysis mass spectrometric patterns. The pyrolysis mass spectrometry-artificial neural network system was evaluated using thirteen fresh streptomycete isolates and found to be capable of long-term identification of the target strains.

PREFACE

To my parents who brought me to this wonderful world

and to Mike who led me to this extraordinary science

"If a man will begin with certainties he shall end in doubts; but if he will be constant to begin with doubts he shall end in certainties." - Francis Lord Bacon

"What does not destroy us, makes us stronger." - Nietzsche

<BLANK>

ACKNOWLEDGMENTS

I would express my sincere gratitude to my supervisor Professor Michael Goodfellow for his guidance, all-round discussions and, most of all, his friendship throughout my stay in Newcastle. Life wasn't always about fashionable things, Mike!

I am also grateful to Dr. Alan Ward for sharing his enthusiasm for programming and artificial neural networks, and to Professor David Minnikin for his friendly and easy-to-understand chemistry lectures.

I am indebted to Professor Yung Chil Hah and academic staff in Seoul National University for their support and encouragement throughout this project.

I also thank my colleagues who shared "the lab. life". To Gilson, thank you for sharing those up-and-down years and for all dishes we had together, or rather precisely cooked by you. I owe so much to Baker, Bongcheol, Chinam, Daejin, Elayne, Hassan, Jeongjin, Joon, Kamil, Martha, Mohammed, Nevzat, Russell, Seonggyun and finally Sam, who shared not only clean benches but also great fun.

Many thanks to members of the Korean Society in Newcastle who constantly provided me with the moral support that kept me going.

I am also indebted to the academic and technical staff in the Medical School and Department of Microbiology of Newcastle University, especially to Bob Riddell and Clive Hetherington for their help in DNA sequencing and pyrolysis mass spectrometry, respectively.

To my parents, thank you for your unselfish, never-ending support and encouragement. Without you, none of me was possible.

I gratefully acknowledge the financial support by the British Council (Seoul).

<BLANK>

PUBLICATIONS RELATED TO THIS THESIS

- **Chun, J., Atalan, E., Ward, A. C. & Goodfellow, M. (1993).** Artificial neural network analysis of pyrolysis mass spectrometric data in the identification of *Streptomyces* strains. *FEMS Microbiology Letters* **107,** 321-326.
- **Chun, J., Atalan, E., Kim, S. B., Kim, H. J., Hamid, M. E., Trujillo, M. E., Magee, J. G., Manfio, G. P., Ward, A. C. & Goodfellow, M. (1993).** Rapid identification of streptomycetes by artificial neural network analysis of pyrolysis mass spectra. *FEMS Microbiology Letters* **114,** 115-120.
- **Hamid, M. E., Chun, J., Magee, J. G., Minnikin, D. E. & Goodfellow, M. (1994).** Rapid characterisation and identification of *Mycobacterium* using fluorogenic enzyme tests. *Zentralblatt für Bakteriologie* **280,** 476-487.
- **Goodfellow, M., Chun, J., Stubbs, S. & Toboli, A. S. (1994).** Transfer of *Nocardia amarae* Lechevalier and Lechevalier to the genus *Gordona* as *Gordona amarae* comb. nov. *Letters in Applied Microbiology* **19,** 401-405.
- **Goodfellow, M., Chun, J., Atalan, E. & Sanglier, J. J. (1994).** Curie point pyrolysis mass spectrometry and its application to bacterial systematics. In *Bacterial Systematics and Diversity* pp. 87-104. F. G. Priest, A. Ramos-Cormenzana & B. Tindall, eds. Prenum Press: London.
- **Chun, J. & Goodfellow, M. (1995).** A phylogenetic analysis of the genus *Nocardia* with 16S rRNA gene sequences. *International Journal of Systematic Bacteriology* **45,** 240-245.

<BLANK>

CONTENTS

Abstract i

Preface iii

Acknowledgements v

Publications related to the thesis vii

Table of contents ix

List of tables xv

List of figures xvii

Chapter I. Modern bacterial systematics: theoretical background and development of software tools 1

A. Introduction 1

1. Phylogenetic reconstruction from molecular sequence data 1
   1.1. Choice of molecule 2
   1.2. Alignment 4
   1.3. Construction of phylogenetic trees 6
   1.4. Rooting the phylogenetic tree 20
   1.5. Evaluation of evolutionary trees 22
   1.6. Simulation studies on different phylogenetic inference methods 23
   1.7. Misconceptions in bacterial phylogenies: a review of the recent literature 25
2. Computer-assisted classification and identification based on phenotypic/phenetic properties 29
   2.1. Computer-assisted classification 29
   2.2. Computer-assisted identification 42
3. Software tools 45

B. Materials and Methods 49

1. Computers and compilers 49
2. The AL16S program 49
   2.1. Description 49
   2.2. Main functions 50
3. The X program 54
   3.1. Description 54
   3.2. Main Functions 54

C. Results and Discussion 67

1. The AL16S program 67
2. The X program 67

Chapter II. Phylogenetic analysis of the family *Nocardiaceae* and related actinomycetes using 16S ribosomal RNA gene sequences 89

A. Introduction 89
1. Phylogeny inferred from 16S ribosomal ribonucleic acid 89
1.1. 16S rRNA as molecular chronometer 89
1.2. Molecular biological properties of ribosomal RNA gene clusters 89
1.3. Methods used to determine 16S rRNA sequences 93
1.4. Species concept and place of 16S rRNA sequencing in bacterial systematic .. 104
2. Chemosystematics 109
2.1. Introduction 109
2.2. Specific chemical markers 110
3. The order *Actinomycetales* 122
3.1. The order *Actinomycetales* Buchanan 1917[AL] 122
3.2. Suprageneric classification of actinomycetes and related taxa 127
4. Family Nocardiaceae and related taxa 139
4.1. Suprageneric classification of mycolic acid-containing taxa 139
4.2. The family *Nocardiaceae* Castellani and Chalmers 1919[AL] 141

B. Materials and Methods 157
1. Test strains 157
2. Chemotaxonomy 157
2.1. Preparation of biomass 157
2.2. Analysis of menaquinones 157
3. Sequencing of 16S rRNA gene 160
3.1. Preparation of biomass 160
3.2. Reagents and basic procedures 160
3.3. Isolation of chromosomal DNA 163
3.4. Amplification of 16S rRNA genes 164
3.5. Isolation of 16S rDNA 166
3.6. *Cloning of 16S rDNA* 166
3.7. Isolation of plasmids 167
3.8. Sequencing of 16S rDNA 168
3.9. Alignment of 16S rDNA sequences 168
3.10. Data analysis 168

C. Results 171
1. Chemotaxonomy 171
2. 16S rRNA gene sequences 171
2.1. Sequencing errors 171
2.2. Phylogeny of the family *Nocardiaceae* and related actinomycetes 174
2.3. Phylogeny of the genus *Nocardia* 192

D. Discussion 201
1. Suprageneric classification of mycolic acid-containing actinomycetes and related strains 201
2. Phylogeny of *Nocardia* 203
3. Taxonomic status of *Nocardia pinensis* 204
4. Taxonomic status of *Tsukamurella wratislaviensis* 205

Chapter III. Isolation and characterisation of actinomycetes associated with activated sludge foaming 207

A. Introduction 207
1. Activated sludge treatment 207
2. Bulking and Foaming 211
3. Actinomycetes associated with activated sludge foaming 212

B. Materials and Methods 215
1. Test strains 215
1.1. Sample collection 215
1.2. Isolation and maintenance of strains 215
1.3. Test strains 215
2. Chemotaxonomy 218
2.1. Preparation of biomass 218
2.2. Analysis of fatty acids, including mycolic acids 218
2.3. Analysis of menaquinones 221
2.4. Analysis for diaminopimelic acid isomers 221
2.5. Analysis of whole-organism sugars 222
3. Pyrolysis mass spectrometry 223
3.1. Test strains 223
3.2. Growth conditions 223
3.3. Preparation of samples 223
3.4. Curie-point pyrolysis mass spectrometry 223
3.5. Multivariate statistical analysis 224
4. Whole-organism protein electrophoresis 224
4.1. Cultivation of test strains and sample preparation 224
4.2. Denaturing polyacrylamide gel electrophoresis 224
4.3. Data analysis 227
5. Numerical phenetic classification 228
5.1. Data acquisition 228
5.2. Morphology and staining 229
5.3. Degradation tests 229
5.4. Nutritional tests 229
5.5. Resistance to antibiotics 230
5.6. Growth in the presence of chemical inhibitors 230
5.7. Growth at 45 ºC 231
5.8. Data analysis 231
6. Sequencing of 16S rDNA 231
6.1. Test strains and procedure 231
6.2. Data analysis 232

C. Results 233
1. Selective isolation 233
2. Morphology and staining 233
3. Chemotaxonomy 233
4. Numerical phenetic classification 241
4.1. Final data matrix and calculation of test error 241
4.2. Cophenetic correlation coefficient 242
4.3. Numerical phenetic classification based on the S_{SM} coefficient and the unweighted pair group method with arithmetic averages algorithm 244
4.4. Other numerical phenetic classifications 244
4.5. Principal component analysis 248
5. Molecular systematics 248
6. Chemical fingerprints 252
6.1. Curie point pyrolysis mass spectrometry 252
6.2. Whole-organism protein electrophoresis 253

D. Discussion 259

CHAPTER IV. **Identification of Streptomycetes Using Curie Point Pyrolysis Mass Spectrometry and Artificial Neural Networks** 267

A. Introduction **267**
1. Pyrolysis mass spectrometry 267
1.1. Introduction 267
1.2. Pyrolysis mass spectrometry 270
1.3. Reproducibility of pyrolysis mass spectrometric data 275
2. Artificial neural networks 279
2.1. Introduction 279
2.2. History of artificial neural networks 280
2.3. Structure of artificial neural networks 281
2.4. Application of artificial neural networks to biology and chemistry 286
2.5. Artificial neural networks and microbial systematic 288

B. Materials and Methods **293**
1. Test strains 293
2. Datasets 293
2.1. *Streptomyces* (target) strains 293
2.2. Non-*Streptomyces* (non-target) strains 293
3. Curie-point pyrolysis mass spectrometry 295
3.1. Growth conditions and sample preparation 295
3.2. Pyrolysis mass spectrometry 295
4. Statistical analyses 296
5. Artificial neural network analysis 299
5.1. Computer programs 299
5.2. Pre-processing 299
5.3. Training of artificial neural networks 300
5.4. Identification of mass spectra 300
5.5. Generation of artificial neural networks 302

C. Results and Discussion **307**
1. Pyrolysis mass spectra 307
2. Effect of scaling on learning 307
3. Comparison of datasets 1A and 1B in multivariate statistical and artificial neural network analyses 307
4. Comparison of simple and complex artificial neural networks 313
5. Statistical comparison of pyrolysis mass spectral data from datasets 1A, 2 and 3 316
6. Artificial neural network comparison of pyrolysis mass spectral data of datasets 1A, 2 and 3 318
7. Spectral identification of data in datasets 4, 5 and 6 using 'operational fingerprinting' and artificial neural network analysis 322

References **335**

Appendix A. Culture media and buffers **381**

Appendix B. IUB nucleotide code **385**

Appendix C. 16S rRNA numbering system based on *Escherichia coli* **386**

Appendix D. Helix nomenclature based on *Escherichia coli* **387**

Appendix E. 16S rRNA sequence similarity and corresponding DNA relatedness values for mycolic acid-containing actinomycetes **388**

Appendix F. A phylogenetic tree depicting relationships between representatives of actinomycetes and related taxa 392

Appendix G. Accession numbers of 16S rRNA sequences for mycolic acid-containing actinomycetes and related taxa 398

Appendix H. Aligned 16S rDNA sequences of nocardiae and the related actinomycetes determined in the present study 401

Appendix I. Aligned 16S rDNA sequences of sludge isolate N1171 and *Tsukamurella paraumetabola* strains JC7^T, N663 and M334 409

Appendix J. Test strains held in the non-target database 412

<BLANK>

LIST OF TABLES

Chapter I

Table 1-1. Performance of numerical analyses implemented in the ***X*** program 68

Table 1-2. Clusters used to evaluate the ***X*** program 69

Table 1-3. Unit characters used to evaluate the ***x*** program 70

Table 1-4. Overlap values estimated from the identification matrix 87

Table 1-5. Identification of ten randomly chosen strains 88

Chapter II

Table 2-1. Comparison of 16S rRNA sequences obtained by using the reverse transcriptase and DNA sequencing methods 100

Table 2-2. Sequencing errors found in entries to the GenBank/EMBL database 103

Table 2-3. Fatty acid patterns of actinomycetes 112

Table 2-4. Classification of peptidoglycans 117

Table 2-5. Major phospholipids found in actinomycetes 120

Table 2-6. Classification of actinomycete phospholipids 121

Table 2-7. Classification of actinomycetes based on wall chemotypes 123

Table 2-8. Distribution of some chemical markers, cultural and morphological properties of actinomycetes and related organisms classified on the basis of 16S rRNA sequence data 130

Table 2-9. Actinomycete taxa of uncertain phylogenetic affiliation 136

Table 2-10. Phenotypic characteristics of wall chemotype IV actinomycetes with mycolic acids and some related organisms 142

Table 2-11. Chemotaxonomic characters which can be used to differentiate between wall chemotype IV actinomycetes with mycolic acids and between them and amycolate taxa 143

Table 2-12. Validly described species classified in the family *Nocardiaceae* 144

Table 2-13. Test strains 158

Table 2-14. Oligonucleotide primers used in polymerase chain reaction amplification and sequencing of 16S rDNA 165

Table 2-15. Predominant menaquinones and nucleotide sequence accession numbers of the test strains 172

Table 2-16. Comparison of 16S rRNA sequences of the type strains of *Nocardia* species determined in three independent studies 173

Table 2-17. Effects of different outgroup strains on positions of the root in phylogenetic trees based on mycolic acid-containing and related actinomycetes 185

Table 2-18. Mean 16S rRNA sequence similarities found between different phyletic lineages (intergroup similarity) 187

Table 2-19. Group specific nucleotide signatures that differentiate between the phylogenetic lineages detected within representatives of the mycolic acid-containing taxa 190

Table 2-20. Similarity values for the 16S rDNA sequences of the representative *Nocardia* strains 194

Table 2-21. 16S rDNA sequence signatures that differentiate between the representative *Nocardia* strains 199

Chapter III

Table 3-1. Causes and effects of activated sludge separation problems 209

Table 3-2. Designation and source of the reference strains 217

Table 3-3. Composition of SDS-PAGE gels 226

Table 3-4. Composition of cellular fatty acids of sludge isolates and representative strains of *Tsukamurella paurometabola* 234

Table 3-5. Comparison of dendrograms and corresponding similarity matrices using the cophenetic correlation coefficient 243

Table 3-6. Distribution of positive characters to clusters defined at the 80% similarity level in the S_{SM}, UPGMA analysis 246

Table 3-7. Matrix of similarity values (lower part) and the number of differences (upper part) in the 16S rDNA sequences of sludge isolate N1171 and *Tsukamurella paurometabola* strains JC7^T, M334 and N663 249

Chapter IV

Table 4-1. Some applications of artificial neural networks in biology and chemistry 287

Table 4-2. Application of artificial neural networks in microbial systematics 289

Table 4-3. *Streptomyces* strains and datasets 294

Table 4-4. Artificial neural networks 305

Table 4-5. Identification results obtained for test set 1 using NET-1AS based on triplicate pyrolysis mass spectra 312

Table 4-6. Comparison of identification results for networks NET-1AS and NET-1AA 315

Table 4-7. Cross identification results from networks NET-1, NET-2 and NET-3. 321

LIST OF FIGURES

Chapter I

Figure 1-1. Additive and ultrametric trees 8

Figure 1-2. Comparison of phylogenetic trees based on analyses of full nucleotide positions and polymorphic sites 27

Figure 1-3. Major steps involved in numerical classification and identification 31

Figure 1-4. Major steps and software tools used for analysis of 16S rRNA sequence data . 51

Figure 1-5. Major steps and software tools used in the analysis of numerical phenotypic data 55

Figure 1-6. A diagrammatic presentation of the taxon-radius model based on the assumption that the taxon is a hyperspherical normal cluster 61

Figure 1-7. Illustration of the method used to derive *W* statistics 63

Figure 1-8. Computer generated output of the module OVERMAT obtained with 26 *Streptomyces* taxa 77

Figure 1-9. Computer generated output from the CHARSEP analysis 79

Figure 1-10. Computer generated output from the DIACHAR analysis..... 80

Figure 1-11. Computer generated output from the module 'Hypothetical Median Organism' . 83

Chapter II

Figure 2-1. Comparison of 16S rRNA sequence similarity and DNA relatedness values. Based on data taken from studies on mycolic acid-containing actinomycetes . 107

Figure 2-2. Structures of (a) unsaturated menaquinones characteristic of *Tsukamurella* strains, (b) partially saturated menaquinones characteristic of *Streptomycs* strains and (c) cyclic menaquinones characteristic of *Nocardia* strains 114

Figure 2-3. Fragments of the primary structure of a typical peptidoglycan 116

Figure 2-4. An abridged phylogenetic tree showing the actinomycete line of descent based on 1371 16S rRNA nucleotide positions on two hundred and ninety-five strains 128

Figure 2-5. The major steps in sequencing 16S rRNA genes..... 161

Figure 2-6. Sequencing gel output generated for the type strain of *Nocardia asteroides*... 175

Figure 2-7. An unrooted evolutionary tree derived from the 16S rDNA sequences. The tree was generated by using Jukes and Cantor's correction (Jukes & Cantor, 1969) and the neighbor-joining method (Saitou & Nei, 1987) 178

Figure 2-8. An unrooted evolutionary tree derived from 16S rRNA sequences. The tree was generated by using the Jukes and Cantor distance method (Jukes & Cantor, 1969) and the least squares algorithm (Fitch & Margoliash, 1967) 179

Figure 2-9. An unrooted maximum likelihood tree depicting relationships between representatives of the mycolic acid-containing taxa 180

Figure 2-10. The most parsimonious tree generated by using the maximum parsimony criterion 181

Figure 2-11. A phylogenetic tree depicting relationships between representatives of mycolic acid-containing taxa and selected outgroup strains 184

Figure 2-12. Summary of the regional variation analysis on five phyletic clades 188

Figure 2-13. Secondary structural differences in 16S rRNA sequences of *Nocardia pinensis* N1164^T and related mycolic acid-containing actinomycetes. 193

Figure 2-14. Phylogenetic trees depicting relationships found between representative nocardiae 196

Figure 2-15. Oligonucleotide signatures that distinguish the two *Nocardia* rRNA subgroups 198

Chapter III

Figure 3-1. Schematic representation of a typical activated sludge treatment plant 208

Figure 3-2. Stoke Bardolph Sewage Works 216

Figure 3-3. Mass spectral fragmentation pathways characteristic of mycolic acid methyl esters 236

Figure 3-4. Partial mass spectrum of the methyl esters of mycolic acids extracted from isolate N1171 237

Figure 3-5. Pyrolysis gas chromatographic analysis of mycolic acid methyl esters extracted from isolate N1171 238

Figure 3-6. High-performance liquid chromatogram showing the menaquinone profile of isolate N1171 240

Figure 3-7. (a) Dendrogram and (b) ordination plot showing relationships between the sludge isolates and the *Tsukamurella paurometabola* strains based on 45 phenotypic characters 245

Figure 3-8. Differences in the secondary structure of the 16S rRNA of isolate N1171 and *Tsukamurella paurometabola* strains JC7^T, M334 and N663 250

Figure 3-9. An unrooted phylogenetic tree showing relationships between isolate N1171 and representatives of mycolic acid containing taxa 251

Figure 3-10. Pyrolysis mass spectra derived from the analysis of (a) *Tsukamurella paurometabola* strain JC7^T, (b) *Tsukamurella paurometabola* strain M334, (c) *Tsukamurella paurometabola* strain N663, and (d) sludge isolate N1171 ...254

Figure 3-11. (a) Dendrogram representing relationships found between the sludge isolates and *Tsukamurella paurometabola* strains based on pyrolysis mass spectral data. (b) Three dimensional ordination plot derived from the pyrolysis mass spectral analysis 256

Figure 3-12. Dendrogram showing relationships between the sludge isolates and the *Tsukamurella paurometabola* strains. The dendrogram is based on mean correlation coefficient (*r*) values which were grouped by using the unweighted pair group method with arithmetic averages algorithm 257

Figure 3-13. Differences in secondary structure found between 16S rDNA sequences of *Tsukamurella paurometabola* JC7^T when data from the present study were compared with corresponding data from Rainey *et al.* (1995a)..... 262

Chapter IV

Figure 4-1. Schematic representation of an automated Curie point pyrolysis mass spectrometer..... 269

Figure 4-2. Structural comparison of the biological neuron and the processing element (PE) 282

Figure 4-3. The sigmoid and step transfer functions 284

Figure 4-4. Typical structure of a multi-layered artificial neural network which consists of three layers, namely input, hidden and output layers 285

Figure 4-5. Major steps involved in multivariate statistical analysis of pyrolysis mass spectra..... 297

Figure 4-6. The backpropagation algorithm 301

Figure 4-7. Test topologies of the neural networks 303

Figure 4-8. The procedure used for the identification of *Streptomyces* group A, B and C strains using an artificial neural network with an additional neuron..... 304

Figure 4-9. Pyrolysis mass spectra of representatives of *Streptomyces* groups A, B and C..... 308

Figure 4-10. Training curves of six neural networks. Three were trained with the normalised data and the other three with the scaled training data in each case using the forty-eight pyrolysis mass spectra from dataset 1A 309

Figure 4-11. Dendrograms showing relationships between representative *Streptomyces* strains based on: (a) dataset 1A and (b) dataset 1B..... 310

Figure 4-12. Training curves of network NET-1AS 314

Figure 4-13. Training curves of network NET-1AA 317

Figure 4-14. Results of multivariate statistical analysis of the combined dataset which was composed of datasets 1A, 2 and 3 319

Figure 4-15. Dendrogram obtained from the PC-CVA analysis of information held in dataset 4..... 323

Figure 4-16. Dendrogram showing relationships between representative *Streptomyces* strains based on multivariate statistical analysis of the pyrolysis mass spectra results held in dataset 5. 325

Figure 4-17. Dendrogram showing relationships based on multivariate statistical analysis of the pyrolysis mass spectra of dataset 6..... 326

Figure 4-18. Principal component-canonical variate analysis of the combined dataset comprising the mass spectra of datasets 1A and 6..... 327

<BLANK>

CHAPTER I.

MODERN BACTERIAL SYSTEMATICS: THEORETICAL BACKGROUND AND DEVELOPMENT OF SOFTWARE TOOLS

A. Introduction

1. Phylogenetic reconstruction from molecular sequence data

Phylogeny is the study of the evolutionary history of organisms. Cladistic relationships indicate the degree of relatedness between microorganisms as shown by pathways of ancestry (Cain & Harrison, 1960). Consequently, classifications which are based on perceived evolutionary relationships between organisms reflect the extent of change over time. Phylogenetic relationships between organisms are represented by evolutionary trees and are inferred from various types of phenetic relationships based on assumptions of how evolution occurs. Evolutionary systems are sometimes seen merely as simple branching over time but this is an oversimplification as *in vivo* hybridisation and lateral gene transfer lead to conceptual as well as to computational difficulties (Sneath 1974a, 1975; Maynard Smith, 1990).

Bacterial systematics is increasingly being based on phylogenetic information, notably that derived from macromolecules such as DNA, RNA and proteins (Woese, 1987; Stackebrandt, 1992; Ludwig *et al.*, 1993; Ludwig & Schleifer, 1994; Olsen *et al.*, 1994a). Classification, the one of basic disciplines of bacterial systematics, is becoming increasingly dependent on the use of molecular sequence data, notably on information generated from 16S rRNA analyses (Stackebrandt, 1992; Woese, 1992). Since classification is a prerequisite of

accurate identification it is perhaps not surprising that identification of unknown bacteria is also increasingly focused on molecular sequence data. These revolutionary changes in microbial systematics are mainly based on improvements in nucleotide sequence acquisition and on the development of methods for phylogenetic analysis. Classifications based on molecular sequence data are often considered to be "*natural*" (Woese, 1992).

Nucleotide sequence similarity values now widely serve as one of the standard taxonomic criteria used at the species level along with estimates of DNA relatedness values (Stackebrandt & Goebel, 1994). Phylogenetic relationships at higher taxonomic rank are mainly determined by constructing phylogenetic trees, that is, by what is known as phylogenetic reconstruction. This procedure involves three sequential steps, namely choice of macromolecule, alignment and construction of phylogenetic trees. A detailed review of ribosomal RNA techniques used in microbial systematics is given in Chapter II. This chapter is limited to considerations of numerical methods used to handle molecular sequence data.

1.1. Choice of molecule

An ideal molecular chronometer have meet the following specifications: (*i*) sequence changes should occur as randomly as possible, that is, by clock-like behaviour; (*ii*) rates of change should be commensurate with the spectrum of evolutionary distances measured; and (*iii*) the macromolecule should be large enough to provide sufficient information and should be a "*smooth-running*" chronometer (Woese, 1987).

Sequence information that is not under any selective constraint is ideal but only for comparisons between recently evolved organisms since mutation rates are so high. Woese (1987) pointed out that small and large subunit ribosomal (r) RNA sequences provide the most useful and suitable chronometer for phylogenetic

analyses as they: (*i*) show a high degree of functional constancy which assures relatively good clockwise behaviour; (*ii*) occur in all organisms; (*iii*) show different rates of mutation along different parts of the macromolecule thereby allowing distant phylogenetic comparisons to be made; and (*iv*) can be sequenced directly using the reverse transcriptase sequencing technique (Lane *et al.*, 1985). The development and application of the polymerase chain reaction (PCR; Saiki *et al.*, 1988) has made the reverse transcriptase redundant.

Prokaryotes contain three types of ribosomal RNA, namely 5S, 16S and 23S rRNAs. 5S rRNA is not usually considered to be a suitable molecule for phylogenetic analyses since its small size, that is, *ca.* 120 nucleotides, does not allow statistically significant sampling (Woese, 1987; Hillis & Dixon, 1991). Nevertheless, comparative 5S rRNA sequencing studies can be used to clarify fine evolutionary relationships between closely related prokaryotes (Hori & Osawa, 1986; Van den Eynde *et al.*, 1990) as witnessed by studies on coryneform bacteria (Park *et al.*, 1987), micrococci (Dekio *et al.*, 1982), mycobacteria (Dams *et al.*, 1987), streptomycetes (Park *et al.*, 1991), streptosporangias (Kim, 1995) and members of the family *Microbacteriaceae* (Park *et al.*, 1993).

Until recently, deductions about the evolution of prokaryotes were almost exclusively based on data derived from 16S rRNA; this macromolecule consists of approximately 1500 nucleotides. Relatively little work has been done on the more complex 23S rRNA which consists of approximately 3300 nucleotides. Reasonably good agreement has been found between evolutionary trees based on 16S and 23S rRNA sequence data (Ludwig *et al.*, 1992). To date, 16S rRNA from more than 2000 prokaryotes has been examined with much of the resultant information held in the GenBank/EMBL database and by the ribosomal database project (RDP; Larsen *et al.*, 1993).

It is important to check phylogenies generated from analyses of 16S rRNA in order to ensure that they reflect the evolution of organisms and not their own evolutionary history. Alternative macromolecules, usually proteins, have been used to generate phylogenies. These alternative molecules include elongation factors (Ludwig *et al.*, 1990; Morden *et al.*, 1992; Ludwig *et al.*, 1993), ATPase subunits (Amman *et al.*, 1988; Iwabe *et al.*, 1989; Gogarten *et al.*, 1989; Klugbauer *et al.*, 1992; Morden *et al.*, 1992; Ludwig *et al.*, 1993), RNA polymerases (Pühler *et al.*, 1989; Zillig *et al.*, 1989), and ribosomal proteins (Ochi *et al.*, 1993; Liao & Dennis, 1994; Ochi & Hiranuma, 1994; Ochi, 1995). In addition, phylogenies have been generated by analyses of *GroEL* (Viale *et al.*, 1994) and *gyrB* (Yamamoto & Harayama, 1995) gene products. Phylogenies deduced from these macromolecules showed good congruence with those derived from ribosomal RNAs (Ludwig & Schleifer, 1994).

The choice of the molecule to be sequenced depends on the aims of the study, notably on the taxonomic rank under consideration. Thus, highly variable genes, such as the intergenic spacer region in ribosomal RNA operons, are most appropriate for comparisons between strains at species and subspecies levels (Frothingham & Wilson, 1993). In contrast, more conserved genes, such as 16S rDNA and those coding for elongation factors and RNA polymerases, are used to unravel relationships at higher taxonomic rank. Genes coding for proteins generally show more variability than ones coding for ribosomal RNA (Olsen & Woese, 1993; Yamamoto & Harayama, 1995).

1.2. Alignment

The first step in phylogenetic reconstruction from molecular sequence data is sequence alignment. This involves finding homologous sites, that is, positions derived from the same ancestral organism, in the molecules under study. A set of

sequences can be aligned against one another by introducing '**alignment gaps**'. The general rule underpinning multiple sequence alignment is that the increase in sequence similarity due to the introduction of alignment gaps must be greater than that which would be expected due to random alignment (Olsen, 1988). Several workers have developed multiple sequence alignment procedures (Bains, 1986; Higgins & Sharp, 1988, 1989; Barton & Sternberg, 1987; Feng & Doolittle, 1987; Santibenez & Rohde, 1987; Sobel & Martinez, 1986; Taylor, 1987). Since these alignment methods are purely computational it is necessary to check alignments manually and, if possible, in light of biological function in order to clarify ambiguous regions (e.g., Vohra *et al.*, 1992; Viale *et al.*, 1994).

In general, protein sequences are obtained indirectly by determining the responsible gene sequences though small proteins may be sequenced directly (e.g., ribosomal protein AL-30; Ochi, 1995). Phylogenies can be inferred from both types of molecular sequences, namely amino acid and nucleotide sequences. However, studies based on amino acid sequences are preferable since the evolutionary behaviour and constraint on individual codon positions in protein coding genes may vary (e.g., Klenk and Zillig, 1994).

Most attention has been given to studying sequence alignment of RNAs, especially ribosomal RNAs (Olsen, 1988; James *et al.*, 1989). It is possible to fit sequenced regions which have little primary structural similarity into a common secondary structure since the secondary structure of rRNAs are largely known (Woese *et al.*, 1983; Noller, 1984; Gutell, 1993; Olsen & Woese, 1993). This biologically meaningful alignment procedure has been widely used because of the ready availability of pre-aligned rRNA sequences held by the ribosomal database project (RDP; Larsen *et al.*, 1993). The secondary structures of 16S and 23S rRNA of prokaryotes have recently been compiled (Gutell, 1993; Gutell *et al.*, 1993; Neefs

et al., 1993). A method for automated detection of secondary, or higher order structure, in rRNA sequences was developed by Winker *et al.* (1990).

1.3. Construction of phylogenetic trees

The results of phylogenetic analyses are usually presented in the form of an '**unrooted tree**' in which the earliest point in time, the location of the common ancestor, is not identified. Members of present day taxa correspond to '**terminal nodes**', or tips, while branching points within a tree are called '**internal nodes**'. A strictly bifurcating unrooted tree has *T* terminal nodes, which correspond to taxa, and *T-2* internal nodes. Such a tree has *2T-3* branches. The total number of possible unrooted trees for *T* taxa is:

$$Number\ of\ trees = \prod_{i=3}^{T}(2i-5)$$ (Felsenstein, 1978a)

where *T* is the number of taxa.

The number of possible rooted trees is increased by a factor of *2T-3* since the root of the tree can be placed along any of the *2T-3* branches. The aim of phylogenetic reconstruction is to find the best estimated tree. Three tree-making procedures are commonly used in bacterial systematics, namely the distance, parsimony and maximum likelihood methods.

1.3.1. Methods based on pairwise distances

These methods involve two consecutive procedures, namely transformation of sequence similarity data to evolutionary distances and construction of trees from information in distance matrices. Evolutionary distance data are generally categorised into two types, namely additive and ultrametric distances (Swofford & Olsen, 1990). Additive distances mathematically satisfy the four point condition. Thus, for four taxa *A*, *B*, *C* and *D*:

$$max(d_{AB}+d_{CD}, d_{AC}+d_{BD}, d_{AD}+d_{BC}) = mid(d_{AB}+d_{CD}, d_{AC}+d_{BD}, d_{AD}+d_{BC})$$

where d_{ij} is the distance between taxa *i* and *j*, *max* is the maximum value function, and *mid* is the middle value (median) function. This formula can be recast using the more standard *min* function as:

$$2\max(d_{AB}+d_{CD}, d_{AC}+d_{BD}, d_{AD}+d_{BC}) + \min(d_{AB}+d_{CD}, d_{AC}+d_{BD}, d_{AD}+d_{BC}) = d_{AB}+d_{CD}+d_{AC}+d_{BD}+d_{AD}+d_{BC}$$

Additive distances can be fitted to an unrooted tree so that all pairwise distances are equal to the sum of the branch lengths that connect the respective taxa or external nodes (Figure 1-1a). Relationships are displayed in an unrooted format as additive trees are not based on any assumptions about tree rooting. Many of the available tree-making methods are used to estimate the additive model, these include the neighbor-joining (Saitou & Nei, 1987) and weighted least-squares methods (Fitch & Margoliash, 1967).

Ultrametric distances are the most constrained and are defined mathematically by meeting the three-point condition. Ultrametric inequality requires that for any three taxa, such as *A*, *B* and *C*:

$$d_{AC} \leq \max(d_{AB}, d_{BC}) \text{ or}$$

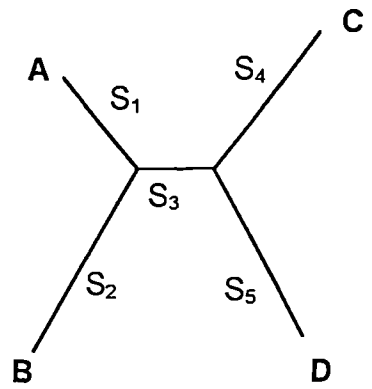
$$\max(d_{AB}, d_{BC}, d_{AC}) = mid(d_{AB}, d_{BC}, d_{AC})$$

, that is, the two greatest distances are equal. This equation can be recast as:

$$2\max(d_{AB}, d_{BC}, d_{AC}) + \min(d_{AB}, d_{BC}, d_{AC}) = d_{AB}+d_{BC}+d_{AC}$$

Ultrametric distances are fitted to form a tree so that the distance between any two taxa is equal to the sum of the branches joining them and the tree can be rooted so that all of the taxa are equidistant from the root (Figure 1-1b). This latter assumption is based on the universal '**molecular clock**' theory (Kimura, 1983), namely that all

(a)



Additive properties:

$P_{AB} = S_1 + S_2$

$P_{AC} = S_1 + S_3 + S_4$

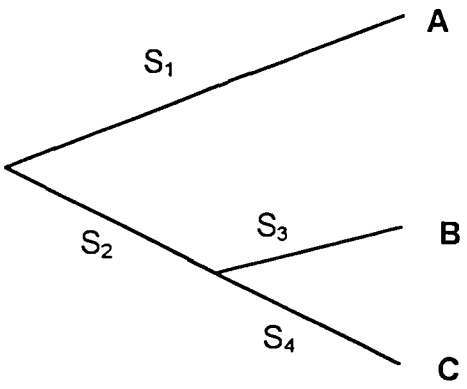
$P_{AD} = S_1 + S_3 + S_5$

$P_{BC} = S_2 + S_3 + S_4$

$P_{BD} = S_2 + S_3 + S_5$

$P_{CD} = S_4 + S_5$

(b)



Additive properties:

$P_{AB} = S_1 + S_2 + S_3$

$P_{AC} = S_1 + S_2 + S_4$

$P_{BC} = S_3 + S_4$

Ultrametric properties

$S_3 = S_4$

$S_1 = S_2 + S_3 = S_2 + S_4$

FIGURE 1-1. ADDITIVE AND ULTRAMETRIC TREES.

lineages are equally diverged. Methods based on ultrametric distances include cluster analysis techniques such as the unweighted pair group method using arithmetic averages (UPGMA; Sneath & Sokal, 1973) and the weighted pair group method using arithmetic averages (WPGMA). These clustering algorithms have been used to analyse data on antibody cross-reactivity, DNA/DNA hybridisation, pairwise sequence similarities and 16S rRNA catalogue similarities. These methods, which would work perfectly if data were only generated by a clock-like evolution (Felsenstein, 1988), are relatively sensitive to lineage-to-lineage differences in evolutionary rates (Colless, 1970).

Transformation of sequence data to evolutionary distances. The simplest way of representing relationships between members of two taxa is by determining their molecular sequence similarity. The most frequently used method for calculating similarities from aligned sequences is based on the equation:

$$S = M / L \text{ (Olsen, 1988)}$$

where $L = M + U + w_G G$, S is the sequence similarity, M is the number of alignment positions with synonymous residues, L is an effective sequence length, U is the number of alignment positions with nonsynonymous residues, w_G is the weight given to alignment gaps, and G is the number of alignment positions with a gap in one juxtaposed with a residue in the other sequence. The weighting value (w_G) varies from zero (ignoring gaps) to one (equivalent to a substitution). A value of 0.5 is a reasonable compromise though gaps are usually ignored in phylogenetic analyses of bacteria (Swofford & Olsen, 1990).

Different alignment positions can be differentially weighted. Stahl *et al.* (1984) used different weights (w_G) in an analysis of 5S rRNA sequence data in which nucleotide positions representing Watson-Crick pairing were assigned half the weight of unpaired nucleotides thereby reflecting their lack of independence. The

same principle can be applied to assign different weights on the basis of inferred substitution rates of individual positions thereby allowing emphasis to be placed on the more conserved regions (Ludwig & Schleifer, 1994). Similarities based on sequence data provide a simple way of recognising closely related taxa. However, it is generally more informative to infer an evolutionary tree from nucleotide sequence data with the exception of identical or nearly identical sequences (Olsen, 1988).

The observed number of nucleotide sequence differences is not synonymous with the number of fixed mutations. This is primarily a consequence of the occurrence of multiple mutations at a sequence position. The one-parameter model proposed by Jukes and Cantor (1969) assumes that there is independent change at all of the positions hence there is an equal probability of ending up with each of the other three bases. This model is also based on the assumption that base composition does not vary over time. The transformed evolutionary distance (*d*), which represents the number of substitutions per sequence position, is given in the following equation:

$$d = -b\log\left(1-\frac{D}{b}\right)$$

where *D* is the dissimilarity between two sequences, that is, (1-*S*), and *b* is a coefficient that varies with the particular model and data type. In the Jukes and Cantor model, *b* is 3/4, which means that the fractions and substitution frequencies of the four different bases are equal throughout evolution and thereby evolutionary distance cannot be calculated when dissimilarity is over 0.75.

Several other distance transformation methods based on this equation have been generated. The two-parameter model (Kimura, 1980) provides for differences between transition and transversion rates. In the Kimura model, evolutionary distance is expressed as:

$$d = -\frac{1}{2}\log\left[(1-2P-Q)\sqrt{1-2Q}\right]$$

$$P = U_P / L$$

$$Q = U_Q / L$$

$$L = M + U_P + U_Q$$

where P is the fraction of sequence positions differing by transition; Q is the fraction of sequence positions differing by transversion; U_P is the number of positions differing by transition; U_Q is the number of positions differing by transversion; and M, the number of positions in which two sequences have identical nucleotides.

Each of the models outlined above is based on the assumption that all four nucleotides occur with equal frequency. A more general formula can be applied when the four nucleotides are present in unequal proportions:

$$d = -b\ln(1-\frac{D}{b}) \text{ and } b = 1-\sum_{i\in R} f_i^2 \text{ (Tajima \& Nei, 1984)}$$

where R is a set of possible residue types (e.g., {A,C,G,T} for the DNA sequence) and f_i is the frequency of the ith type of residue in the sequences under comparison. Base composition values can be determined separately for each pair of sequences (Woese *et al.*, 1990a) or can be the mean composition for all of the analysed sequences (Swofford & Olsen, 1990). When transversions are considered alone , the transversion distance is:

$$d = -b\ln(1-\frac{D}{b}) \text{ (Woese } et\ al.\text{, 1991)}$$

and

$$b = 1-[(f_A+f_G)^2+(f_C+f_T)^2]$$

where D is the dissimilarity between two sequences and f_N is the frequency of the base N. Jin and Nei (1990) developed a method based on the Kimura model of

base substitution where the rate of substitutions was assumed to vary from site to site according to a gamma distribution.

Another approach used to estimate evolutionary distances with consideration of gaps was introduced by Van de Peer *et al.* (1990). In this method:

$$d = -\frac{3}{4}\ln\left[1-\frac{4}{3}\left(\frac{U}{M+U}\right)\right]\left(1-\frac{G}{L}\right)+\frac{G}{L}$$

where M is the number of identical nucleotides, U is the number of positions showing a substitution, G is the number of gaps in one sequence with respect to the other, and L is the sum of I, S and G. This approach has been applied to infer the phylogenies of strains of *Candida* (Hendriks *et al.*, 1991), basidiomycetous yeasts (Van de Peer *et al.*, 1992) and all life forms (Neefs *et al.*, 1993; Van de Peer *et al.*, 1994).

Neighbor-joining method. The neighbor-joining method (Saitou & Nei, 1987) is an algorithm for inferring additive trees. It is theoretically related to clustering methods, such as the UPGMA, but is not based on the assumption that data are ultrametric and that all lineages have equally diverged. In contrast to cluster analysis, the neighbor-joining method keeps track of nodes on the tree rather than taxa or clusters of taxa. The evolutionary distance matrix is provided as input data and a modified distance matrix is constructed in which the separation between each pair of nodes is adjusted on the basis of their average divergence from all of the other nodes. This procedure leads to the normalisation of the divergence of each taxon for its average clock rate. As the neighbor-joining algorithm seeks to represent the data by an additive tree, a negative length can be assigned to a branch. The computational steps involved in this method are shown below:

(a) Given a matrix of pairwise distances (d), for each terminal node i calculate its net divergence (r_i) from all other taxa using the formula:

$$r_i = \sum_{k=1}^{N} d_{ik}$$

where N is the number of terminal nodes in the matrix.

(b) Create a rate-corrected distance matrix (M) in which the elements are defined by the equation:

$$M_{ij} = d_{ij} - (r_i + r_j)/(N-2)$$

for all i and with $j>i$ (the matrix is symmetrical). Only the values i and j, for which M_{ij} is minimum, need to be kept for the next step.

(c) Define a new node, u, with the three branches that join nodes i, j and the rest of the tree. Define the lengths of the tree branches from u to i and j using the equation:

$$S_{iu} = \frac{d_{ij}}{2} + \frac{(r_i - r_j)}{2(N-2)}$$

$$S_{ju} = d_{ij} - S_{iu}$$

where S_{iu} is the length of the path connecting nodes i and u.

(d) Define the distance from u to each of the other terminal nodes (for all $k \neq i$ or j):

$$d_{ku} = \frac{(d_{ik} + d_{jk} - d_{ij})}{2}$$

(e) Remove distances to nodes i and j from the data matrix and decrease N by 1.

(f) Go back to step 1 if more than two nodes remain. Otherwise, the tree is finally defined apart from the length of the branch joining the two remaining nodes (i and j). Let this remaining branch be $S_{ij} = d_{ij}$.

Least-squares methods. A variety of techniques are available for fitting a given set of pairwise evolutionary distance estimates to an additive tree. A concrete definition of the net disagreement between the tree and the original data, as an objective function to be minimised, is required. Fitch and Margoliash (1967) introduced the first distance matrix method; Cavalli-Sforza and Edwards (1967) independently invented a similar algorithm.

The error (E) of fitting distance estimates to a tree is defined as:

$$E = \sum_{i=1}^{T-1} \sum_{j=i+1}^{T} w_{ij} \left| d_{ij} - p_{ij} \right|^{\alpha}$$ (Swofford & Olsen, 1990)

where T is the number of taxa, w_{ij} is the weight applied to the separation of taxa i and j, d_{ij} is the distance between taxa i and j, p_{ij} is the length of the path connecting i and j in a given tree, and α is 1 or 2. The value of α and a weighting scheme must be chosen for the analysis.

Setting α to 2 represents a weighted least-squared criterion, that is, the weighted square deviation of the tree path lengths from the distance estimates will be minimised. If α is 1, the weighted absolute differences will be minimised. The least-squares criterion is preferred if the errors in the distance estimates are distributed uniformly across the data (Swofford & Olsen, 1990).

Four possible weighting schemes can be used:

$w_{ij} = 1$ (Cavalli-Sforza & Edwards, 1967), $w_{ij} = d_{ij}^{-1}$ (Fitch & Margoliash, 1967), $w_{ij} = d_{ij}^{-2}$, and $w_{ij} = \sigma_{ij}^{-2}$

where σ is the expected variance of measurements of d_{ij}.

Two procedures are generally required to find the phylogenetic tree with the lowest E value, namely optimising the branch lengths given a tree topology and finding a tree topology with the lowest E value of all of the possible trees. De Soete

(1983) proposed a method in which these two steps were combined into a single step.

The equation can be recast in terms of the branch lengths in order to find the branch lengths that minimise the *E* value. There are 2*T*-3 independent branches that define the p_{ij} values in an unrooted tree with *T* external nodes. Let *A* be a matrix such that the element $A_{(ij)k}$ is equal to 1 if branch *k* is part of the path connecting taxa *i* and *j*; otherwise $A_{(ij)k}$ is equal to 0. Using this definition:

$$p_{ij} = \sum_{k=1}^{2T-3} A_{(ij)k} S_k$$

Combining this equation with $E = \sum_{i=1}^{T-1} \sum_{j=i+1}^{T} w_{ij} \left| d_{ij} - p_{ij} \right|^{\alpha}$ yields,

$$E = \sum_{i=1}^{T-1} \sum_{j=i+1}^{T} w_{ij} \left| d_{ij} - \sum_{k=1}^{2T-3} A_{(ij)k} S_k \right|^{\alpha}$$

The branch lengths, S_k, remain undetermined after the value of α and a weighting scheme have been chosen. The solution for the minimal *E* value involves linear or quadratic programming either by using iterative successive refinement techniques or by solving a set of simultaneous equations using ordinary linear algebra (Olsen, 1988).

1.3.2. Parsimony methods

If each site in a set of nucleotide sequences has changed only once in the evolution of a group then the newly-arisen base will be shared by all members of species descended from the lineage in which the change occurred. If this were the case at all of the sites, the sets of species having the new bases would be either perfectly nested or disjointed, that is, they would not overlap unless one set of species was included in the other. It would ,therefore, be possible to construct a

phylogenetic tree on which the evolutionary traits of the group could be explained with only a single change at each site. There would be a conflict between the information provided by different sites if some of these sequences overlapped without being nested. The different parsimony methods provide a means to resolve these problems.

The concept of parsimony, introduced by Edwards and Cavalli-Sforza (1964) using gene frequency data, was first applied to molecular sequence data by Eck and Dayhoff (1966). In general, parsimony methods for inferring phylogeny operate by selecting trees that minimise the total tree length, that is, the number of evolutionary steps (base substitutions) required to explain a given set of data.

Different parsimony methods apply different constraints on character-state changes. The parsimony method developed by Fitch (1971) was based on the simplest model and applied disordered multistate-characters, such as nucleotide and protein sequences. In contrast, Wagner parsimony, formalised by Kluge and Farris (1969) and Farris (1970), assumes that transformation of one character state to another implies a transformation through any intervening state, as defined by the ordering relationship. Both methods permit the reversibility of the tree, that is, transformation in character states can be in either direction between nodes. The trees generated by these methods , therefore, are unrooted and the different rootings do not cause changes in the branch lengths, as represented by the number of steps.

The Fitch and Wagner parsimony criteria are based on the assumption that the probabilities of character changes are symmetrical (e.g., the probabilities of transformations from character 0 to 1 and 1 to 0 are the same). The Dollo parsimony method (Farris, 1977) is used to find the most parsimonious tree using an asymmetrical criterion on transformation. This model is based on the assumption

that if a hypothetical ancestor is included in the analysis each character state originates once on the tree and that any required **homoplasy** (*i.e.*, convergence, parallelism or reversal) takes the form of reversals to a more ancestral condition. The Dollo parsimony method is usually applied in the analysis of restriction site data in which the gain of bands is more difficult than the loss of bands (DeBry & Slade, 1985). It is also possible to construct an unrooted tree using the Dollo parsimony method.

Transformations based on transversion can be weighted by coding pridine (A and G) as *R* and pyrimidine (C and T) as *Y* so that the transitions are not considered. This procedure is not appropriate when closely related taxa are compared (Swofford & Olsen, 1990).

Parsimony methods are subject to systematic errors when large lineage-to-lineage variations in evolution rate are combined with large amounts of sequence change (Felsenstein, 1978b).

1.3.3. Maximum likelihood method

The most general method of deriving statistical estimates is the maximum likelihood method (Felsenstein, 1988). The concept was introduced by Cavalli-Sforza and Edwards (1967) for restriction data and by Felsenstein (1981a) for molecular sequence data. The method has been reviewed by Saitou (1990).

The likelihood of a tree (*T*) can be defined as *P(D;T,M)* given a model of evolution (*M*) and actual data (*D*). Likelihood can, therefore, be considered as a function of the tree. The probability of all possible sets of data must add up to one. However, the different values of *P(D;T,M)* need not add up to one when the data are held constant and the tree is varied. In such cases, the probabilities are called likelihoods (Felsenstein, 1988). A maximum likelihood approach to phylogenetic

inference is used to evaluate the net likelihood that the given evolutionary model will yield the observed sequences; the inferred trees are those with the highest likelihood.

The least-squares and parsimony methods define the optimal tree in terms of an optimal criterion when an evolutionary tree is inferred from multiple sequences. In the maximum likelihood method, the tree is given and the task is to determine how good it is. The method for evaluating the likelihood of a given tree proceeds from a hypothetical root node at any convenient location in the tree and it combines the likelihoods of each of its daughter trees, that is, descendant trees. The choice of root location does not change the likelihood of the tree.

The mathematical expression of a model of evolution can be given as a table of substitution rates per unit evolutionary distance at each site, *R*. In order to convert rates to a model of evolution each matrix is defined as a set of four simultaneous, linear differential equations:

$$\frac{\partial}{\partial t}\begin{bmatrix}\mathbf{A}\\ \mathbf{C}\\ \mathbf{G}\\ \mathbf{T}\end{bmatrix}=\begin{bmatrix}\mathbf{R}_{AA}\mathbf{R}_{CA}\mathbf{R}_{GA}\mathbf{R}_{TA}\\ \mathbf{R}_{AC}\mathbf{R}_{CC}\mathbf{R}_{GC}\mathbf{R}_{TC}\\ \mathbf{R}_{AG}\mathbf{R}_{CG}\mathbf{R}_{GG}\mathbf{R}_{TG}\\ \mathbf{R}_{AT}\mathbf{R}_{CT}\mathbf{R}_{GT}\mathbf{R}_{TT}\end{bmatrix}\begin{bmatrix}\mathbf{A}\\ \mathbf{C}\\ \mathbf{G}\\ \mathbf{T}\end{bmatrix}$$

where *A*, *C*, *G* and *T* are the probabilities that the nucleotide at the given site is A, C, G and T, respectively. These probabilities are analytically integrated with respect to evolutionary distance, *d*, to give a matrix, *M(d)*, in which element $M_{ij}(d)$ is the probability that a nucleotide of initial identity *i* has identity *j* after evolving through a distance *d*. This integration accounts for all possible series of substitutions linking the initial and final residues, providing an intrinsic correction for multiple substitutions. For instance, the integrated substitution matrix for Jukes and Cantor's one parameter model (1969) is:

$$M(d)=\begin{bmatrix} a & b & b & b \\ b & a & b & b \\ b & b & a & b \\ b & b & b & a \end{bmatrix}$$

where $a=\frac{1+3e^{-4\alpha d}}{4}$ and $b=\frac{1-e^{-4\alpha d}}{4}$.

When two nucleotide sequences are involved, the likelihood of observing that a nucleotide x_{1j} in the first sequence will be changed to a nucleotide x_{2j} in the second sequence at a site j under the condition that the sequences are separated by distance d can be calculated as:

$$L(x_{1j},x_{2j};d)=f_{x1j}M_{x1j,x2j}(d)$$

where f_{x1j} is the probability that the first sequence has a nucleotide x_{1j}.

If all of the sites are independent, the overall likelihood of finding sequence x_1 and sequence x_2 separated by distance d is the product of the likelihoods at each position:

$$L(x_1,x_2;d)=\prod_{j=1}^{N} f_{x1j}M_{x1j,x2j}(d)$$

The equation is usually presented in its logarithmic form:

$$\log L(x_1,x_2;d)=\sum_{j=1}^{N}\log\left[f_{x1j}M_{x1j,x2j}(d)\right]$$

where N is the number of sites in the compared sequences. For cases involving three or more sequences overall likelihoods between nodes are computed for a given tree topology and the combination shows the highest likelihood chosen as the solution (the best tree or maximum likelihood tree).

Methods for finding optimal branch lengths have been proposed by Felsenstein (1981b). Fukami and Tateno (1989) demonstrated that for the Jukes and Cantor model of evolution there was no problem of local minima while finding optimal branch lengths. Tillier (1994) developed a multiple-parameter model of the maximum likelihood method and applied it to the three-parameter case of RNA sequence data.

The maximum likelihood method is the most statistically sound way of reconstructing phylogeny (Felsenstein, 1988). However, the use of the method is mainly hampered by its computational cost as demonstrated by Olsen *et al.* (1994b). The number of taxa analysed at a time is normally below twenty (Olsen *et al.*, 1994b). The significance of this approach to bacterial systematics has been emphasised by Ludwig and Schleifer (1994).

1.4. Rooting the phylogenetic tree

Most methods outlined above do not specify the root in the tree, that is, they are used to generate unrooted trees. If a rooted tree is desired, the root must be located using extrinsic information (Swofford & Olsen, 1990). The most commonly used method for rooting an unrooted tree is to include single or many outgroups that are assumed to lie cladistically outside a presumed monophyletic group (Olsen, 1988; Ludwig & Schleifer, 1994). If there is a single branch on the unrooted tree that partitions the ingroup taxa from the outgroup taxa then the tree is consistent with the assumption of ingroup monophyly.

It is well known that most methods of estimating pairwise distance tend to underestimate long branches and hence cause fundamental errors (Swofford & Olsen, 1990). It is, therefore, important for distance methods to include an outgroup, which is not too far away from the ingroup and to use the minimum number of

outgroups. In contrast, for parsimony methods, the inclusion of multiple outgroups can be more effective than the use of a single outgroup (Swofford & Olsen, 1990).

The appropriate rooting procedure may be essential in bacterial classification where it is important that the ingroup is monophyletic. There is some confusion on this matter. It was mentioned earlier that the outgroup taxa used to identify the root should be as close as possible to the ingroup (Swofford & Olsen, 1990). However, Ludwig and Schleifer (1994) argued that only a moderately related taxon should be used as the outgroup though they did not clearly specify why this was so. It is far from clear what is meant by terms such as close and moderately close. In addition, it is not always possible to find the closest outgroup. In practice, it is sound to evaluate tree topologies with various combinations of outgroup strains in order to identify the position of the root.

A major rooting problem is apparent in the case of relationships between Archaea, Bacteria and Eukarya (Woese *et al.,* 1990b). In this case, the position of the root, that is, the universal ancestor, cannot be identified by the approach outlined above as it is clearly impossible to use outgroup taxa. Iwabe *et al.* (1989) and Gogarten *et al.* (1989) suggested that the universal root be placed between the Archaea and Bacteria on the basis of duplicated genes, namely those coding for elongation factors and ATPases; these genes are believed to have evolved before the separation of the three domains. The three-domain concept is widely accepted amongst biologists though minor disagreements are found between different molecular data and phylogenetic analyses (Lake, 1987a,b; Rivera & Lake, 1992).

An alternative method for rooting trees is midrooting where the root is simply the mid-point within the unrooted tree. This method is not valid when sampling nucleotide sequences that are not distributed equally and hence its application to most bacterial phylogenetic studies is not to be recommended.

1.5. Evaluation of evolutionary trees

An evolutionary tree should be evaluated to assess the significance of the tree topology and the length of branches. The procedures underlying statistical tests of phylogenies have been extensively reviewed (Felsenstein, 1988; Li & Gouy, 1990; Swofford & Olsen, 1990).

Bootstrap analysis is the most frequently used method for evaluating phylogenies. The technique was developed by Efron and Gong (1983) and introduced to phylogenetic studies by Felsenstein (1985). This method is also called the **'resampling method'** as it involves the generation of new data sets by random resampling of positions in the original data set. Some positions in the original data set are included one or more times in the derived data set but some are not considered at all. Felsenstein (1985) suggested that bootstrap values of over 95% can be regarded as significant support for monophyly.

The number of resamplings is an important factor in bootstrap analyses. Li and Gouy (1990) argued that the use of 100 replications was not sufficient to provide confidence of the test on the 95% bootstrap limit. Using binomial distribution, Hedges (1992) found that the minimum number of resamplings needed to obtain ±1% of significance at a bootstrap level of 95 % was 2000 replications and at a bootstrap level of 99% it was 400 replications. Resampling 1000 times is a possible compromise between accuracy and efficacy on cost of computing and is generally accepted by most bacterial taxonomists (Ludwig & Schleifer, 1994).

Several methods have been proposed to evaluate the significance of likelihood values. Felsenstein (1988) suggested that a one-degree of freedom χ^2 test on twice the difference in log-likelihoods was a conservative test. Kishino and Hasegawa (1989) provided an alternative test and obtained the same confidence interval as with a bootstrap analysis. The bootstrap approach for estimating

confidence intervals of branch lengths in an evolutionary tree was proposed by Dopazo (1994). It is widely accepted that the bootstrap method is the most adequate technique for evaluating branching patterns in an evolutionary tree, especially one that contains more than four taxa (Li & Gouy, 1990).

The **jackknife** method (Miller, 1974) is closely related to the bootstrap approach. In this method, the resampling of the original sequence dataset is dropped by *k* positions at a time and the level of support on each branch pattern calculated from the trees based on the resampled data. A typical value for *k* is one. The jackknife approach has rarely been used and has not been studied extensively.

1.6. Simulation studies on different phylogenetic inference methods

Huelsenbeck and Hillis (1993) examined the consistencies between sixteen phylogenetic inference methods under three different evolutionary processes by using computer simulations of a four-taxon case (in this instance, consistency means convergence on the true tree as the sample size becomes infinite). They found that the neighbor-joining and least-squares methods were consistent when evolutionary processes match assumptions made on evolutionary distances but when these assumptions were violated the two methods show inconsistencies. The parsimony method gave inconsistencies throughout the simulation thereby confirming the results of a previous study (Felsenstein, 1978b). The UPGMA method also gave inconsistent results irrespective of the distance model used. The inconsistencies of the different methods were generally increased with the number of sites included in the simulation.

Kim (1993) carried out a simulation study on an eight-taxon case using three methods, namely the UPGMA, neighbor-joining and maximum parsimony methods. He considered that a tree topology given by three different methods was likely to be a '*true tree*' and went to on to argue that sequence data should be analysed by

more than one method. Neither Huelsenbeck and Hillis (1993) nor Kim (1993) examined the maximum likelihood method.

Kuhber and Felsenstein (1994) examined five methods, namely the parsimony, compatibility, maximum likelihood, Fitch-Margoliash and neighbor-joining methods. They found that all of the methods performed well given equal rates of evolution, a fundamental assumption for all of these approaches. The maximum likelihood method was superior to the other methods given unequal rates of evolution per branch or per site though the distance methods performed better with short nucleotide sequences. Nevertheless, none of the methods provided accurate estimations of branch lengths under unequal evolution rates per site. It was also noted that the parsimony method gave the worst estimation under low and unequal evolutionary rates.

It is generally accepted from simulation studies that the neighbor-joining method provides a better estimation of phylogenies than the parsimony method given unequal rates of evolution per branch (Li *et al.*, 1987; Sourdis & Nei, 1988; Jin & Nei, 1990; *Kuhber & Felsenstein, 1994*). Similarly, the maximum likelihood method slightly outperforms the other methods, especially given unequal rates of evolution in branches (Hasegawa & Yano, 1984; Saitou & Imanish, 1989; Kuhber & Felsenstein, 1994).

Comparative studies of phylogenetic methods based on simulated or actual four-taxon data sets are often misleading because the performance of a method may be quite different for a four-taxon data set than for a larger data set. Generally, the results from simulation studies should be carefully evaluated since the evolutionary process is much more complex than simulated ones !

1.7. Misconceptions in bacterial phylogenies: a review of the recent literature

Felsenstein (1988) pointed out that most molecular evolutionists who use methods for inferring phylogenies do not take much interest in discussions of the properties of the mathematical methods since they focus on the difficult task of collecting data, that is, nucleic acid sequencing. Bacterial taxonomists often seem to regard phylogenetic analyses as a **'blackbox'** integrated into computer software. Examples of this phenomenon can readily be found in the literature. The following examples of inaccurate or incomplete presentation of methods used for phylogenetic analyses are taken from the International Journal of Systematic Bacteriology, volume 44, 1994:

(a) *"The ODEN program package was used to align the sequences, and phylogenetic distances were calculated by using both the unweighted pair group method and the neighbor-joining method"* (Ezaki *et al.*, 1994, page 130). In this study, the authors did not say how evolutionary distances were calculated.

(b) *"Evolutionary distance values were determined by using the neighbor-joining method"* (Briglia *et al.*, 1994, page 494). The method used in this study was designed for constructing trees from distance matrices not for generating distances !

(c) *"The dendrogram was constructed by using the software program PILEUP obtained from the Genetic Computer Group, Inc."* (Robertson *et al.*, 1994, page 836). The authors did not give any information on how data were analysed.

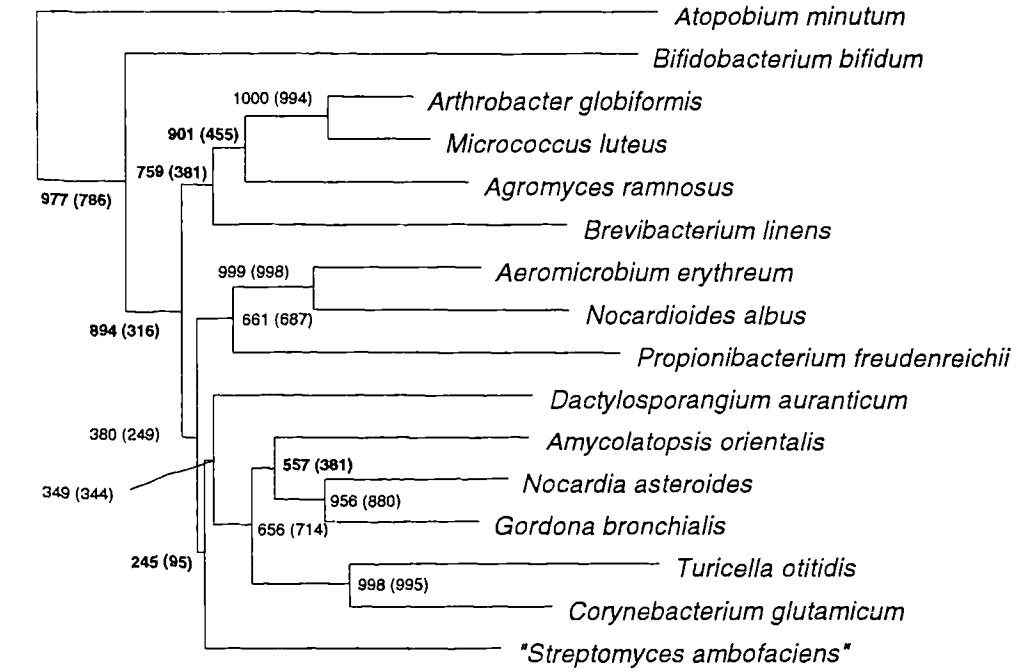
(d) *"An unrooted phylogenetic tree was produced by using the DNADIST and FITCH programs in the PHYLIP package"* (Dupuy *et al.*, 1994, page 461). It is not clear which distance method was used as the ***DNADIST*** program contains four different distance methods.

(e) A common omission in several papers was the failure to specify the distance models used (Cai & Collins, 1994, page 583; Collins *et al.*, 1994a, page 523; Collins *et al.*, 1994b, page 674; Collins *et al.*, 1994c, page 812; Enright *et al.*, 1994, page 387; Jagoueix *et al.*, 1994, page 379; Postic *et al.*, 1994, page 743; Takewaki *et al.*, 1994, page 159; Willems & Collins, 1994, page 591).

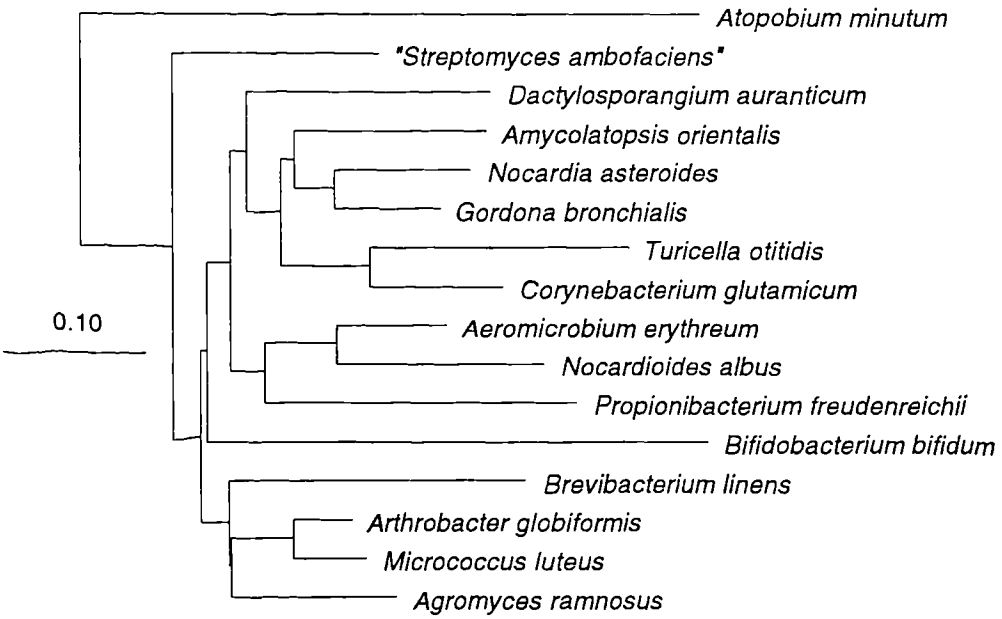
(f) Evolutionary distance and sequence similarity were confused by Enright *et al.* (1994, page 387).

Confusion is also apparent over the use of bootstrap analyses. Two levels of replications, namely 100 and 1000 resamplings, are most widely used. It is clear from the heuristic nature of bootstrap analyses that as the numbers of resamplings increase the accuracy of evaluation on the resultant phylogenies also increases. However, Ruimy *et al.* (1994a) justified the use of 100 replications over 1000 resamplings since "*their* ***experience*** *showed that increasing the number of bootstrap replications above 100 usually had only a small influence on the results compared with the choice of species*".

Some investigators have constructed phylogenetic trees using all of the aligned sites and applied only polymorphic sites for bootstrap evaluation (Rainey *et al.* 1995a; Yurkov *et al.*, 1994). It is evident from the algorithms that the parsimony method is not affected by the selection of polymorphic sites though this is not the case for the distance matrix and maximum likelihood methods. The effect of reducing data size by selecting polymorphic sites for phylogenetic analysis is illustrated in Figure 1-2. It is evident that the neighbor-joining tree based on all of the aligned sites (Figure 1-2a) differs significantly from the one based on the polymorphic sites (Figure 1-2b), especially with respect to the positions of



(a)



(b)

FIGURE 1-2. COMPARISON OF PHYLOGENETIC TREES BASED ON ANALYSES OF FULL NUCLEOTIDE POSITIONS AND POLYMORPHIC SITES. DISTANCE MATRICES WERE GENERATED ACCORDING TO JUKES AND CANTOR (1969), TREES CONSTRUCTED USING THE NEIGHBOR-JOINING METHOD (SAITOU & NEI, 1987) WITH *BACILLUS SUBTILIS* AS THE OUTGROUP. BOOTSTRAP VALUES WERE BASED ON 1000 RESAMPLINGS AND VALUES DERIVED FROM THE POLYMORPHIC SITES ARE INDICATED IN PARENTHESES. (a) ANALYSIS OF ALL POSSIBLE SITES (1371 NUCLEOTIDES) AND (b) ANALYSIS OF POLYMORPHIC SITES (528 NUCLEOTIDES). MAJOR DIFFERENCES ARE INDICATED IN BOLD TYPE.

Bifidobacterium bifidum and "*Streptomyces ambofaciens*". It is, therefore, clear that the bootstrap values derived from the different tree topologies are not the same. The bootstrap values that were significantly reduced are indicated in bold type (Figure 1-2a). It is evident that in two cases the bootstrap values were reduced from over 95% to below 90%. This misuse of bootstrap evaluation may be due to the confusion seen between the parsimony and distance matrix methods whereby the former method gives identical results irrespective of whether all or just the polymorphic sites are used but the latter method does not.

Over one hundred methods are available to estimate phylogenetic relationships but it is not always clear what their strengths and weaknesses are or which method should be used in a given situation (Swofford & Olsen, 1990; Hills *et al.*, 1993). Simulation studies can be used to address such questions but the final choice has to be made by molecular systematists. However, it is clear that bacterial phylogenies should be interpreted with care as all estimates of phylogeny are based on relatively simple assumptions when compared with the complexity of the natural evolutionary process. It has been shown that the assumption of symmetry of nucleotide substitutions is violated for actual sequence data (Gojobori *et al.*, 1982; Li *et al.*, 1984). It has also been shown that compensatory mutations in stem regions of ribosomal DNA may violate the assumption of character independence (Wheeler & Honeycutt, 1988; Dixon & Hillis, 1993). Those engaged in generating phylogenies should remember the words of Francis Lord Bacon (1605), as quoted by Sneath (1986):

> *"If a man will begin with certainties he shall end in doubts; but if he will be constant to begin with doubts he shall end in certainties."*

2. Computer-assisted classification and identification based on phenotypic/phenetic properties

2.1.Computer-assisted classification

Classification is the process of ordering organisms into groups (taxa) on the basis of their relationships (not confined to relationships by ancestry). The product is an orderly arrangement or system of classification designed to express interrelationships of organisms and to serve as an information storage and retrieval system. Early bacterial classifications relied heavily on a few morphological and staining properties that were sometimes supplemented by a small number of physiological features. Taxonomies based on single characters, or a series of single characters, are termed monothetic classifications (Sneath, 1962). Such artificial classifications are usually very unreliable as they have a low information content and cannot accommodate strain variation or test error.

The structural weaknesses of monothetic classifications led some bacterial systematists to believe that stable taxonomies could only be achieved when classifications were derived from the analysis of large numbers of bacteria for many properties. Numerical taxonomies have a high information content and are often described as general purpose classifications *sensu* Gilmour (1937) since they are of potential value to many bacteriologists.

A reliable way of establishing centers of variation in poorly circumscribed genera is to examine many strains for large numbers of equally weighted characters. This is the foundation of the numerical taxonomic procedure introduced to bacteriology by P. H. A. Sneath (1957a,b) and subsequently widely applied (Sneath & Sokal, 1973; Goodfellow & Dickinson, 1985; McDonnell & Colwell, 1985; Sackin & Jones, 1993). In essence, numerical classifications are formed when large numbers of strains are examined for many characters then classified on the basis of

overall similarity (Figure 1-3). Classifications generated in this way are polythetic as they have a high information content and are based on a complete set of recorded characters not on the presence or absence of single subjectively weighted features. Numerical classifications can generally accommodate strain variation and are objective in the sense that they are not overtly sensitive to the addition of more strains or characters. The relationships between the test strains and any hierarchies based on them are phenetic not phylogenetic (Goodfellow & O'Donnell, 1993).

The strength of the numerical taxonomic procedure relies on the ability of computers to handle large amounts of data on large numbers of strains. Thus, the taxonomic groups produced should be, and generally are, stable because they are based on large amounts of information with no single character being either essential for membership or sufficient to exclude a strain from a group (Jones & Sackin, 1980). Theoretically, the more phenetic characters that are examined the better will be the measure of the expressed, and hence probably the genomic, relatedness between test strains.

Choice of strains. The objects to be classified, usually strains, are called operational taxonomic units (OTU; Sneath & Sokal, 1973). Where possible, strain collections should include type strains, representatives of additional well studied cultures, marker strains outside the area of study and duplicated cultures to provide an internal check on test reproducibility. It is also advisable to include newly isolated strains since organisms which have been repeatedly subcultured may not be good representatives of established taxa (Logan, 1994; Goodfellow, 1995a). At least ten percent of strains should be duplicated and treated as separate operational taxonomic units. In theory, about twenty-five strains are needed to accurately define the centre and radius of a numerically defined species (Sneath, 1977).

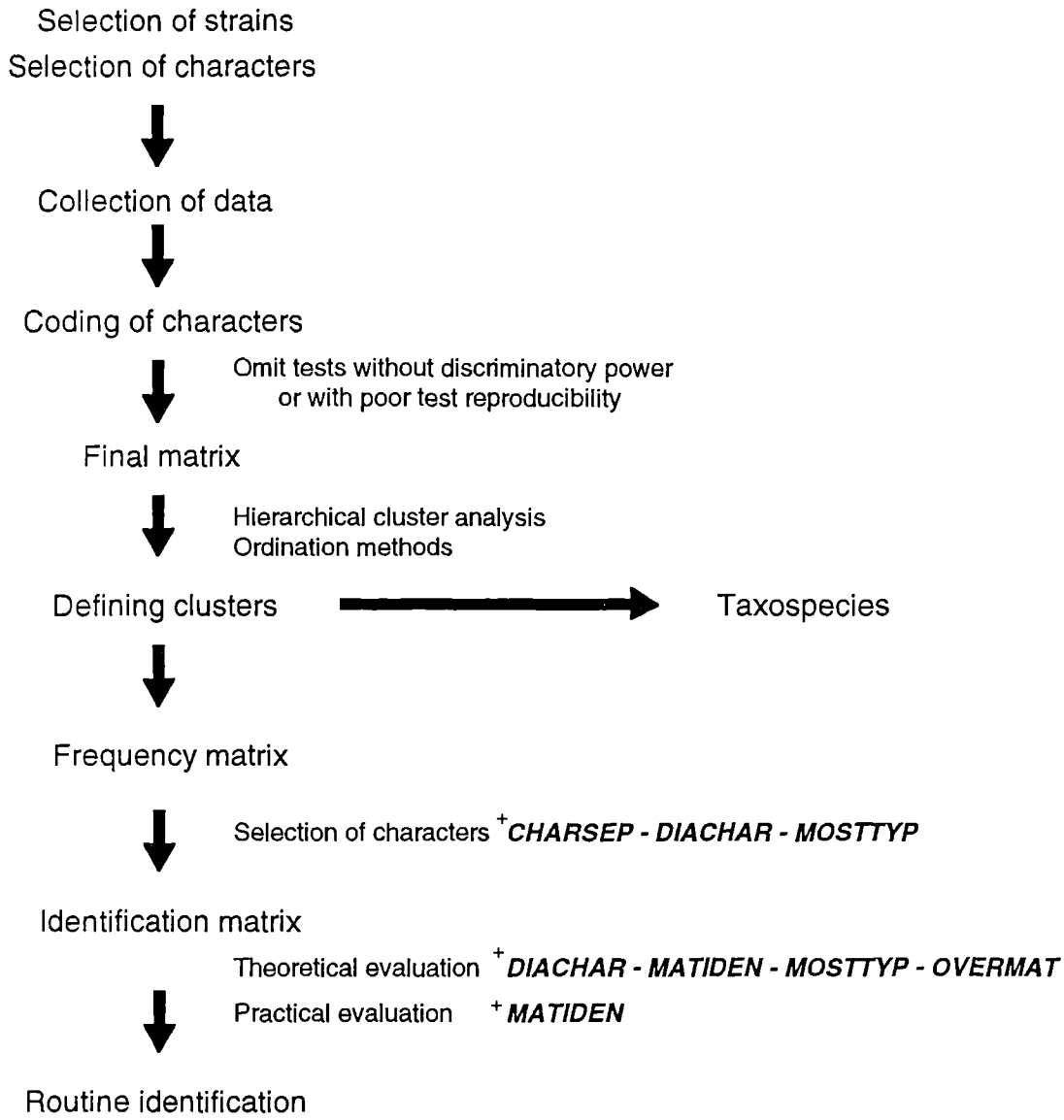


FIGURE 1-3. MAJOR STEPS INVOLVED IN NUMERICAL CLASSIFICATION AND IDENTIFICATION. +, PROGRAMS WRITTEN BY SNEATH (1979d, e, 1980a, b, c).

Choice of characters. The characters used in numerical taxonomy are based on results of tests performed on test organisms. It is important to use characters that are genetically stable and not overtly sensitive to experimental or observational uncertainties. The usual procedure is to take a selection of biochemical, cultural, morphological, nutritional and physiological characters to represent the entire phenome, that is, the genotype and the phenotype. It is also important to have sufficient information to discriminate between taxa. Sackin and Jones (1993) recommended at least fifty unit characters but preferably several hundred tests are needed though with high numbers of features any gain in information falls off disproportionally to the effort involved in securing data. In addition, tests need to be carefully chosen and performed under rigorously standardised conditions (O'Brien & Colwell, 1987; Logan, 1994; Goodfellow, 1995a).

Coding of data. Data must be coded into a format suitable for computation. The results obtained from various tests are usually coded as binary form. In general, unit characters are coded as two state characters where the possession of the character is scored as plus (+) or 1 and its absence as minus (-) or 0. Multistate characters, such as colonial colour, are converted into two states by using the mutually exclusive method of coding (Sneath & Sokal, 1973) where an OTU possessing a particular state for a property is coded positive (1) for that character state and negative (0) for all of the remaining character states. A disadvantage of this method of coding is that the weight given to each complex character increases proportionately with the number of character states. It is for this reason that the number of mutually exclusive characters are usually kept to a minimum in numerical taxonomic studies (Goodfellow *et al.*, 1990; White *et al.*, 1993).

Estimation of test error. Determination of test reproducibility is essential given the effect of test error on the structure of numerical classifications (Sneath, 1972,

1974b; Sneath & Johnson, 1972; Goodfellow, 1977; Sackin & Jones, 1993). By examining duplicated cultures, initially under code, and treating them as individual OTUs, the mean probability of test error can be estimated from an analysis of test variance. The reproducibility of a two-state test *i*, can be estimated using the following equation:

$$p_i = \frac{1}{2}\left(1 - \sqrt{1 - 4s_i^2}\right)$$ (Sneath & Johnson, 1972; equation 4)

where p_i is the estimated error rate of the test and s_i^2 is its estimated variance which is given by:

$$s_i^2 = \frac{n}{2t}$$ (Sneath & Johnson, 1972; equation 15)

where *n* is the number of OTUs with discrepancies and *t* is the total number of duplicated strains. Individual test variances may be averaged to give the pooled variance (S^2):

$$S^2 = \frac{1}{N}\sum_{j=1}^{N} S_j^2$$

where *N* is the total number of tests.

Test error has the general effect of lowering similarities between strains and when high of eroding taxonomic structure. Sneath and Johnson (1972) found that there is a rapid erosion of taxonomic structure when $p > 0.1$. They also pointed out that it is generally better to employ many tests even if they are not as reproducible as desired rather than to use only a few extremely reproducible tests. Tests of historical importance can be included in final data matrices even though they are not reproducible (Barrett & Sneath, 1994). It is important to balance the detrimental effect on taxonomic structure of individual tests with a high error against information loss of test results deleted from data matrices.

Values for test error range from less than four percent for within laboratory studies (McCarthy & Cross, 1984; Goodfellow *et al.*, 1991; Whitham *et al.*, 1993; Barrett & Sneath, 1994) to as high as fifteen percent for results between laboratories (Sneath, 1974b). Clearly, standardised methods are required to generate inter-laboratory databases for diagnostic purposes (Sneath, 1974b).

Computation of resemblance. A number of ways of measuring similarity and dissimilarity between a pair of OTUs have been devised (Sneath & Sokal, 1973; Austin & Colwell, 1977) but only those commonly used in bacterial taxonomy are considered here.

The simple matching coefficient (S_{SM}), which is only valid with binary data, is simply the proportion of characters that two organisms have in common. The symbols *a* and *d* are used to account for the number of shared positive and negative matches with *b* and *c* accounting for the number of differences between a pair of OTUs:

		Results for OTU1	
		+	-
Results for OTU2	+	*a*	*b*
	-	*c*	*d*

The S_{SM} coefficient is calculated as:

$$S_{SM} = \frac{a+d}{a+b+c+d}$$

This coefficient has the great virtue of simplicity and is equal to $1-d^2$ where *d* is the taxonomic distance. Principal coordinate analysis can be performed on a matrix based on S_{SM} coefficients so that the resulting values are taxonomic distances between OTUs (Sackin & Jones, 1993). The S_{SM} coefficient also satisfies the

conditions for the statistical examination of cluster overlap using the method described by Sneath (1979b).

Similarly, Jaccard's coefficient (S_J) may be defined as:

$$S_J = \frac{a}{a+b+c}$$

that is, the ratio of the total number of positive matches to the total number of characters minus the sum of the negative matches. The S_J coefficient is often applied to ensure that relationships detected using the S_{SM} coefficient are not based on negative correlations. Negative matches are not necessarily a measure of similarity as some strains may fail to give a positive response whereas others may simply be unable to do so under the test conditions. The S_J coefficient is particularly useful in studies where relatively fast- and slow-growing organisms are compared (Goodfellow & Wayne, 1982; Ridell & Goodfellow, 1983; Whitham *et al.*, 1993).

Gower's similarity coefficient (S_G) is a weighted average of all similarity values between pairs of OTUs and can be used to analyse binary, quantitative and disordered multistate data. For binary data, the S_G coefficient is equivalent to the S_J coefficient whereas for disordered multistate characters it is identical to the S_{SM} coefficient. In the case of quantitative data, the S_G coefficient is defined as:

$$S_G = 1 - \frac{v1 - v2}{r}$$

where *r* is the range of values for the character and *v1* and *v2* the values of OTU1 and OTU2, respectively.

Sneath (1968) considered that the total difference between OTUs was a reflection of two components, namely pattern and vigour differences. The latter may be equated with shape and size differences, respectively, in animals and plants. The aim of separating these two components was to reduce the apparent

dissimilarity between OTUs that could be attributed to differences in growth rate, that is, the vigour difference. Thus, slow-growing strains may be scored as negative in a test, such as acid production from sugars, not because they lack the necessary catabolic pathway but because they do not grow fast enough for a positive reaction to be recorded within a given period of time. The pattern coefficient (D_P) is useful when strains of widely different metabolic activity are being studied.

The vigour of an OTU is defined as the proportion of its characters that show a positive response in the tests. The vigour difference, D_V, is the difference in vigour between two OTUs:

$$D_V = \frac{c-b}{a+b+c+d} \text{ (Sneath, 1968)}$$

using the notation described earlier. Similarly, the pattern difference, D_P, is defined as:

$$D_P^2 = D_T^2 - D_V^2 \text{ (Sneath, 1968)}$$

where D_T is the total difference. The latter can be defined either as $\sqrt{1-S_{SM}}$ or $\sqrt{1-S_J}$. Applying the former definition, the pattern difference (D_P) can be derived as:

$$D_P = \frac{2\sqrt{bc}}{a+b+c+d}$$

The D_P and D_V coefficients are dissimilarity coefficients which range from 0 to 1. The D_P coefficient has been successfully applied in numerical phenetic studies of bacteria that grow slowly, particularly when comparisons are to be made with more vigorous organisms (e.g., Goodfellow *et al.*, 1976; Whitham *et al.*, 1993). However, this coefficient tends to be unreliable when OTUs show a disproportionate number of positive and negative matches (Goodfellow, 1995a).

Sackin (1981) derived an alternative form of the pattern difference coefficient (D_{P2}) from the S_G coefficient. The D_{P2} coefficient is defined as:

$$D_{P2}^2 = D_{T2}^2 - D_{V2}^2 \text{ (Sackin \& Jones, 1993)}$$

where $D_{T2}^2 = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i|$, $D_{V2}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)$, x_i and y_i are the *i*th character states, transformed into the range 0 to 1, in the two OTUs under comparison, and *n* is the total number of characters.

The Euclidean distance coefficient, *D,* is useful for analyses of quantitative data. It is defined as:

$$D^2 = \sum(x_i - y_i)^2$$

where x_i and y_i are the values of the *i*th character for the two OTUs. The taxonomic distance, often referred to as *d,* is the Euclidean distance divided by the square root of the number of characters and hence ranges from 0 (OTUs identical) to 1 (maximum dissimilarity). Euclidean distances satisfy the conditions for many forms of statistical analysis such as overlap statistics (Sackin & Jones, 1993).

Character weighting. Numerical taxonomists are generally in agreement in giving all unit characters equal weight when creating taxonomic groups (Sneath & Sokal, 1973). However, there are few cases in which unequal weighting can be considered. A way to weight characters according to reproducibility has been developed (Sackin & Jones, 1993). The weight of the *i*th character is defined as:

$$w_i = (1 - 2p_i)^2 \text{ (Sackin \& Jones, 1993)}$$

where p_i is the estimated error rate. An advantage of this procedure is that it is not necessary to choose an arbitrary cut-off point for estimated test error.

Milligan (1989) recommended the weighting scheme of De Soete (1986, 1988) when Euclidean distance is applied to hierarchical clustering. This method attempts to yield distances:

$$d = \sqrt{w_i \sum (x_i - y_i)^2}$$

that approximate to ultrametric inequality:

$$d_{xy} \leq \max(d_{xz}, d_{yz})$$

where w_i is the weight for the ith character, x_i and y_i are the values of ith character for OTUs x and y, and d_{xy}, d_{xz} and d_{yz} are the distances between OTUs x, y and z. Milligan (1989) found that this method gave the best recovery of hierarchical structure in a simulated dataset as it downweighted characters that are noisy in relation to the hierarchy.

In practice, it is not necessary to weight unit characters in analyses which are based on a large amount of binary data. It is also unlikely that a few correlated features, or noisy data, will affect the taxonomic structure derived from analyses of a hundred or more unit characters (Sackin & Jones, 1993).

Hierarchical clustering. The ordering of OTUs into groups of high overall phenetic similarity is usually achieved by means of one of several commonly used sequential, agglomerative, hierarchic, non-overlapping (SAHN *sensu* Sneath and Sokal 1973) clustering methods. In general, these begin by searching similarity matrices for the highest values between any pair of OTUs. This pair forms a group or cluster. The similarities between this group and each of the remaining OTUs are considered for the next agglomeration. This process proceeds so that at each cycle OTUs are added to clusters or clusters join until all of the OTUs are included in a single cluster. Consequently, *t-1* cycles are required to cluster *t* OTUs.

Clustering techniques vary in terms of the definition of the similarity between an OTU and a group and, more generally, between groups. The single linkage (SL; Sneath, 1957a) technique, which is also known as the nearest neighbor method, defines the similarity between two groups as the similarity of the two most similar OTUs, one in each group. In contrast, the average linkage (AL; Sokal & Michener, 1958) technique takes the average of all of the similarities across the groups. The most popular variant of the AL method, the unweighted pair group method with arithmetic averages (UPGMA; Sokal & Michener, 1958), takes the simple arithmetic average of the similarities across two groups, each similarity having equal weight. The weighted pair group method with arithmetic averages (WPGMA; Sneath & Sokal, 1973) differs from the UPGMA algorithm by weighting the OTU most recently admitted to a cluster equal with all previous members. This algorithm shares the properties of the UPGMA method but distorts overall taxonomic relationships in favor of the most recent addition to a cluster.

The similarity levels at which clusters are defined are influenced by the clustering algorithms (Goodfellow, 1995a). In general, clusters formed using the AL algorithm are more compact than those based on the SL method (Sackin & Jones, 1993). The strengths and weaknesses of SAHN methods have been considered in detail (Sneath & Sokal, 1973; Everitt, 1980).

Hierarchical clustering methods impose a structure on data which may or may not be a true representation of relationships between OTUs as shown by their observed similarity values. Intuitively, data are suitably hierarchical if a dendrogram can be constructed in which most or all of the original resemblance scores between the OTUs have values close to the corresponding scores, that is, the cophenetic values, derived from the dendrogram alone. This is the basis of the cophenetic

correlation coefficient (*r*; Sokal & Rolf, 1962), which is the most common measure of hierarchicalness of the SAHN clustering methods.

Typical cophenetic correlation values are in the range of 0.6 to 0.95 (Sackin & Jones, 1993; Goodfellow, 1995a). Values at or above 0.8 are usually considered to be good whereas those below 0.7 suggest that only limited confidence can be placed in relationships presented in dendrograms. In practice, complete agreement between dendrograms and resemblance matrices cannot be achieved given the taxonomic distortion introduced when representing multidimensional data in two dimensions. The UPGMA algorithm tends to give higher cophenetic correlation values than other SAHN clustering methods. It is for this reason that this algorithm is usually the method of choice in numerical phenetic surveys (Jones & Sackin, 1980; Sackin & Jones, 1993; Goodfellow, 1995a).

The results of SAHN clustering are usually presented in the form of a dendrogram where the tips of the branches represent OTUs and the axis at right angles to the tips is the similarity axis which shows the similarity values at which groups form. The results can also be used to produce ordered similarity matrices where squares are shaded according to their similarity values so that the highest values receive the darkest shading. The shaded diagram is a useful vehicle for highlighting major groups and subgroups within a set of OTUs. It also gives a visible diagrammatic representation of intra- and inter-group similarities (Goodfellow, 1977; Sneath, 1978; Jones & Sackin, 1980).

Ordination methods. Ordination is the placement of *t* OTUs in an *A*-space of dimensionality that varies from one to the number of unit characters or *t*-1 whichever is less (Sneath & Sokal, 1973). With ordination methods, it is possible to view relationships between OTUs directly in terms of taxonomic space. In contrast to hierarchical clustering methods, OTUs are not divided into convenient groups using

ordination methods although groups may be recognised by careful examination of output. Two ordination methods, namely principal component analysis (PCA) and principal coordinate analysis (PCO), are commonly used in bacterial taxonomy. Principal component analysis (PCA) is a well known means of reducing dimensionality from multivariate data. The guiding concept of PCA is to reduce the multidimensional aspect of the distribution of OTUs in the *A*-space to just two or three dimensions so that the positions of OTUs can be visualised. The theoretical background of PCA is well documented (Dunn & Everitt, 1982; Manly, 1986).

Principal coordinate analysis, an alternative method to PCA, was developed by Gower (1966). In contrast to PCA, which uses the raw data matrix ($n \times t$), the starting point for PCO is the dissimilarity matrix. Further unlike PCA, which is only relevant when an Euclidean metric is considered, PCO is an ordination which is applicable to relationships between a set of OTUs in taxonomic space irrespective of whether distances are Euclidean or not (Alderson, 1985). However, when the observed proximities are Euclidean the results of PCO are equivalent to those of PCA (Gower, 1966; Everitt, 1980). Principal coordinate analysis is especially useful when only dissimilarity matrices are available (e.g., DNA-DNA pairing data). The method has been successfully applied to bacterial classification (Logan & Berkeley, 1981; Bridge & Sneath, 1983; Alderson *et al.*, 1984; Barrett & Sneath, 1994).

Ordination techniques have been successfully used to represent relationships between large groups but such analyses can distort affinities between close neighbors (Alderson, 1985). In contrast, hierarchical clustering methods are reliable when depicting relationships between closely related, large heterogeneous groups (Sneath & Sokal, 1973; Sneath, 1978, Sackin & Jones, 1993). Results from ordination studies are generally represented as two or three dimensional plots. Cluster overlap, due to reduction of dimensions, can be a limitation in the

interpretation of results of ordination analyses as clusters that are distant in full hyperspace may overlap in low dimensional plots (Sneath & Sokal, 1973; Clifford & Stephensen, 1975; Sneath, 1980c; Logan, 1994).

2.2. Computer-assisted identification

Identification is the end-product of the taxonomic process and as such is clearly dependent on the accuracy and data content of classification systems and on the predictive value of names assigned to taxa. In general, the characters chosen for identification schemes should be easy to perform and few in number. Numerical taxonomic surveys provide data on test reactions of strains within each taxon circumscribed in the classification. Results are usually expressed as frequency matrices which consist of the percentage of strains in each cluster that give a positive result for the unit characters. Diagnostic characters can then be selected from percentage positive frequency tables, that is, by *a posteriori* weighting, and used to generate dichotomous keys, diagnostic tables and computer identification matrices. Computer-assisted identification is preferred to conventional keys and tables as it is relatively quick and simple (Lapage *et al.*, 1973; Hill, 1974; Priest & Williams, 1993).

Beers and Lockhart (1962) were first to suggest that bacterial identification could be based on a mathematical model. The theoretical studies that followed culminated in the introduction of a computer-based probabilistic system for the identification of enterobacteria (Lapage *et al.*, 1970, 1973; Bascomb *et al.*, 1973; Willcox *et al.*, 1973). Since then, probabilistic identification schemes based on numerical classifications have been introduced for certain actinomycetes, including slowly-growing mycobacteria (Wayne *et al.*, 1980, 1984), neutrophilic streptomycetes (Williams *et al.*, 1983b; Langham *et al.*, 1989a; Kämpfer &

Kroppenstedt, 1991), streptosporangias (Kim, 1993; Whitham *et al.*, 1993) and streptoverticillias (Locci *et al.*, 1986).

The first step in the generation of an identification matrix is the selection of a small number of diagnostic tests that are sufficient to differentiate all of the taxa in the numerical taxonomic database. Programs available for this purpose include ***CHARSEP*** (Sneath, 1979e) and ***DIACHAR*** (Sneath, 1980a). The ***CHARSEP*** program is used to find the value of different unit characters in separating groups where the percentage of positive values are known. Five separation indices can be calculated for each unit character, namely Gyllenberg's sum of C, Gyllenberg's R, Niemelä's index, the separation potential (VSP) index and the character separation potential (CSP) index. These indices can be used to determine the value of each unit character in turn in separating clusters in a frequency matrix. The ***DIACHAR*** program (Sneath, 1980a) is used to calculate the diagnostic scores of each test for each cluster given the frequency matrix; the tests are then ranked in order of descending scores. An identification matrix should contain sufficient information to define each taxon by several diagnostic properties.

The importance of evaluating frequency matrices has been repeatedly stressed (Sneath, 1978; Williams *et al.*, 1983b; Priest & Williams, 1993). The computer program ***OVERMAT*** (Sneath, 1980c) can be used to determine the degree of overlap between clusters represented in frequency or identification matrices. Unknown strains cannot be unambiguously identified when there is considerable overlap between taxa. The ***OVERMAT*** program determines both the disjunction index (W) for each pair of taxa and the corresponding nominal overlap (V_G) from the percentage positive data. Additional tests selected by using the ***DIACHAR*** program can be added to identification matrices in order to reduce cluster overlap.

A homogeneous taxon will contain strains that fall into a well defined taxonomic space. Strains which do not fall within the boundary of a cluster are sometimes referred to as **'outliers'**. These organisms presumably represent atypical strains or centers of variation that represent new taxospecies. The reliability of identification matrices may be impaired if outliers are included in clusters. Sneath and Langham (1989) developed a program for detecting outliers. This program, ***OUTLIER***, can be used to estimate the distances of strains from the centre of a taxon and it draws a $^{2}\chi$ distribution in the resultant graph. The Kolmogorov-Smirnov test is used to assess the fitness of data to the $^{2}\chi$ distribution and to list strains that lie outside the taxon.

Sneath (1980b) introduced the computer program ***MOSTTYP*** to calculate identification scores for the most typical organisms, that is, the hypothetical median organism (HMO), in each of the clusters included in an identification matrix. When identification matrices are sound the HMO of each cluster will be identified to its taxon with high identification scores. Probabilistic identification matrices can be further assessed by treating strains included in the original numerical taxonomic study as known organisms then calculating identification scores using the original classification data obtained for the diagnostic tests.

Practical evaluations of frequency matrices are also important as they allow an assessment of test error (Priest & Williams, 1993). The usual approach to practical evaluation is to first identify reference strains then field isolates. In general, reference strains are identified with high scores to the correct taxon (Bascomb *et al.*, 1973; Williams *et al.*, 1983b; Locci *et al.*, 1986; Priest & Alexander, 1988; Whitham *et al.*, 1993). The second step involves identification of fresh isolates (Priest & Alexander, 1988, Kim, 1993; Whitham *et al.*, 1993).

Two types of computer-assisted identification schemes can be recognised (Sneath, 1979a). The Bayesian model, is based on the concept that the probability of an unknown strain belonging to a specific taxon is a function of the probability that an individual chosen at random possesses a given series of character states. A number of investigators have developed this model (Bascomb *et al.*, 1973; Lapage *et al.*, 1973; Willcox *et al.*, 1973). This method involves the calculation of Willcox probabilities, that is, the likelihood of unknown character-state values against a particular group divided by the sum of the likelihoods against all of the other groups. Scores approaching 1.0 indicate a good fit between unknown organisms and a group in the identification matrix. Most commercial identification systems based on phenotypic properties use Willcox probabilities (Priest & Williams, 1993).

The second approach to numerical phenetic identification is based on the taxon-radius model, where a taxon is represented as a hypersphere in a space of many dimensions, one dimension per unit character (Sneath, 1979a). This approach allows the taxon to be circumscribed by a critical radius drawn about the centre. The centre represents the most typical individual and the circumference the limits of possible variation. This method is only applicable when it is known that clusters are close to being hyperspherical (Sneath, 1979a). Identifications are achieved by calculating the distance between unknown OTUs and cluster centroids.

The two approaches to computer-assisted identification outlined above have been implemented in the BASIC program developed by Sneath (program MATIDEN, 1979d); the computational details are given in the Materials and Methods section of this chapter.

3. Software tools

Bacterial taxonomists have been engaged in two difficult tasks in recent years, namely the generation of taxonomic information and the development of

suitable data handling techniques. The first objective is increasingly being realised due to the development of automated systems for the acquisition of phenotypic (Logan & Berkeley, 1981; Garland & Mills, 1991; Kämpfer *et al.*, 1991, 1992; Hamid *et al.*, 1994), nucleic acid sequence (Xia *et al.*, 1994; Rainey *et al.*, 1995a) and chemotaxonomic data (Stead, 1992; Vainshtein *et al.*, 1992; Haack *et al.*, 1994). However, less attention has been devoted to the development of adequate software tools.

In the late 1970's and 1980's, P. H. A. Sneath wrote several computer *programs for the analysis of phenotypic data. Software packages* currently recommended for the classification of bacteria have been reviewed by Sackin and Jones (1993). The ***NTSys-pc*** (James Rohlf; Exeter Software, 100 North Country Road, Building B, Setauket, New York 11733, USA) includes routines for most cluster analysis and multivariate statistical methods, such as principal component and principal coordinate analyses. The classical ***CLUSTAN*** package (Wishart, 1987) also contains several procedures relevant for bacterial classification.

Probabilistic identification of microorganisms can be achieved by using different software packages, such as ***Bacterial Identifier*** (Bryant, 1991), ***MICRO-IS*** (Bello, 1989) and ***TAXON*** (A. C. Ward, Department of Microbiology, University of Newcastle upon Tyne, England, UK; unpublished). The development of software for bacterial identification has been extensively reviewed (Sackin, 1987; Sackin & Jones, 1993). In addition to these programs, most commercially available phenotypic identification systems, such as BIOLOG (BIOLOG, 1992), implement the identification routine in their packages.

Nucleotide sequence data, unlike the phenotypic data, can be used to construct a universal database. Such data are currently managed by three different gene databases, namely DDBJ (DNA Databank of Japan), EMBL (European

Molecular Biology Laboratory) and GenBank (USA). The ribosomal database project (RDP; Larsen *et al.*, 1993), headed by Carl Woese, provides aligned 16S and 23S rRNA sequence data with software packages for alignment and phylogenetic analysis. A detailed review of the databanks mentioned in the global international network has been given by Canhos *et al.* (1993).

It is very interesting that the use of different types of data handling routines is largely correlated with the availability of the corresponding software tools. The use of fatty acid composition data, for example, was enhanced by the introduction of automated systems, notably the MIDI identification system (Microbial ID, Inc., Newark, Delaware. USA). This system, which is especially useful in clinical diagnostic laboratories, includes an automated data handling system suitable for non-taxonomists. Similarly, the application of whole-organism protein fingerprinting and Curie point pyrolysis mass spectrometry is increasing due to the availability of suitable software tools (Magee, 1993, 1994; Pot *et al.*, 1994). The availability of the ***PHYLIP*** package (Felsenstein, 1993) has provided a platform for a variety of phylogenetic analyses of amino acid and nucleotide sequence data. It is not surprising that the value of many statistical and numerical methods introduced for the analysis of various types of taxonomic data have not been appreciated given the lack of easy access to the necessary software.

In the present study, two programs specifically designed to handle taxonomic data were written for personal computers. The first one, the ***AL16S*** program, was developed for the management and comparative analysis of 16S rRNA sequence data and the second, the ***X*** program, for handling phenotypic data.

<BLANK>

B. Materials and Methods

1. Computers and compilers

The software developed in the present study was written in the C++ language (Stroustrup, 1991) and complied by using Borland C++ version 3 (Borland International, 1800 Green Hills Road, Scotts Valley, CA 95067-0001, USA). The operating system was Microsoft-DOS version 6.0 (Microsoft Corporation, One Microsoft Way, Redmond, WA 98052-6399, USA). The software was performed on IBC-PC compatible personal computers.

2. The AL16S program

2.1. Description

The ***AL16S*** program is integrated software that allows editing, alignment and comparative analysis of 16S rRNA sequence data based on the sequence format used in the ribosomal database project (RDP version 4.0; Larsen *et al.*, 1993). This format comprises 2846 nucleotide positions per sequence which allows alignment of almost all available small subunit rRNA sequences (more than 2000 sequences). The numbering of the nucleotide positions can be chosen by the user; *Escherichia coli* numbering system (Appendix C) is set as default. For convenience, thymine (T) is used for uracil (U). The use of degenerate bases, according to IUB codes (Appendix B), are also possible with most analyses included in the ***AL16S*** program.

The ***AL16S*** program can be used to handle up to 150 16S rRNA sequences at a time. However, there is no limit on the number of nucleotide sequences that can be considered at any one time using the neighbor-joining method. Some analyses are based on groups which correspond to clusters in numerical phenetic classifications, the composition of which can be defined by the user. The management of the 16S RNA database is carried out by using an accessory

database program, namely the ***SSUrRNA*** program which is written in the FOXPRO database language (Microsoft Corporation, One Microsoft Way, Redmond, WA 98052-6399, USA). The main steps in the analysis of 16S rRNA sequence data are given in Figure 1-4.

2.2. Main functions

Importing sequence data. 16S rRNA sequences in a variety of file formats can be imported to the program for alignment. These include the GenBank/EMBL formats and text formats from automatic sequencers.

Alignment. Individual nucleotide positions of bacterial 16S rRNA are usually indicated by using the *Escherichia coli* 16S rRNA gene of the *rrnB* cistron (accession number J01695; Brosius *et al.*, 1978; Appendix C). The numbering systems from other bacterial strains can be also applied, if necessary. The helices found in 16S rRNA molecules of Archaea and Bacteria can be named after the nomenclature system of Neefs *et al.* (1993; Appendix D).

The alignment of 16S rRNA sequences is achieved manually by using the secondary structural information incorporated into the program. The secondary structural information from *Arthrobacter globiformis* (accession number M23411), *Bacillus subtilis* (K00673), *Escherichia coli* (J01695) and "*Streptomyces coelicolor*" strain A3(2) [Y00411] obtained from the RDP (Larsen *et al.*, 1993) was implemented in the program. The program is used to display information on helical structures and in tertiary structures in 16S rRNA molecules (e.g., positions 506-507 are complementary to positions 524-526 of the *Escherichia coli* numbering system, see Appendix D).

Masking/selecting positions. The appropriate nucleotide positions should be selected by the user prior to any kind of sequence analysis, especially for phylogenetic inference analyses. It is important to select positions that have been

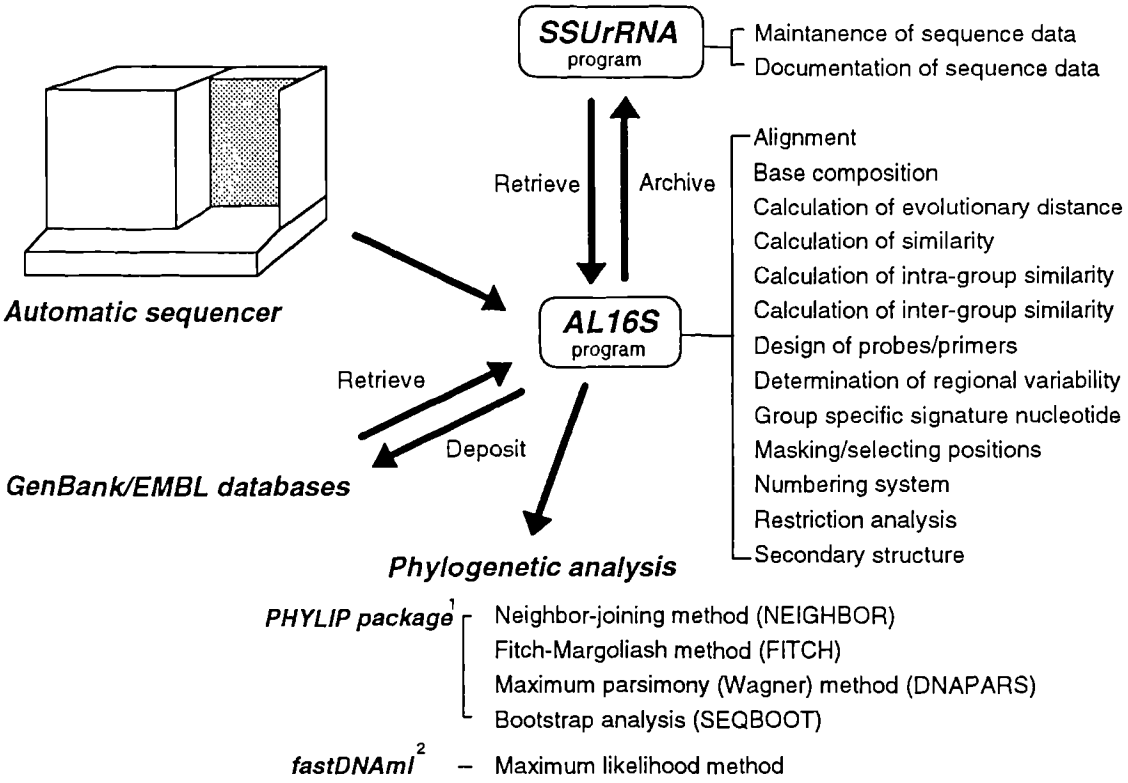


FIGURE 1-4. MAJOR STEPS AND SOFTWARE TOOLS USED FOR ANALYSIS OF 16S RRNA SEQUENCE DATA. [1], ***PHYLIP*** PACKAGE (FELSENSTEIN, 1993) AND [2], ***FASTDNAML*** (OLSEN *et al.*, 1994b).

aligned with confidence. When distantly related sequences are compared many of the aligned positions can be masked and hence excluded from subsequent analyses. However, all of the positions for sequence analysis can be included in comparisons of closely related organisms.

Calculation of sequence similarity. The program can be used to calculate the pairwise sequence similarity from the selected positions (module '***Similarity***'). The mean of the pairwise sequence similarities between members of the same group can be determined by using the module '***Intra-group similarity***'. The results generated from this module provide a means of assessing the sequence variability in different phyletic lines. In addition, the program contains the module '***Inter-group similarity***' that calculates the mean similarities between members of different groups.

Calculation of evolutionary distances. The program can be used to compute two evolutionary distances, namely the Jukes and Cantor (Jukes & Cantor, 1969) and Olsen models (Swofford & Olsen, 1990) as incorporated in the module '***Distance***'. The output of this module can be used for tree-making methods implemented in the ***PHYLIP*** package (Felsenstein, 1993).

Signature nucleotides. Some nucleotide positions in 16S rRNA can be used to classify bacteria at different taxonomic ranks (Woese, 1987). The module '***Group signature***' can be used to produce the consensus nucleotide sequence from a group of organisms and hence can be applied to find group specific nucleotides, that is, signature nucleotides. The consensus base is presented as a degenerated base according to IUB codes (Appendix B) when different bases occur in a given nucleotide position. The output of this module is equivalent to the frequency matrix (percent positive table) of numerical phenetic taxonomy.

Regional variability. The degree of conservation in different parts of 16S rRNA molecules varies in different phylogenetic lineages (Woese, 1987; Stackebrandt & Goebel, 1994). The degree of sequence variability of specific parts of the 16S rRNA molecule within a group can be determined by using the module '***Regional variability***'. This module is used to calculate the mean value of pairwise similarities within a certain region of 16S rRNA (e.g., between positions 401-500) from a selected group.

Design of specific primers and probes. One of the useful end-products of 16S rRNA sequence analyses is that specific group primers and oligonucleotide probes are highlighted (Woese, 1987; Amman *et al.*, 1994, 1995). The program ***AL16S*** can be used to check oligonucleotide sequences against 16S rRNA databases. Since the program can be used to detect the secondary structure of the priming region, the user can consider the interference made by intramolecular pairing.

Restriction analysis (module restriction analysis). The sizes of DNA fragments derived from restriction analyses of PCR-amplified 16S rRNA genes vary among organisms if appropriate restriction endonucleases are used. The user can append or edit restriction sites. This module, which produces lists of restriction positions in amplified 16S rDNA and displays simulated agarose gel images, can be used to detect suitable endonucleases for the separation of members of different species.

Listing of raw nucleotide sequences. The module '***Profile***' can be used to produce lists of aligned sequences. The output file is useful for visual inspection of nucleotide sequences and for the designation of oligonucleotide probes.

3. The X program

3.1. Description

This user-friendly, menu-driven software is suitable for the analysis of phenotypic data for numerical taxonomic purposes. The software is designed both to provide an integrated platform for the various numerical analyses that are involved in the classification and identification of bacteria and to serve as an automated interface with other software packages.

The program can hold binary data for up to 500 strains, 500 tests, 200 clusters (groups of strains) and 200 testsets (groups of unit characters). These limits can easily be extended by recompiling the original source codes. The program only contains modules specific for bacterial systematics as several commercial software packages are available for multivariate statistical analyses. The steps and corresponding software tools are shown in Figure 1-5.

3.2. Main Functions

Generation of frequency matrix. The module '***Percentage positive table***' can be used to produce percentage positive tables (frequency matrices) that are the basis for most statistical analyses. The module '***Plus table***' can be used to list the number of positive results if percentages are not required.

Estimating test reproducibility. The module '***Test error***' can be used to calculate the individual reproducibility of tests. Both the variances (s_i^2) and estimated error rates (p_i) are computed according to Sneath and Johnson (1972). The overall reproducibility is determined from the pooled variance.

Centre of clusters. The centre of a cluster can be represented in two different ways, namely as a point representing an actual organism in taxonomic space or as a point representing a hypothetical organism (Sneath & Sokal, 1973). The former,

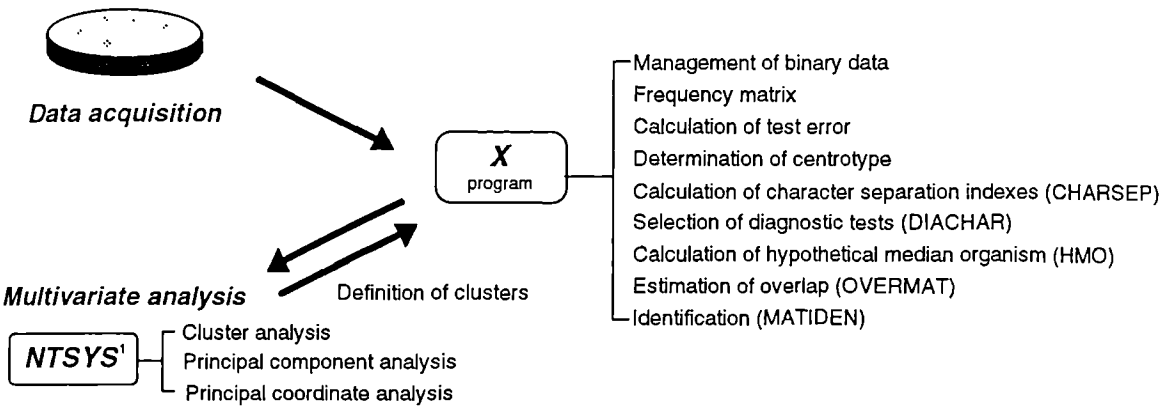


FIGURE 1-5. MAJOR STEPS AND SOFTWARE TOOLS USED IN THE ANALYSIS OF NUMERICAL PHENOTYPIC DATA. [1], COMMERCIALLY AVAILABLE FROM EXETER SOFTWARE (100 NORTH COUNTRY ROAD, BUILDING B, SETAUKET, NEW YORK 11733, USA).

the centrotype (Silvestri *et al.*, 1962), represents the OTU with the highest mean resemblance to all of the other OTUs in the cluster. It is also the nearest OTU to the centroid of the Euclidean distance model. The second representation, the centroid, is a point in phenetic space whose coordinates are the mean values of each character over the given cluster of OTUs, it can be derived from the frequency matrix (Sneath & Sokal, 1973). The hypothetical median organism (HMO; Liston *et al.*, 1963) is commonly used with binary data. This hypothetical organism possesses the commonest state for each character hence it is sometimes called the hypothetical modal organism (Sneath & Sokal, 1973).

The module '***Centrotype***' can be used to provide information for detecting the centrotype of a given cluster. The calculation of the mean pairwise dis/similarity within a cluster (intra-cluster dis/similarity) and the mean values of three different dis/similarities, namely the S_{SM}, S_J and D_P coefficients, for each OTU against the rest of OTU's in a cluster are generated by using this module. The OTU that shows the largest S_J and S_{SM} coefficients and the smallest D_P coefficient can be selected as the centrotype of a given cluster. The module can also be used to calculate the taxonomic distance of each OTU from the HMO; in such a case, the OTU that is the closest to the HMO is the centrotype. Usually only one OTU satisfies all four criteria.

Hypothetical median organism. The module '***Hypothetical median organism***' is used to display the HMOs from a frequency matrix and to identify the HMOs of given clusters against all of the clusters. This procedure has been implemented in the program ***MOSTTYP*** (Sneath, 1980b) and provides a useful means of evaluating identification matrices (Priest & Williams, 1993). A sound identification matrix should give excellent identification scores for the HMOs against their own clusters (Sneath, 1980b).

Character separation indices. Sneath (1979e) developed a BASIC program, namely ***CHARSEP***, for determining the value of each character as a potential separator of groups defined in frequency matrices. The module ***'CHARSEP'*** is a C++ version of Sneath's ***CHARSEP*** program (Sneath, 1979e). This module is used to compute five character separation indices:

(a) Gyllenberg's sum of *C(i)* (Gyllenberg, 1963) is $\sum_{J=1}^{q}(0.5+|0.5-P_{iJ}|)$,

where *i* is the character, *q* is the number of OTU's, P_{iJ} is the proportion of positive values for the *i*th character in cluster *J*.

(b) Gyllenberg's rank measure *R(i)* (Gyllenberg, 1963) is $q_0 q_1 (Sum\, C(i))$,

where q_0 and q_1 are the numbers of clusters for which character *i* is able to allocate an unknown OTU either to a cluster for which the character is negative or to one for which it is positive. This measure uses the given cutoff level *F* as the criterion of whether a cluster is considered positive or negative. That is, q_0 is the number of clusters for which $P_{iJ} < (1-F)$ and q_1 is the number of clusters for which $P_{iJ} \geq F$, where *F* lies between 0.5 to 1.0. Sneath (1979e) recommended the cutoff level of 0.85 since characters with frequencies of positive values between 15 and 85% were usually of little value in achieving a successful identification (Lapage *et al.*, 1973). This cutoff level was successfully used in a numerical taxonomic study of the genus *Streptomyces* (Williams *et al.*, 1983b).

(c) Niemelä's separation index is $\ln(q_0+q_1)! - \ln q_0! - \ln q_1!$ (Niemelä *et al.*, 1968).

(d) The index VSP is $\sum_{J=1}^{q} P_{iJ}^2 / q - \left(\sum_{J=1}^{q} P_{iJ} / q\right)^2$ (Sneath, 1979e).

(e) The index CSP is $1 - 4\left(\sum_{J=1}^{q} P_{iJ} - \sum_{J=1}^{q} P_{iJ}^2\right) / q$ (Sneath, 1979e).

High usefulness of a character *i* is indicated by the following criteria (Sneath, 1979e): high sum of *C(i)*, which ranges from *q/2* to *q*; high value of the product of q_o and q_1 and of *R(i)*, which ranges from 0 to $q^3/4$; high value of Niemelä's separation index, which ranges from 0 to high numbers depending on *q*; and high value of the CSP and VSP indices, which range from 0 to 1. The latter indices are independent of the cutoff value (*F*).

The output file of the module ***CHARSEP*** consists of the values of q_o, q_1, and the separation indices outlined above.

Selection of diagnostic characters. Sneath (1980a) developed a program, called ***DIACHAR***, to determine diagnostic scores from frequency matrices. The diagnostic scores are given as $\sqrt{C_{iJ}D_{iJ}}$, where C_{iJ} is the constancy of the character *i* in cluster *J* and D_{iJ} the average difference between the proportion for character *i* in cluster J and the proportions in all of the other clusters. The consistency is calculated as:

$$C_{iJ} = (2P_{iJ} - 1)^2$$

and the average difference is given as:

$$D_{iJ} = \sum_{K=1}^{q}(P_{iJ} - P_{iK})2/(q-1).$$

The module ***'Diagnostic'*** is the C++ version of Sneath's ***DIACHAR*** program (Sneath ,1980a).

Identification of unknown strains. Sneath (1979d) compiled various identification coefficients in a BASIC program, called ***MATIDEN.*** The module ***'Identification'***, a modified version of the original program ***MATIDEN***, contains two basic identification systems, namely the Baysian and taxon-radius models.

Willcox probability, a Baysian approach, is the likelihood of an unknown strain (*u*) against cluster *J* divided by the sum of the likelihoods of *u* against all *q* clusters. The likelihood L_{uJ} of *u* against cluster *J* is given as:

$$L_{uJ} = \prod_{i=1}^{m} |u_i + P_{iJ} - 1|$$

where *m* is the number of characters under consideration. The Willcox probability is calculated as shown below:

$$Willcox\ probability = L_{uJ} / \sum_{J=1}^{q} L_{uJ}$$ (Willcox *et al.*, 1973).

The module can also be used to calculate several coefficients that are based on the taxon-radius model. In this model, the taxonomic distance (*d*) between an unknown strain *u* and the centroid of taxon *J* is given as:

$$d = \sqrt{[\sum_{i=1}^{q} (u_i - P_{iJ})^2 / m}$$ (Sneath, 1979d)

and the mean of squared distances from OTUs to the taxon centroid, $\bar{d}_J^2$, is calculated as:

$$\bar{d}_J^2 = \sum_{i=1}^{q} P_{iJ}(1 - P_{iJ}) / m$$

which is the variance in hyperspace about the centroid uncorrected for the number of OTUs in the sample used to represent the taxon (Sneath, 1974b). The mean of distances from OTUs in taxon *J* to the centroid ($\bar{d}_J$), and its standard deviation (SD_J), can be approximated on the assumption that the taxa are hyperspherical normal clusters as:

$$\bar{d}_J = \sqrt{[(2m-1)\bar{d}_J^2 / 2m]}$$

and

$$SD_J = \sqrt{\bar{d}_J^2 / 2m}$$ (Sneath, 1979d).

The standard error score of d, s.e.(d), is the constant c in the equation:

$$d = \bar{d}_J + cSD_J.$$

The module can also be used to calculate the Gaussian integral of s.e.(d), where Gau(s.e.(d)) is over 0.5 for the negative s.e.(d). The Gaussian integration is achieved by using the algorithm of Hill (1985) which was originally developed by *Adams (1969)*.

Sneath (1979d) also provided an algorithm for calculating the pattern distance coefficient, d_p, from frequency matrices:

$$d_P = \sqrt{D_P}$$

$$D_P^2 = D_T^2 - D_V^2$$

where D_P is taxonomic distance (d), and D_V is $(u_i - P_{iJ})|u_i - P_{iJ}| / m$.

The 95% taxonomic radius can be defined on the assumption that there is a normal distribution of OTUs in hyperspherical space. The 95% taxonomic radius of taxon J is given as:

$$d = \bar{d}_J + 1.645 \times SD_J.$$

This radius is specific to the taxon and hence independent of unknown strains. A diagrammatic representation of the taxon-radius model is given in Figure 1-6.

Willcox probability scores provide the straight-forward answers for the identification of unknown strains whereas careful considerations are required for the coefficients derived from the taxon-radius model. The level of Willcox probability

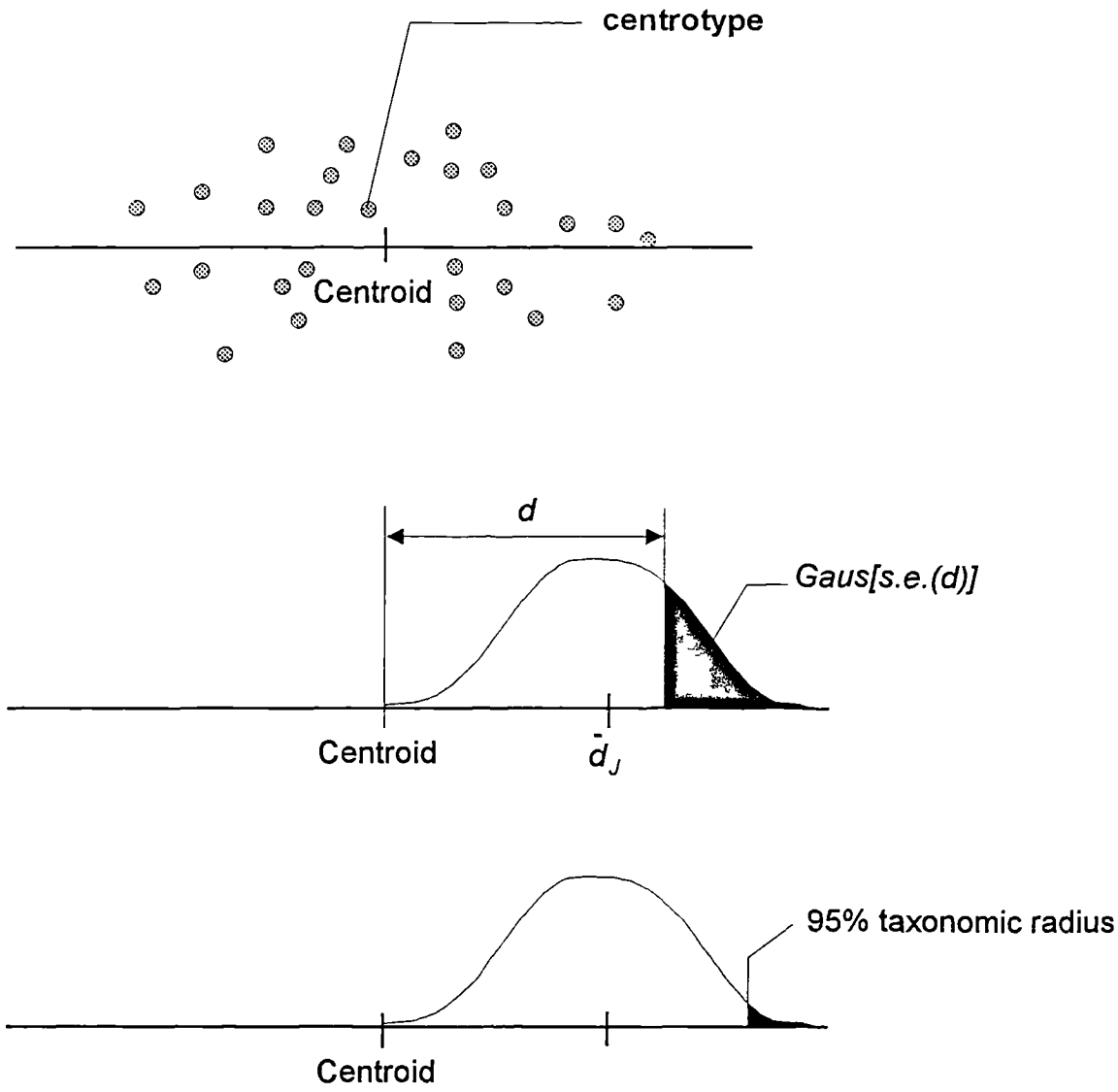


FIGURE 1-6. A DIAGRAMMATIC PRESENTATION OF THE TAXON-RADIUS MODEL BASED ON THE ASSUMPTION THAT THE TAXON IS A HYPERSPHERICAL NORMAL CLUSTER. FOR DETAILS, SEE TEXT.

needed to achieve a successful identification is somewhat subjective (Williams *et al.*, 1985; Priest & Williams, 1993).

However, in the taxon-radius model, the taxonomic distance from each of the taxa in the identification matrix should be compared with the 95% taxon-radius. In most cases organisms will lie close to, or within 2 to 3 standard deviations, of the mean, that is, *s.e.(d)* ranges from 2 to 3. Since a *s.e.(d)* score of 2 implies that very few strains (2.275%; equivalent to the 97.275% taxon radius) can be expected to lie further away from the taxon centroid, scores below 2, or preferably negative scores give successful assignment to a taxon. The Gaussian integration of the *s.e.(d)* scores means that the percentage of OTUs in a taxon lie further away from the centroid than the unknown organism. These scores should be regarded primarily as a test of excluding the null hypothesis that an unknown strain belongs to a given taxon. It is important to note that even if a high Willcox probability score is obtained for a taxon it is still possible that the unknown strain lies further away from the 95% taxonomic radius. It is, therefore, important to consider both identification systems.

Estimation of overlap. The module ***'OVERMAT'*** incorporated in the ***X*** program is a direct translation from the original BASIC version of Sneath's ***OVERMAT*** (Sneath, 1980c) to the C++ version. This module can be used to calculate the phenetic overlap between clusters or taxa as they are represented by the vectors of character values together with a test of significance of the observed overlap (Sneath, 1980c).

The statistical background of overlap between clusters in taxonomic space has been well documented (Sneath, 1977). It is assumed that there are two samples of OTUs which form two clusters, *L* and *M*, that contain n_L and n_M members, respectively. The projections, *q*, of the OTUs onto the line joining the sample centroids of clusters *L* and *M* yield two *q*-distributions (q_L and q_M) [Figure 1-7].

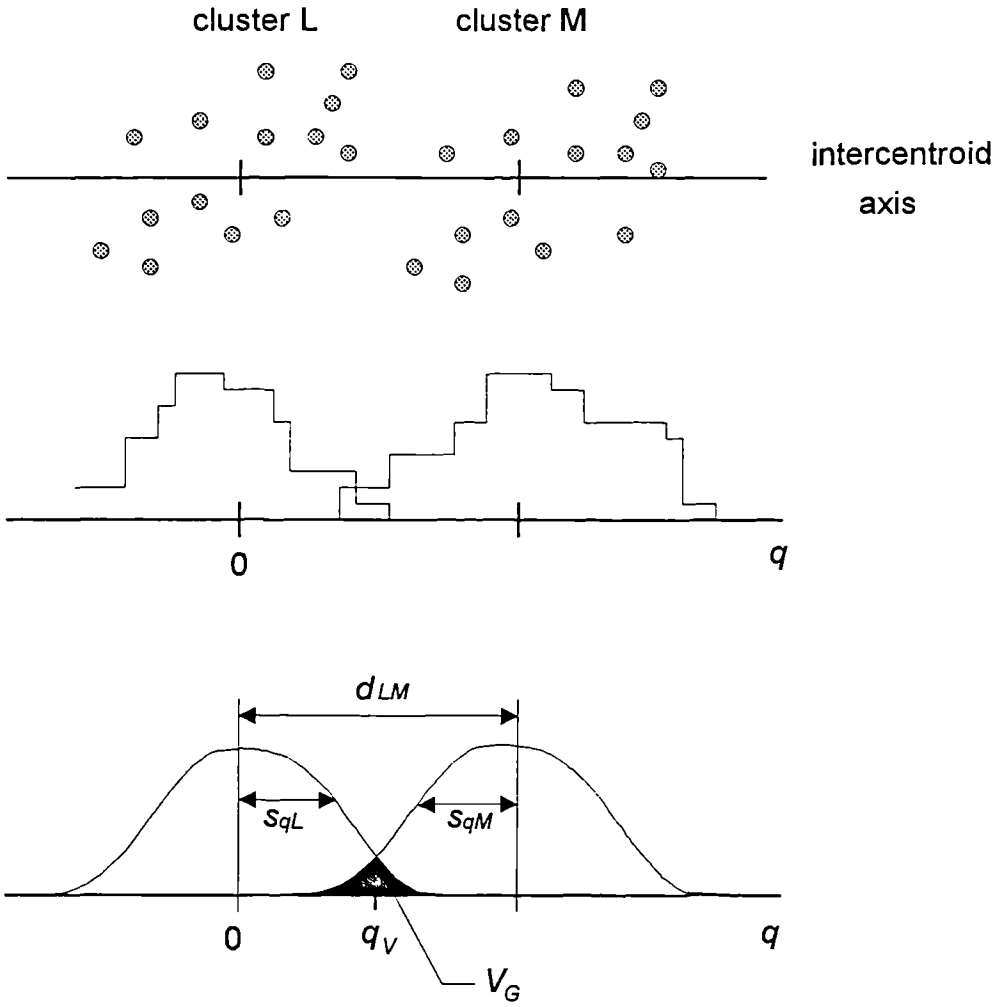


FIGURE 1-7. ILLUSTRATION OF THE METHOD USED TO DERIVE *W* STATISTICS. FOR SYMBOLS, SEE TEXT. WHEN s_{qL} IS EQUAL TO s_{qM}, *W* IS q_V. MODIFIED FROM SNEATH (1977).

If these distributions do not overlap, it is evident that there cannot be overlapping in the hyperspace. The *q* value for OTU *j* of cluster *L* is given as:

$$q_j = (d_{LM}^2 + d_{jL}^2 - d_{jM}^2) / 2d_{LM}$$

and similarly, the *q* value for each OTU *k* of cluster *M* is calculated as:

$$q_k = (d_{LM}^2 + d_{kL}^2 - d_{kM}^2) / 2d_{LM}$$

where the q_j and q_k values are the distributions q_L and q_M, respectively and d_{LM} is the intercentroid distance between taxa *L* and *M*. The mean of q_L will be zero and that of q_M will be d_{LM}. The standard deviations of q_L and q_M, s_{qL} and s_{qM}, are also required for estimating overlapping clusters (Figure 1-7)

A convenient index of overlap in the *q*-distributions is the area of q_L that lies to the right of the point of intersection of the curves, plus the area of q_M that lies to the left of this point in relation to the total area of both curves (Sneath, 1977). When applying normal distribution statistics, these curves will intersect at the point $q_V = s_L d_{LM} / (s_L + s_M)$ and the overlapping area between two curves can be defined as $2P(q_V)$ where $P(q_V)$ is the one-tailed Gaussian integral of q_V/s_{qL}. This index of overlap, V_G, varies from 1 when the means coincide, to zero when they are infinitely apart.

Sneath (1977) also derived an index of disjunction, *W*, which corresponds to the critical value for the two-tailed normal distribution provided variances of two given clusters, namely s_{qL}^2 and s_{qM}^2, are identical. For example, V_G is 0.01 when *W* is 2.576. When two variances are unequal this correspondence is not exact, but for the ratio of variances up to 1:10 it is fairly close (Sneath, 1977). In such cases, *W* is equivalent to a value of V_G that is greater than $2P(q_V)$, so that *W* will be conservative in the sense that it will not mislead one into believing the overlap is less than the true value.

The module ***'OVERMAT'*** is used to compute the variances and standard deviations for each taxon as follows:

$$\mathrm{var}(J) = n_J[\sum_{i=1}^{m} P_{iJ}(1-P_{iJ})]/m(n_J-1)$$

and

$$sd(J) = \sqrt{\mathrm{var}(J)}$$

where m is the number of characters and n_J is the number of OTUs in taxon J. The index of disjunction between taxa L and M (Sneath, 1977) is given as:

$$W_{LM} = \frac{d_{LM}}{\sqrt{(n_L+n_M)(s_{qL}^2/n_L + s_{qM}^2/n_M)}}$$

and

$$d_{LM} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(P_{iL}-P_{iM})^2}$$

where d_{LM} is the intercentroid distance between taxa L and M, and s_{qL}^2 and s_{qM}^2 are taken as *var(L)/m* and *var(M)/m*, respectively. The latter calculations are based on the assumption that the clusters are hyperspherical and hence the variance of a cluster J along any single axis, the intercentroid axis, is *var(J)/m*. Since the intercentroid distance, d_{LM}, is biased by application of finite numbers of OTUs, a corrected estimate of d_{LM}^2 can be achieved by subtracting a correction factor which is:

$$ns_{qL}^2/n_L + ns_{qM}^2/n_M$$

In cases where distances are corrected to give negative values, distances are set to zero.

The index of disjunction, *W*, ranges from 0 for complete overlap to infinity for complete disjunction. The index of overlap, V_G, is twice the standardised Gaussian integral for *W* which means that index V_G ranges from one for complete overlap to zero for complete disjunction.

The significance of *W* is tested by a noncentral *t*-test using Welch's *t*-test (Welch, 1947) and the approximation method of Johnson and Welch (1937). The value of *t* corresponding to the observed W_{LM} is:

$$t_W = W_{LM}\sqrt{n_L + n_M}$$

This is tested by using *F* "effective degrees of freedom" where

$$F = \frac{1}{c^2/(n_L - 1) + (1-c)^2/(n_M - 1)}$$

and

$$c = \frac{n_M \operatorname{var}(L)}{n_M s_{qL}^2 + n_L s_{qM}^2}$$

The module is used to calculate the noncentral t_0 values for the critical V_0 ,given by the user, at the confidence levels of *P*=0.90, 0.95 and 0.99. For example, if one considers that the critical V_0 is given as 0.05 (5%), the observed t_w is 8.72 and the t_0 values are 7.71, 8.76 and 9.34 for *P* of 0.90, 0.96 and 0.99, respectively. The result indicates that the overlap is significantly less than 5% at the 90% confidence level, but t_w does not reach the t_0 values required for confidence levels 95 or 99%.

C. Results and Discussion

1. The AL16S program

The ***AL16S*** program is not evaluated in detail in this Chapter as most of the functions of the program were used in the molecular systematic studies described in *Chapters II* and *III*.

2. The X program

The functionality and accuracy of the ***X*** program were evaluated by using a small dataset given by Sneath (1979d, 1979e, 1980a, 1980b, 1980c) and a larger dataset generated by Williams *et al.* (1983a). The outputs from the former dataset were identical to those given in the original publications of Sneath (data not shown). The latter dataset is considered in detail. The performances of each of the analyses are summarised in Table 1-1. The times required for the numerical analyses implemented in the ***X*** program were reasonably short, that is, within three minutes. This level of performance is appropriate for the analysis of multiple datasets with different choices of strains.

The clusters and unit characters used to evaluate the ***X*** program are shown in Tables 1-2 and 1-3. The frequency matrix generated from the original data matrix consisted of 307 strains and 139 unit characters; the strains were assigned to 26 taxa, that is, 18 major (> 5 strains), 5 minor clusters and 3 subclusters. All of the data were stored as a file which served as the source of input to most of the numerical analyses.

TABLE 1-1. PERFORMANCE OF NUMERICAL ANALYSES IMPLEMENTED IN THE ***X*** PROGRAM[1].

Analysis	Data size	Time (second)
Percent positive table (frequency matrix)	307 OTUs; 139 unit characters; 26 clusters	40
Plus table	307 OTUs; 139 unit characters; 26 clusters	38
Centrotype	307 OTUs; 139 unit characters; 26 clusters	90
CHARSEP	26 clusters; 139 unit characters	0.77
Diagnostic (DIACHAR)	26 clusters; 139 unit characters	2.31
OVERMAT	26 clusters; 139 unit characters	0.66
Hypothetic median organism	26 clusters; 139 unit characters	2.97
Identification	26 clusters; 139 unit characters	2.21[2]

[1], The results were obtained using an IMB-PC compatible personal computer (486-DX2; 50 MHz).

[2], Time required for identification of ten unknown strains.

TABLE 1-2. CLUSTERS USED TO EVALUATE THE ***X*** PROGRAM.

Cluster number	Cluster name*	Number of strains
1A	*Streptomyces albidoflavus*, subcluster A	20
1B	*Streptomyces anulatus*, subcluster B	38
1C	*Streptomyces halstedii*, subcluster C	13
3	*Streptomyces atroolivaceus*	9
5	*Streptomyces exfoliatus*	18
6	*Streptomyces violaceus*	8
10	*Streptomyces fulvissimus*	9
12	*Streptomyces rochei*	26
15	*Streptomyces chromofuscus*	9
16	*Streptomyces albus*	6
17	*Streptomyces griseoviridis*	6
18	*Streptomyces cyaneus*	38
19	*Streptomyces diastaticus*	20
20	*Streptomyces olvaceoviridis*	7
21	*Streptomyces griseoruber*	9
23	*Streptomyces microflavus*	5
29	*Streptomyces lydicus*	11
30	*Streptomyces filipinensis*	4
31	*Streptomyces antibioticus*	5
32	*Streptomyces violaceoniger*	6
33	*"Streptomyces chromogenus"*	5
37	*Streptomyces griseoflavus*	6
40	*Streptomyces phaeochromogenes*	6
42	*Streptomyces rimosus*	7
61	*Streptomyces lavendulae*	12
65	*Streptomyces (Kitasatoa) sp.*	4

*, Clusters defined according to Williams *et al.* (1983a). Names in inverted commas were not included in the Approved Lists of Bacterial Names (Skerman *et al.*, 1980) and have not been validly published subsequently.

TABLE 1-3. UNIT CHARACTERS USED TO EVALUATE THE *X* PROGRAM[+].

Code	Unit characters
	Morphology and pigmentation:
SPS	Presence of spores on aerial mycelium
RFS	Rectiflexibiles
RAS	Retinaculiaperti
SPI	Spirales
BIV	Verticillati
SMO	Smooth spores
WRT	Warty spores
SPY	Spiny spores
HRY	Hairy spores
RUG	Rugose spores
AMY	Aerial mycelium production
RED	Red aerial spore mass
YEL	Yellow aerial spore mass
GRY	*Gray aerial spore mass*
GRN	Green aerial spore mass
BLU	Blue aerial spore mass
VIO	Violet aerial spore mass
WHI	White aerial spore mass
YBS	No distinctive substrate mycelial pigments
ROS	Red/orange substrate mycelial pigments
GNS	Green substrate mycelial pigments
BLS	Blue substrate mycelial pigments
VIS	Violet substrate mycelial pigments
PIG	Production of diffusible pigments
ROP	Red/orange diffusible pigments
YBP	Yellow/brown diffusible pigments
GNP	Green diffusible pigments
BLP	Blue diffusible pigments
VIP	Violet diffusible pigments
PHS	Sensitivity of substrate pigment to pH
PHP	Sensitivity of diffusible pigment to pH
MPI	Melanin production on peptone/yeast/iron agar
MTY	Melanin production on tyrosine agar
FRG	Fragmentation of mycelium
SCL	Sclerotia formation
SBS	Sporulation on substrate mycelium
	Enzyme activity:
LEC	Lecithinase (on egg-yolk medium)
PRT	Proteolysis (on egg-yolk medium)
LIP	Lipolysis (on egg-yolk medium)
HIP	Hippurate hydrolysis

TABLE 1-3. CONTINUED.

Code	Unit characters
PEC	Pectin hydrolysis
CHI	Chitin hydrolysis
NO3	Nitrate reduction
H2S	Hydrogen sulphide production
YPG	β-lactamase production on YPG agar
BFS	β-lactamase production on Beecham's FS agar
KLE	Production of *Klebsiella* β-lactamase inhibitor
	Antimicrobial activity against:
SUB	*Bacillus subtilis* NCIB 3610
PSE	*Pseudomonas fluorescens* NCIB 9046
COL	*Escherichia coli* NCIB 9132
LUT	*Micrococcus luteus* NCIB 196
ALB	*Candida albicans* CBS 562
CER	*Saccharomyces cerevisiae* CBS 1171
MUR	*Streptomyces murinus* ISP 5091
NIG	*Aspergillus niger* LIV 131
	Degradation of:
HYP	Hypoxanthine
GUA	Guanine
ELA	Elastin
TYR	L-Tyrosine
ADE	Adenine
XAN	Xanthine
DNA	DNA
RNA	RNA
T80	Tween 80
STA	Starch
XYN	Xylan
CAS	Casein
TES	Testosterone
URE	Urea
ALL	Allantoin
GEL	Gelatin
AES	Aesculin
ARB	Arbutin
	Resistance to antibiotics (μg/ml):
GEN	Gentamicin (100)
NEO	Neomycin (50)
STR	Streptomycin (100)
TOB	Tobramycin (50)
RIF	Rifampicin (50)
CEP	Cephaloridine (100)

TABLE 1-3. CONTINUED.

Code	Unit characters
VAN	Vancomycin (50)
DMC	Dimethylchlortetracycline (500)
OLE	Oleandomycin (100)
LIN	Lincomycin (100)
PEN	Penicillin G (10 i.u.)
	Growth at:
4OC	4°C
10C	10°C
37C	37°C
45C	45°C
4.3	pH 4.3
	Growth in the presence of (%, w/v, v/v):
4NA	Sodium chloride (4)
7NA	Sodium chloride (7)
10N	Sodium chloride (10)
13N	Sodium chloride (13)
01Z	Sodium azide (0.01)
02Z	Sodium azide (0.02)
1OH	Phenylethanol (0.1)
30H	Phenylethanol (0.3)
PHN	Phenol (0.1)
01T	Potassium tellurite (0.001)
1TL	Potassium tellurite (0.01)
TH1	Thallous acetate (0.01)
T01	Thallous acetate (0.001)
XVI	Crystal violet (0.0001)
	Growth on sole nitrogen source (0.1%, w/v):
BUT	DL-α-amino-*n*-butyric acid
POT	Potassium nitrate
CYS	L-Cysteine
VAL	L-Valine
THR	L-Threonine
SER	L-Serine
PHE	L-Phenylalanine
MET	L-Methionine
HIS	L-Histidine
ARG	L-Arginine
HYD	L-Hydroxyproline
	Growth on sole carbon source (1%, w/v):
ARA	L-Arabinose
SUC	Sucrose

TABLE 1-3. CONTINUED.

Code	Unit characters
XYL	D-Xylose
INO	*meso*-inositol
MAN	Mannitol
FRU	D-Fructose
RHA	L-Rhamnose
RAF	Raffinose
MEZ	D-Melezitose
MNS	D-Mannose
LAC	D-Lactose
INU	Inulin
ADO	Adonitol
SAL	Salicin
TRE	Trehalose
MEB	D-Melibiose
DEX	Dextran
GAL	D-Galactose
CEL	Cellobiose
XYT	Xylitol
	Growth on sole carbon source (0.1%, w/v):
ACE	Sodium acetate
CIT	Sodium citrate
MAL	Sodium malonate
PRO	Sodium propionate
PYR	Sodium pyruvate

*, In order of the appearance in the data file.

The extent of overlap between clusters was estimated by using the module ***OVERMAT***. None of clusters showed overlap greater than 0.01% (V_G) using the uncorrected intercentroid distances. However, two pairs of clusters showed overlap over 0.01% (V_G) when the intercentroid distances were corrected for estimated sampling error (Sneath, 1980c); the output is given in Figure 1-8. It is evident that overlap occurred between clusters 18 (*Streptomyces cyaneus*) and 30 (*Streptomyces filipinensis*), and between clusters 18 (*Streptomyces cyaneus*) and 31 (*Streptomyces violaceoniger*). The V_G values for these overlaps, namely 0.0132 and 0.0128, were slightly higher than the given cutoff value (1%). When the critical overlap value (V_O) of 5% was applied, the t_w value from both pairs of clusters failed to reach the t_0 values required for confidence levels of 90, 95 and 99%, respectively.

The module ***CHARSEP*** was used to select tests for the generation of an identification matrix. The output shows the list of unit characters in order of the VSP values (Figure 1-9). The unit characters that represent VSP values of at least 25%, that is, 47 out of the 139 unit characters, were used to construct the identification matrix. The cutoff value for the ***CHARSEP*** analysis, *F*, was set at 0.85.

The frequency matrix was examined by using the module ***DIACHAR*** to select few additional tests to help differentiate poorly defined clusters. The detailed output of this analysis is given in Figure 1-10. The sums of the diagnostic scores ranged from 23.96 (cluster 18; *Streptomyces cyaneus*) to 35.63 (cluster 65; *Streptomyces [Kitasatoa] sp.*). An additional 22 unit characters were added to the identification matrix constructed on the basis of the output from the ***CHARSEP*** analysis.

The identification matrix, which consisted of 26 clusters and 69 unit characters, was evaluated theoretically first by using the module ***Hypothetical median organism.*** The output of this analysis is shown in Figure 1-11. All of the

HMOs were identified correctly to the parental clusters with high Willcox probability scores (> 0.9999) and negative *s.e.(d)* coefficient values thereby indicating that the identification matrix was sound. The matrix was evaluated further by employing the module ***OVERMAT*** using the uncorrected intercentroid option; the results of this analysis are summarised in Table 1-4. Six pairs of clusters showed overlap values over 1% (V_G); the highest overlap value , 2.17%, was found between clusters 18 (*Streptomyces cyaneus*) and 31 (*Streptomyces antibioticus*). One pair of taxa, namely clusters 18 (*Streptomyces cyaneus*) and 19 (*Streptomyces diastaticus*), showed a significant overlap at the 95% confidence level when the 5% critical V_G value was applied.

Ten strains randomly selected from the original *Streptomyces* database (Williams *et al.*, 1983a) were identified using the identification matrix generated in the present study; the results are summarised in Table 1-5. Eight out of the ten strains were correctly identified to the corresponding taxon with high Willcox probability scores (>0.9999) and small *s.e.(d)* values (<1.0). One organism, namely *Streptomyces* strain C1S.S (S. T. Williams lab. no. C1) from cluster 1B (*Streptomyces anulatus*), was correctly identified albeit with a slightly low Willcox probability score of 0.9883 but the *s.e.(d)* coefficient was very small (0.4). The remaining organism, *Streptomyces fumanus* strain 154FU (ISP 5154) from cluster 18 (*Streptomyces cyaneus*), was assigned to cluster 12 (*Streptomyces rochei*) with a Willcox probability score of 0.8926 and a small *s.e.(d)* coefficient of 1.4854. The next best alternative for this strain was its assignment to the parental taxon (cluster 18; *Streptomyces cyaneus*), with a low Willcox probability score (0.0948) though the corresponding *s.e.(d)* value of 1.5127 was very close to that for cluster 12 (*Streptomyces rochei*; 1.4854). It is noteworthy that the *Streptomyces fumanus* strain (154FU) showed little DNA relatedness with two representative strains of

cluster 18 (*Streptomyces cyaneus*) [Labeda & Lyons, 1991]. It would seem, therefore, that this organism is an atypical member of the *Streptomyces cyaneus* cluster.

FIGURE 1-8. COMPUTER GENERATED OUTPUT OF THE MODULE ***OVERMAT*** OBTAINED WITH 26 *STREPTOMYCES* TAXA (18 MAJOR CLUSTERS, 5 MINOR CLUSTERS AND 3 SUBCLUSTERS). THE INTERCENTROID DISTANCES WERE CORRECTED ACCORDING TO SNEATH (1980C).

```
;  Overmat analysis (Max. Matrix : Cluster 100 x Test 500)
Number of taxa    : 26
 Number of tests : 139

OPTION = 1: INTERCENTROID DISTANCE IS
CORRECTED FOR ESTIMATED SAMPLING ERROR

-----------------------------------------------------------
    TAXON    N[J]    VAR[J]      SD[J]
-----------------------------------------------------------
        1A     20    0.086812    0.294638
        1B     38    0.093141    0.305190
        1C     13    0.085551    0.292490
         3      9    0.099454    0.315363
         5     18    0.105691    0.325102
         6      8    0.101952    0.319299
        10      9    0.089100    0.298496
        12     26    0.103926    0.322375
        15      9    0.108446    0.329311
        16      6    0.082055    0.286452
        17      6    0.101672    0.318860
        18     38    0.105089    0.324174
        19     20    0.105099    0.324189
        20      7    0.099045    0.314714
        21      9    0.093026    0.305002
        23      5    0.103251    0.321327
        29     11    0.100781    0.317460
        30      4    0.101662    0.318845
        31      5    0.111172    0.333424
        32      6    0.098523    0.313884
        33      5    0.097763    0.312670
        37      6    0.097335    0.311985
        40      6    0.104763    0.323672
        42      7    0.098747    0.314241
        61     12    0.105517    0.324833
        65      4    0.087339    0.295532
-----------------------------------------------------------

cutoff level of V(G) = 0.010000

-----------------------------------------------------------
  M                  Q                  C
139                  26                  2.575813
-----------------------------------------------------------
CRITICAL OVERLAP V(0) FOR W0 =0.050000
CORRESPONDING TO W(0) OF 1.959967
```

FIGURE 1-8. CONTINUED.

```
 Overlap statistics are printed as follows :
 on 1st line : names of taxa[number of OTU's]
 on 2nd line : D(L,M)    W    V(G)
 on 3rd line : T(w), and three values for T(0) at P = 0.90, 0.95,
0.99, respectively.


(18[38] vs. 30[4])
          D(LM)=  0.2287 W=  2.4780 V=0.0132
          T(w) = 16.0594 , 25.7708, 31.6348, 47.0337
(18[38] vs. 31[5])
          D(LM)=  0.2188 W=  2.4883 V=0.0128
          T(w) = 16.3169 , 22.6789, 26.7892, 37.8363

TOTAL CPU TIME: 0.66 SECOND(S)
```

FIGURE 1-9. COMPUTER GENERATED OUTPUT FROM THE ***CHARSEP*** ANALYSIS OBTAINED USING THE *STREPTOMYCES* DATABASE (WILLIAMS *et al.*, 1983a). ONLY TESTS SHOWING AT LEAST A 20% VSP COEFFICIENT ARE LISTED.

```
;  Character Separation Index Analysis (Max. Matrix : Cluster 100 x Test 500)
 Number of taxa  : 26
 Number of teste : 139
 Cutoff Value    : 0.85
 Sorted on VSP values
```

Name	q0	q1	VSP(%)	CSP	C(i)	R(i)	Niemela 4var		Cons.	Pot.
MPI	11	5	55.78	0.6011	22.670	1246.850	8.382	0.5843	0.6297	0.9546
RFS	12	5	53.80	0.6074	23.000	1380.000	8.730	0.5816	0.6566	0.9250
ADO	10	5	53.52	0.5655	22.360	1118.000	8.007	0.5525	0.5838	0.9687
SPI	6	7	53.25	0.5340	21.410	899.220	7.448	0.5333	0.5348	0.9985
LUT	6	7	52.76	0.5372	21.760	913.920	7.448	0.5327	0.5424	0.9903
GRY	10	4	49.57	0.5166	21.480	859.200	6.909	0.5065	0.5278	0.9786
NO3	6	5	49.16	0.4926	21.470	644.100	6.136	0.4921	0.4931	0.9990
MUR	4	7	47.38	0.4785	21.040	589.120	5.799	0.4760	0.4808	0.9952
ALL	6	5	47.27	0.4727	20.790	623.700	6.136	0.4727	0.4727	1.0000
MTY	11	4	46.95	0.5563	22.510	990.440	7.219	0.5194	0.6155	0.9039
OLE	9	2	45.30	0.4862	21.450	386.100	4.007	0.4692	0.5035	0.9656
7NA	4	5	45.10	0.4570	21.010	420.200	4.836	0.4537	0.4598	0.9940
RHA	2	6	44.55	0.4697	21.470	257.640	3.332	0.4568	0.4817	0.9751
RAF	3	9	44.13	0.4758	21.340	576.180	5.394	0.4577	0.4935	0.9642
PEC	10	4	43.96	0.4744	20.890	835.600	6.909	0.4561	0.4922	0.9639
CER	13	5	41.49	0.5559	23.060	1498.900	9.056	0.4998	0.6698	0.8300
01Z	7	4	41.13	0.4281	20.460	572.880	5.799	0.4184	0.4355	0.9829
NIG	12	3	37.21	0.5014	22.220	799.920	6.120	0.4391	0.5918	0.8473
PIG	14	4	36.71	0.5207	22.580	1264.480	8.026	0.4529	0.6425	0.8104
LEC	13	2	36.57	0.4732	21.430	557.180	4.654	0.4167	0.5392	0.8775
PHN	3	11	36.26	0.4553	21.300	702.900	5.897	0.4044	0.5078	0.8966
SUB	5	3	36.00	0.3809	19.780	296.700	4.025	0.3678	0.3892	0.9786
ALB	15	4	35.81	0.5429	23.550	1413.000	8.263	0.4741	0.7188	0.7554
45C	9	2	35.72	0.4546	21.360	384.480	4.007	0.4011	0.5105	0.8906
BUT	5	4	35.66	0.3726	20.040	400.800	4.836	0.3625	0.3788	0.9837
VAL	3	4	34.63	0.3545	19.500	234.000	3.555	0.3492	0.3574	0.9918
BFS	4	5	34.21	0.3468	19.280	385.600	4.836	0.3437	0.3485	0.9952
4OC	10	3	33.61	0.4534	21.310	639.300	5.656	0.3889	0.5246	0.8642
PEN	1	10	33.06	0.4213	20.890	208.900	2.398	0.3677	0.4686	0.8991
HIP	10	2	32.95	0.3915	20.370	407.400	4.190	0.3529	0.4193	0.9336
HIS	2	8	32.84	0.4437	21.580	345.280	3.807	0.3787	0.5117	0.8671
LIP	1	13	32.78	0.4643	21.810	283.530	2.639	0.3917	0.5548	0.8369
PHE	2	5	32.32	0.3439	19.690	196.900	3.045	0.3302	0.3513	0.9789
SUC	3	3	31.91	0.3226	19.410	174.690	2.996	0.3202	0.3237	0.9965
INO	2	8	31.67	0.3978	20.680	330.880	3.807	0.3476	0.4367	0.9109
01T	2	7	31.31	0.3854	20.120	281.680	3.584	0.3398	0.4182	0.9216
MEB	1	8	31.08	0.3977	20.830	166.640	2.197	0.3439	0.4400	0.9039
CYS	2	3	30.93	0.3197	19.170	115.020	2.303	0.3126	0.3231	0.9895
HYD	6	2	30.90	0.3752	20.510	246.120	3.332	0.3327	0.4040	0.9288
XAN	2	12	30.74	0.4774	22.600	542.400	4.511	0.3927	0.6100	0.7827
TH1	10	3	30.70	0.4033	20.490	614.700	5.656	0.3442	0.4520	0.8921
RIF	2	9	30.54	0.3984	20.650	371.700	4.007	0.3407	0.4446	0.8961
YPG	1	6	29.67	0.3033	19.120	114.720	1.946	0.2986	0.3053	0.9934
SMO	3	17	29.50	0.5073	23.430	1194.930	7.039	0.4250	0.7309	0.6941
10C	2	13	29.45	0.4753	22.490	584.740	4.654	0.3860	0.6231	0.7629
MEZ	2	2	29.16	0.2963	18.970	75.880	1.792	0.2930	0.2976	0.9953
ELA	1	10	26.75	0.3438	19.450	194.500	2.398	0.2918	0.3750	0.9168
H2S	3	15	24.14	0.4500	22.920	1031.400	6.704	0.3432	0.6398	0.7034
DEX	10	1	22.92	0.3895	21.260	212.600	2.398	0.2866	0.4871	0.7996
ACE	1	4	22.41	0.2322	18.230	72.920	1.609	0.2260	0.2341	0.9919
CHI	9	0	21.26	0.3432	19.990	0.000	0.000	0.2514	0.4060	0.8455
YBP	20	3	21.14	0.4584	24.620	1477.200	7.479	0.3814	0.8272	0.5542
MAN	2	19	20.85	0.4487	23.790	904.020	5.347	0.3480	0.7488	0.5992

```
Total CPU time : 0.88 second(s)
```

FIGURE 1-10. COMPUTER GENERATED OUTPUT FROM THE ***DIACHAR*** ANALYSIS OBTAINED USING THE *STREPTOMYCES* DATABASE (WILLIAMS *et al.*, 1983a)*.

```
;  DIACHAR analysis (Max. Matrix : Cluster 100 x Test 500)
 Number of Taxa  : 26
 Number of Tests : 139
 Number of unit characters to be listed : 5
```

Taxon : 1A (*S. albidoflavus*)

Character	State	Score
TH1	+	0.716554
SPI	-	0.626547
RAF	-	0.588427
LUT	-	0.571904
GRY	-	0.547805
Sum of all scores		29.495970

Taxon : 1B (*S. anulatus*)

Character	State	Score
4OC	+	0.694738
RFS	+	0.517379
GRY	-	0.511014
SPI	-	0.489543
TH1	+	0.459932
Sum of all scores		25.714769

Taxon : 1C (*S. halstedii*)

Character	State	Score
OLE	+	0.675484
TH1	+	0.560010
MPI	-	0.541084
GRY	+	0.520892
MTY	-	0.491710
Sum of all scores		28.667791

Taxon : 3 (*S. atroolivaceus*)

Character	State	Score
4OC	+	0.742395
RFS	+	0.733349
INO	-	0.703517
SPI	-	0.626547
MEZ	-	0.589615
Sum of all scores		28.454195

Taxon : 5 (*S. exfoliatus*)

Character	State	Score
MAN	-	0.858272
RFS	+	0.733349
SPI	-	0.626547
INO	-	0.591281
ADO	-	0.546801
Sum of all scores		24.301828

Taxon : 6 (*S. violaceus*)

Character	State	Score
MTY	+	0.738390
NO3	+	0.615125
PHE	+	0.506352
MPI	+	0.479528
PHN	+	0.457441
Sum of all scores		26.465982

Taxon : 10 (*S. fulvissimus*)

Character	State	Score
OLE	+	0.675484
ROS	+	0.644697
NEO	+	0.624288
VAL	+	0.611143
PHS	+	0.608466
Sum of all scores		30.746105

Taxon : 12 (*S. rochei*)

Character	State	Score
DEX	+	0.511940
MTY	-	0.491710
MPI	-	0.488426
7NA	+	0.487585
RHA	+	0.477929
Sum of all scores		23.717859

Taxon : 15 (*S. chromofuscus*)

Character	State	Score
OLE	-	0.524229
HYD	-	0.458309
CER	-	0.453078
NIG	-	0.443107
PHE	-	0.433234
Sum of all scores		25.356756

Taxon : 16 (*S. albus*)

Character	State	Score
10C	-	0.796057
45C	+	0.727437
ADO	+	0.687147
01Z	+	0.642312
7NA	+	0.626615
Sum of all scores		33.343311

FIGURE 1-10. CONTINUED.

Taxon : 17 (*S. griseoviridis*)

Character	State	Score
YBP	+	0.879291
RED	+	0.827626
PIG	+	0.783157
LUT	+	0.650732
01Z	+	0.642312

Sum of all scores 27.676605

Taxon : 18 (*S. cyaneus*)

Character	State	Score
MPI	+	0.663091
RAF	+	0.520025
SUC	+	0.455069
RFS	-	0.452835
HIP	-	0.433620

Sum of all scores 23.965408

Taxon : 19 (*S. diastaticus*)

Character	State	Score
MUR	-	0.545979
01Z	-	0.461229
RHA	+	0.460712
CER	-	0.453078
LEC	-	0.450736

Sum of all scores 23.123320

Taxon : 20 (*S. olvaceoviridis*)

Character	State	Score
ARB	-	0.848189
MUR	-	0.627506
LUT	-	0.571904
RHA	+	0.531904
OLE	-	0.524229

Sum of all scores 28.528778

Taxon : 21 (*S. griseoruber*)

Character	State	Score
PHP	+	0.903630
PHS	+	0.859797
YBS	-	0.849847
PIG	+	0.783157
PEC	+	0.675057

Sum of all scores 30.408241

Taxon : 23 (*S. microflavus*)

Character	State	Score
PEC	+	0.675057
LUT	+	0.650732
BFS	-	0.600599
ALL	-	0.598526
YPG	-	0.596620

Sum of all scores 29.133127

Taxon : 29 (*S. lydicus*)

Character	State	Score
H2S	-	0.816161
NIG	+	0.760818
LUT	+	0.650732
SPI	+	0.595114
MUR	+	0.570962

Sum of all scores 27.885382

Taxon : 30 (*S. filipinensis*)

Character	State	Score
SPY	+	0.884596
URE	-	0.849872
SMO	-	0.832474
ALB	+	0.813249
CER	+	0.780297

Sum of all scores 32.307411

Taxon : 31 (*S. antibioticus*)

Character	State	Score
H2S	-	0.816161
ALB	+	0.813249
CER	+	0.780297
LIP	-	0.758945
ELA	-	0.688937

Sum of all scores 27.857637

Taxon : 32 (*S. violaceoniger*)

Character	State	Score
RUG	+	0.960400
SMO	-	0.832474
10C	-	0.796057
XAN	-	0.787579
PHN	-	0.723869

Sum of all scores 29.726505

Taxon : 33 ("*S. chromogenus*")

Character	State	Score
YBP	+	0.879291
ALB	+	0.813249
PIG	+	0.783157
CER	+	0.780297
MTY	+	0.738390

Sum of all scores 30.774645

Taxon : 37 (*S. griseoflavus*)

Character	State	Score
RAS	+	0.868756
SMO	-	0.832474
XAN	-	0.787579
XVI	-	0.770048
CIT	-	0.765020

Sum of all scores 31.162279

FIGURE 1-10. CONTINUED.

Taxon : 40 *S. phaeochromogenes*)

Character	State	Score
XYT	-	0.845646
INU	-	0.811426
ADO	-	0.687147
MUR	-	0.627506
VAL	+	0.611143

Sum of all scores 29.751490

Taxon : 42 (*S. rimosus*)

Character	State	Score
STR	+	0.904310
NEO	+	0.878351
NIG	+	0.760818
ADO	-	0.687147
RHA	-	0.661270

Sum of all scores 32.140282

Taxon : 61 (*S. lavendulae*)

Character	State	Score
LEC	+	0.738754
ADE	-	0.718133
MPI	+	0.708054
MAN	-	0.679657
STA	-	0.590818

Sum of all scores 29.239372

Taxon : 65 (*S. [Kitasatoa] sp.*)

Character	State	Score
LAC	-	0.893001
YBP	+	0.879291
XYL	-	0.807118
PIG	+	0.783157
HIS	-	0.738943

Sum of all scores 35.629147

Total CPU time: 0.25 second(s)

*, Details on the unit characters are given in Table 1-3.

FIGURE 1-11. COMPUTER GENERATED OUTPUT FROM THE MODULE '*HYPOTHETICAL MEDIAN ORGANISM*' OBTAINED WITH THE *STREPTOMYCES* DATABASE (WILLIAMS *et al.*, 1983a).

; Hypothetic Median Organism (Max. Matrix : Cluster 100 x Test 500)
Number of taxa : 26
Number of tests : 69
Coefficient for sorting = Willcox Probability

Worked with 1A (*Streptomyces albidoflavus*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
1A	**0.9999**	**0.3525**	**0.2071**	**0.2028**	**-3.86**	**0.9999**
1B	0.0000	0.4057	0.4041	0.3966	1.59	0.0557
1C	0.0000	0.3656	0.4134	0.4132	3.39	0.0003

Worked with 1B (*Streptomyces anulatus*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
1B	**0.9999**	**0.4057**	**0.2267**	**0.2265**	**-4.25**	**1.0000**
3	0.0000	0.4236	0.3472	0.2719	0.72	0.2351
1C	0.0000	0.3656	0.3880	0.3841	2.47	0.0068

Worked with 1C (*Streptomyces halstedii*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
1C	**0.9999**	**0.3656**	**0.1798**	**0.1774**	**-5.14**	**1.0000**
1B	0.0000	0.4236	0.3704	0.3701	0.48	0.3145
3	0.0000	0.3525	0.3941	0.3623	2.40	0.0082

Worked with 3 (*Streptomyces atroolivaceus*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
3	**0.9999**	**0.4245**	**0.2173**	**0.2167**	**-3.93**	**1.0000**
1C	0.0000	0.3525	0.3657	0.3237	1.65	0.0493
1B	0.0000	0.4236	0.3921	0.3474	1.20	0.1158

Worked with 5 (*Streptomyces exfoliatus*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
5	**0.9999**	**0.3865**	**0.2605**	**0.2544**	**-3.44**	**0.9997**
1B	0.0000	0.3525	0.4026	0.3840	1.54	0.0613
6	0.0000	0.4210	0.4027	0.3818	2.21	0.0137

Worked with 6 (*Streptomyces violaceus*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
6	**0.9999**	**0.3711**	**0.2331**	**0.2298**	**-3.65**	**0.9999**
5	0.0000	0.4033	0.3875	0.3856	0.58	0.2796
1B	0.0000	0.3525	0.4607	0.4601	3.46	0.0003

Worked with 10 (*Streptomyces fulvissimus*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
10	**0.9999**	**0.3736**	**0.2111**	**0.2089**	**-4.29**	**1.0000**
18	0.0000	0.4057	0.5173	0.5166	4.07	0.0000
5	0.0000	0.4210	0.5056	0.4990	4.33	0.0000

FIGURE 1-11. CONTINUED.

Worked with 12 (*Streptomyces rochei*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
12	**0.9999**	**0.3865**	**0.2549**	**0.2444**	**-3.67**	**0.9999**
15	0.0000	0.3525	0.3710	0.3500	0.58	0.2822
1C	0.0000	0.4210	0.4220	0.4125	3.71	0.0001

Worked with 15 (*Streptomyces chromofuscus*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
15	**0.9999**	**0.3381**	**0.2252**	**0.2206**	**-4.25**	**1.0000**
20	0.0000	0.3716	0.4072	0.3918	2.42	0.0077
12	0.0000	0.4057	0.4479	0.3881	2.41	0.0079

Worked with 16 (*Streptomyces albus*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
16	**0.9999**	**0.3636**	**0.1933**	**0.1779**	**-3.79**	**0.9999**
12	0.0000	0.4033	0.4601	0.4273	2.80	0.0026
15	0.0000	0.3525	0.4501	0.4500	3.19	0.0007

Worked with 17 (*Streptomyces griseoviridis*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
17	**0.9999**	**0.3656**	**0.2276**	**0.2173**	**-3.80**	**0.9999**
12	0.0000	0.4378	0.4505	0.4503	2.49	0.0063
15	0.0000	0.3802	0.4733	0.4346	3.96	0.0000

Worked with 18 (*Streptomyces cyaneus*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
18	**0.9999**	**0.4210**	**0.2627**	**0.2614**	**-3.69**	**0.9999**
19	0.0000	0.4057	0.3993	0.3970	0.85	0.1963
12	0.0000	0.3525	0.4381	0.4381	2.10	0.0177

Worked with 19 (*Streptomyces diastaticus*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
19	**0.9999**	**0.3907**	**0.2474**	**0.2433**	**-3.92**	**1.0000**
20	0.0000	0.4236	0.3780	0.3767	1.41	0.0793
18	0.0000	0.3936	0.3957	0.3804	0.36	0.3590

Worked with 20 (*Streptomyces olvaceoviridis*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
20	**0.9999**	**0.3137**	**0.2297**	**0.2290**	**-3.74**	**0.9999**
15	0.0000	0.3730	0.3904	0.3884	1.22	0.1114
19	0.0000	0.3636	0.4321	0.4263	1.89	0.0297

Worked with 21 (*Streptomyces griseoruber*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
21	**0.9999**	**0.3842**	**0.2014**	**0.2013**	**-4.31**	**1.0000**
19	0.0000	0.3525	0.4308	0.4229	1.84	0.0327
12	0.0000	0.3711	0.4341	0.4318	1.98	0.0240

FIGURE 1-11. CONTINUED.

Worked with 23 (*Streptomyces microflavus*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
23	**0.9999**	**0.4210**	**0.2310**	**0.2297**	**-3.45**	**0.9997**
18	0.0000	0.3936	0.4776	0.4767	2.86	0.0021
19	0.0000	0.4057	0.4694	0.4643	3.06	0.0011

Worked with 29 (*Streptomyces lydicus*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
29	**0.9999**	**0.3736**	**0.2229**	**0.2229**	**-4.04**	**1.0000**
12	0.0000	0.3711	0.4540	0.4489	2.61	0.0046
18	0.0000	0.3882	0.4926	0.4873	3.32	0.0005

Worked with 30 (*Streptomyces filipinensis*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
30	**0.9999**	**0.3758**	**0.2555**	**0.2206**	**-2.51**	**0.9940**
31	0.0000	0.3907	0.4430	0.4250	3.32	0.0004
29	0.0000	0.4236	0.4538	0.4534	3.90	0.0000

Worked with 31 (*Streptomyces antibioticus*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
31	**0.9999**	**0.3848**	**0.2348**	**0.2346**	**-3.74**	**0.9999**
18	0.0000	0.3711	0.4090	0.4076	0.77	0.2213
19	0.0000	0.3716	0.4202	0.4059	1.51	0.0654

Worked with 32 (*Streptomyces violaceoniger*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
32	**0.9999**	**0.3730**	**0.2387**	**0.2236**	**-3.14**	**0.9992**
15	0.0000	0.3610	0.4504	0.4311	3.20	0.0007
29	0.0000	0.4236	0.4608	0.4608	4.14	0.0000

Worked with 33 ("*Streptomyces chromogenus*")

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
33	**0.9999**	**0.3736**	**0.2273**	**0.2271**	**-3.54**	**0.9998**
19	0.0000	0.3721	0.4672	0.4655	2.99	0.0014
15	0.0000	0.3711	0.4952	0.4740	4.69	0.0000

Worked with 37 (*Streptomyces griseoflavus*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
37	**0.9999**	**0.3865**	**0.2105**	**0.2005**	**-3.92**	**1.0000**
12	0.0000	0.3258	0.4582	0.4186	2.74	0.0031
19	0.0000	0.4378	0.4585	0.4457	2.71	0.0033

Worked with 40 (*Streptomyces phaeochromogenes*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
40	**0.9999**	**0.3848**	**0.2231**	**0.2128**	**-3.78**	**0.9999**
12	0.0000	0.3636	0.4284	0.4273	1.80	0.0361
19	0.0000	0.4057	0.4492	0.4367	2.42	0.0077

FIGURE 1-11. CONTINUED.

Worked with 42 (*Streptomyces rimosus*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
42	**0.9999**	**0.3656**	**0.1850**	**0.1642**	**-4.40**	**1.0000**
29	0.0000	0.3721	0.4560	0.3853	3.98	0.0000
10	0.0000	0.3882	0.4717	0.4581	4.86	0.0000

Worked with 61 (*Streptomyces lavendulae*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
61	**0.9999**	**0.3802**	**0.2358**	**0.2225**	**-3.65**	**0.9999**
5	0.0000	0.3258	0.5019	0.4820	4.21	0.0000
33	0.0000	0.3907	0.5019	0.4839	6.33	0.0000

Worked with 65 (*Streptomyces [Kitasatoa] sp.*)

Taxon	Willcox Prob.	95 % Radius	Taxonomic Distance	Dp	s.e.(d)	Gauss(s.e.(d))
65	**0.9999**	**0.3736**	**0.1881**	**0.1758**	**-3.70**	**0.9999**
5	0.0000	0.3636	0.5167	0.4976	4.68	0.0000
61	0.0000	0.4245	0.5035	0.5023	5.50	0.0000

Total CPU time: 2.03 second(s)

TABLE 1-4. OVERLAP VALUES ESTIMATED FROM THE IDENTIFICATION MATRIX GENERATED FROM THE *STREPTOMYCES* DATABASE (WILLIAMS *et al.*, 1983a)[1].

Cluster pair[2]	D_{LM}[3]	W	V_G	Tw	T_0=0.90	T_0=0.95	T_0=0.99
1B(38):3(9)	0.268	2.496	0.0125	17.116	18.627	20.478	24.871
18(38):19(20)	0.248	2.542	0.0110	19.360	17.895	18.838	20.811
18(38):23(5)	0.348	2.505	0.0122	16.427	22.328	26.257	36.784
18(38):30(4)	0.361	2.331	0.0197	15.108	25.676	31.488	46.781
18(38):31(5)	0.334	2.296	0.0217	15.055	22.539	26.577	37.418
18(38):40(6)	0.318	2.482	0.0131	16.464	20.609	23.590	31.309

[1], The identification matrix consisted of 26 clusters and 69 unit characters.

[2], The number of strains are presented in parentheses.

[3], Uncorrected intercentroid distances.

TABLE 1-5. IDENTIFICATION OF TEN RANDOMLY CHOSEN STRAINS TAKEN FROM THE ORIGINAL MATRIX OF WILLIAMS *et al.* (1983a). THE STRAINS WERE IDENTIFIED USING THE IDENTIFICATION MATRIX WHICH CONSISTED WITH 26 CLUSTERS AND 69 UNIT CHARACTERS.

		Coefficients					
Strain (Cluster)	Identified as*	Willcox probability	95% Taxon-radius	Taxonomic distance	D_p	*s.e.(d)*	*Gaus[s.e.(d)]*
537SA(1A)	**1A**	**0.9999**	**0.3525**	**0.2735**	**0.2703**	**-1.3490**	**0.9113**
	1B	0.0000	0.4057	0.3846	0.3755	0.9506	0.1709
	12	0.0000	0.4236	0.4777	0.4759	3.3509	0.0004
C1S.S(1B)	**1B**	**0.9883**	**0.4057**	**0.3680**	**0.3680**	**0.4057**	**0.3425**
	1C	*0.0069*	0.4236	0.4091	0.4080	3.2365	0.0006
	19	0.0038	0.3936	0.4157	0.4016	1.3693	0.0855
T1S.S(5)	**5**	**0.9999**	**0.3730**	**0.3660**	**0.3658**	**-0.0990**	**0.5394**
	1B	0.0000	0.4236	0.4460	0.4450	2.9717	0.0015
	19	0.0000	0.3802	0.4876	0.4864	3.6299	0.0001
470VI(12)	**12**	**0.9999**	**0.4033**	**0.3478**	**0.3155**	**-0.7442**	**0.7716**
	18	0.0000	0.3842	0.4433	0.4320	1.8151	0.0348
	1B	0.0000	0.4236	0.4450	0.4361	2.9395	0.0016
154FU(18)	**12**	**0.8926**	**0.4057**	**0.4185**	**0.4181**	**1.4854**	**0.0687**
	18	0.0948	0.3882	0.4334	0.4311	1.5127	0.0652
	19	0.0126	0.3711	0.4619	0.4579	2.8235	0.0024
309VA(19)	**19**	**0.9999**	**0.3525**	**0.3656**	**0.3622**	**-0.2059**	**0.5816**
	12	0.0003	0.3865	0.4294	0.4289	1.8301	0.0336
	18	0.0000	0.3736	0.4243	0.4229	1.2344	0.1085
357NA(29)	**29**	**0.9999**	**0.3381**	**0.3178**	**0.2965**	**-0.7758**	**0.7811**
	12	0.0000	0.4236	0.4725	0.4411	3.1874	0.0007
	31	0.0000	0.3802	0.4682	0.4296	4.1740	0.0000
563VI(32)	**32**	**0.9999**	**0.3936**	**0.2679**	**0.2654**	**-2.0928**	**0.9818**
	15	0.0000	0.3137	0.4712	0.4512	3.8914	0.0000
	12	0.0000	0.3865	0.4864	0.4841	3.6257	0.0001
098PR(37)	**37**	**0.9999**	**0.4210**	**0.2707**	**0.2667**	**-1.6918**	**0.9547**
	12	0.0000	0.4245	0.4521	0.4167	2.5452	0.0055
	19	0.0000	0.3730	0.4505	0.4405	2.4638	0.0069
K280D(65)	**65**	**0.9999**	**0.3907**	**0.2935**	**0.2921**	**0.7834**	**0.2167**
	5	0.0000	0.3865	0.5228	0.5066	4.8754	0.0000
	61	0.0000	0.3381	0.5033	0.5015	5.4927	0.0000

*, The three taxa showing the highest Willcox probabilities are listed.

CHAPTER II.
PHYLOGENETIC ANALYSIS OF THE FAMILY *NOCARDIACEAE* AND RELATED ACTINOMYCETES USING 16S RIBOSOMAL RNA GENE SEQUENCES

A. Introduction

1. Phylogeny inferred from 16S ribosomal ribonucleic acid

1.1. 16S rRNA as molecular chronometer

It was pointed out in Chapter I that sequencing of 16S rRNA provides a useful means of unravelling phylogenetic relationships between representatives of bacterial taxa (Woese, 1987; Stackebrandt, 1992; Woese, 1992; Ludwig & Schleifer, 1994). The information held in 16S rRNA molecules has been extensively used to determine genealogical relationships between actinomycetes, especially at the suprageneric level (Goodfellow 1989a; Embley & Stackebrandt, 1994).

1.2. Molecular biological properties of ribosomal RNA gene clusters

Ribosomal RNA genes are amongst the most extensively studied genes in bacteria. In members of most bacterial species, genes that encode for the three rRNA types are closely linked in gene sets (operons) in the order 16S-23S-5S rDNAs. However, exceptions to this pattern are found in *Mycoplasma hyopneumoniae* (the 5S rRNA gene is separated from the genes coding for the 16S and 23S RNAs; Taschke *et al.*, 1986), *Thermus thermophilus* (the genes coding for the 16S and 23S rRNAs are separated by several kilobases; Ulbrich *et al.*, 1984) and *Vibrio harveyi* (the gene set is in the order 23S-16S-5S rDNAs; Lamfrom *et al.*, 1978). With respect to actinomycetes, the rRNA operons of *Frankia* (Normand *et al.*,

1992), *Mycobacterium* (Bercovier *et al.*, 1986; Suzuki *et al.*, 1988a; Sela *et al.*, 1989; Ji *et al.*, 1994a,b) and *Streptomyces* strains (Baylis & Bibb, 1988; Suzuki *et al.*, 1988b; van Wezel *et al.*, 1994) have been studied in some detail. All actinomycetes examined so far show the typical bacterial gene organisation, that is, the gene set is in the order of 16S-23S-5S rDNAs.

The number of rRNA operons in bacterial genomes vary from one (*Mycobacterium tuberculosis*) to eleven (some *Bacillus* species; Gottlieb *et al.*, 1985). There seems to be a broad correlation between the rate of growth of a bacterial strain and the number of constituent ribosomes (Winder & Rooney, 1970; Bremer & Dennis, 1987). The rate at which mature rRNAs are produced in bacteria depends on several factors which include the number of rRNA (*rrn*) operons, the strength of their promoters and the efficiency with which the operons are transcribed and processed. Codon *et al.* (1992) examined all seven rRNA operons of *Escherichia coli* and was unable to find any significant difference in their transcriptional efficacy under fast growth conditions though considerable differences were observed in response to stress, including nutritional shift-down. Van Wezel *et al.* (1994) found that one out of six rRNA operons in "*Streptomyces coelicolor*" strain A3(2) contained four different promoters with strengths that varied in response to the growth phase.

There are a few examples where more than one rRNA operon per organism has been sequenced. In most of these cases the nucleotide sequences of the different operons in the same strain have been found to be either identical to or show a low level of heterogeneity (about 0.1% or few differences in nucleotide positions; Maden *et al.*, 1987; Dryden & Kaplan, 1990; Heinonin *et al.*, 1990). In an analysis of bulk 23 rRNA from *Escherichia coli,* only eight nucleotide differences were observed in the seven copies of the rRNA operons; this corresponds to

approximately one heterogeneous nucleotide per operon (Branlant *et al.*, 1981). Dryden and Kaplan (1990) found that the three 16S rRNA gene sequences in the genome of *Rhodobacter sphaeroides* were identical though one nucleotide substitution and three deletions were found in the corresponding 23S rRNA gene sequences. The greatest heterogeneity was found in the 5S rRNA coding region where five nucleotide differences were found in one of the three rRNA operons.

The halophilic archaebacterium, *Haloarcula marismortui*, is exceptional as it has two nonadjacent rRNA operons, namely *rrnA* and *rrnB*, in which the two 16S rRNA coding regions contain 1472 nucleotides but differ in nucleotide substitutions at seventy-four positions (5% nucleotide sequence dissimilarity; Mylvaganam & Dennis, 1992). These investigators found that each of the rRNA operons were transcribed and that the resultant 16S rRNA molecules were present in intact 70S ribosomes. Mylvaganam and Dennis (1992) noted that none of the seventy-four heterogeneous nucleotide positions were related to positions seen as functionally important for interactions with tRNA, mRNA or translational factors during protein synthesis. Oren *et al.* (1988) directly determined the 16S rRNA sequence of *Haloarcula marismortui* by using the reverse transcriptase sequencing method. These workers did not know that the organism contained mixtures of two 16 rRNAs hence their nucleotide sequence comprised twenty-four nucleotides from the *rrnA* operon and twenty-six nucleotides from the *rrnB* operon.

It is evident from the results of the two independent sequencing studies outlined above that nucleic acid sequencing methods which do not include a cloning step may be flawed in cases where rRNA operons are heterogeneous. However, *Haloarcula marismortui* is the only prokaryote found so far which shows appreciable heterogeneity in its 16S rRNA genes. In addition, the rRNA operons of *Plasmodium berghei*, a blood parasite, were found to have a dissimilarity of 3.5% in two different

small subunit rRNA genes (Gunderson *et al.*, 1987). The two operons were found to be differentially regulated; one expressed in sporozoites in the insect host and the other in the asexual stage in mammalian blood streams.

It is evident both from the literature and from the success of direct sequencing of PCR-amplified rRNA genes that multiple copies of 16S rRNA genes found in microorganisms do not differ significantly. This phenomenon of homogenisation is called **'concerted evolution'** (Arnheim *et al.*, 1980). Several biological processes appear to be responsible for concerted evolution, the most important of these seem to be unequal crossing over (Perelson & Bell, 1977; Petes, 1980; Szostak & Wu, 1980) and gene conversion (Nagylaki & Petes, 1982; Nagylaki, 1984; Enea & Corredor, 1991). The relative contributions of these two genetical mechanisms have been discussed (Dover, 1982a, b) but little experimental data are available to discriminate between their impact (Hillis & Dixon, 1991).

Intergenic spacer regions, which are typically found between 16S and 23S rRNAs, can be used as a source of phylogenetic information on closely related organisms, typically those at and below species rank since these regions show more variability than corresponding rRNA coding regions (Frothingham & Wilson, 1993; Postic *et al.*, 1994). Intergenic spacer regions often differ in length and show high nucleotide sequence dissimilarity between operons. In some cases, only certain rRNA operons have the tRNA gene(s) in the intergenic spacer region between the 16S and 23S rRNA genes (e.g., East *et al.*, 1992). The nucleotide sequence information in these regions has been used to underpin phylogenetic relationships between members of closely related taxa, including *Borrelia burgdorferi* (Postic *et al.*, 1994) and members of the "*Mycobacterium avium* complex" (Frothingham & Wilson, 1993). Several molecular identification systems

based on variation in intergenic spacer regions have been developed; these include PCR (Uemori *et al.*, 1992; Dolzani *et al.*, 1994) and restriction endonuclease analyses of PCR-amplified spacer genes (Harasawa *et al.*, 1993).

1.3. Methods used to determine 16S rRNA sequences

DNA-rRNA hybridisation. Hybridisation methods have been developed to measure the degree of binding between rRNA and rRNA cistrons and the resultant data used to determine relatedness at both generic and suprageneric levels (Kilpper-Bälz, 1991; Goodfellow & O'Donnell, 1993). Members of diverse taxa can be compared in DNA-rRNA hybridisation studies since the nucleotide sequences of rRNA cistrons are considered to be more highly conserved than those of most genes forming the bacterial genome (Doi & Igarashi, 1965; Dubnau *et al.*, 1965; Moore & McCarthy, 1967; Nomura *et al.*, 1968). Members of over three hundred and fifty archaeal and bacterial species have been studied by using this method (Stackebrandt, 1992).

Several techniques have been developed for DNA-rRNA hybridisation (Palleroni *et al.*, 1973; Johnson & Francis, 1975; Baharaeen *et al.*, 1983; Klenk *et al.*, 1986). The most widely used procedure is the saturation hybridisation technique (De Ley & De Smedt, 1975) which is a modification of the original membrane hybridisation technique of Gillespie and Spiegelman (1965). The results of DNA-rRNA hybridisation studies are usually presented as two dimensional similarity maps where $T_{m(e)}$ values are plotted against percentage DNA-rRNA binding (e.g., Mordarski *et al.*, 1980a) or as dendrograms based on $T_{m(e)}$ values (e.g., Vandamme *et al.*, 1994).

Highly related strains belonging to the same taxon have been found to have $T_{m(e)}$ values of about 80°C under experimental conditions used by De Ley and his colleagues (De Smedt & De Ley, 1975; Gillis & De Ley, 1980). Good resolution can

be achieved with $T_{m(e)}$ values within the range 65°C to 80°C; this range of $T_{m(e)}$ values usually allows the establishment of intergeneric or interfamily relationships. The limitation of the technique is reached at $T_{m(e)}$ values of about 60 °C where its has been shown that resolution of relationships between the subclasses of the class *Proteobacteria* is not possible (Stackebrandt *et al.*, 1988).

The relatedness between labelled rRNA and DNA is dependent on a number of factors which include genome size and the number of rRNA operons (Kilpper-Bälz, 1991). Only a few reference strains can be examined in DNA-rRNA hybridisation studies so that results are usually presented in an incomplete triangular format. In general, good congruence has been found between DNA-DNA and DNA-rRNA hybridization data for closely related organisms; the $T_{m(e)}$ values for strains of species that show more than 60% DNA homology differ by less than 2 °C (Stackebrandt, 1992).

Few DNA-rRNA hybridisation studies have been focused on actinomycetes. However, ribosomal RNA cistron similarity data show that acid-fast actinomycetes are phylogenetically close (Mordarski *et al.*, 1980a, 1981). It has also been shown that sporoactinomycetes fall into at least three major phylogenetic clades which correspond to the genera *Actinoplanes, Ampullariella* and *Micromonospora*; to the genera *Planobispora*, *Planomonospora* and *Streptosporangium*; and to the genus *Streptomyces* (including *Chainia*, *Elytrosporangium*, *Kitasatoa, Microellobosporia* and *Streptoverticillium*)[Stackebrandt *et al.* ,1981].

The primary impact of rRNA cistron similarity studies has been on the classification of Gram-negative bacteria at family and suprafamily ranks, notably in the rearrangement of taxa within the class *Proteobacteria* (De Vos *et al.*, 1989). In recent times, the DNA-rRNA hybridisation method has been replaced by more

sensitive methods, mostly by the development and application of rRNA cataloguing and sequencing techniques.

16S rRNA oligonucleotide cataloguing. The application of 16S rRNA oligonucleotide cataloguing provided a more exacting way of detecting phylogenetic relationships between prokaryotes (Fox *et al.*, 1977; Woese & Fox, 1977; Stackebrandt & Woese, 1981a, b). The cataloguing method was introduced by Sanger *et al.* (1965) and its value for unravelling evolutionary relationships developed by Woese and coworkers (Woese *et al.*, 1975; Fox *et al.*, 1977). The importance of the 16S rRNA cataloguing method was that it could be used to obtain a lot of phylogenetic information without the need to sequence the whole 16S rRNA molecule. An additional advantage of the method was that it provided a way of generating databases thereby enabling comparisons of new catalogues with those held in databases (Sobieski *et al.*, 1984). Over 600 microorganisms were examined using the 16S rRNA cataloguing method (Stackebrandt, 1992).

The procedure underpinning the cataloguing method involved enzymatic digestion of purified 16S rRNA by the guanosine-specific ribonuclease T_1, separation of oligonucleotide fragments by two-dimensional electrophoresis and alkaline digestion of fragments followed by nucleotide sequencing of fragments longer than five bases by two dimensional thin-layer-chromatography (Fowler *et al.*, 1985; Fox & Stackebrandt, 1987). Oligonucleotide sequences longer than five bases were considered to be position specific within the 16S rRNA molecule and hence were seen to be appropriate for comparisons between homologous nucleotide positions. A typical catalogue consisted of about 80 fragments (7 to 20 nucleotides in length) that were evenly distributed over the primary structure of the 16S rRNA. Consequently, about 35 to 45% of complete nucleotide sequences were compared using the cataloguing procedure.

The relationship between any two given strains was expressed as a Jaccard-type similarity coefficient (S_{AB}) which was calculated on the basis of the proportion of identical oligonucleotides in the respective catalogues:

$$S_{AB} = \frac{2N_{AB}}{N_A + N_B}$$ (Fox *et al.*, 1977)

where N_A is the total number of residues in oligonucleotides of length at least L in catalogue A, N_B is the total number of residues in oligonucleotides of length at least L in catalogue B and N_{AB} is the number of residues represented by all of the coincident oligonucleotides between two catalogues, A and B, of length at least L. The choice of L was governed by statistical considerations; a figure of six was taken for all but the most closely related organisms (Fox *et al.*, 1977). A matrix of S_{AB} values was constructed and the similarity values sorted using the UPGMA algorithm (Sneath & Sokal, 1973).

Data derived from 16S rRNA cataloguing studies suggested that Gram-positive bacteria formed a distinct phyletic line that could be split into two branches on the basis of DNA base composition values (Stackebrandt & Woese, 1981b). The actinomycete-coryneform line included bacteria with a guanine(G) plus cytosine(C) content above about 55 mol% and the *Clostridium-Bacillus-Streptococcus* branch strains with relatively low G+C values (below 50 mol%).

Several taxa previously associated with actinomycetes were found to belong to the *Clostridium-Bacillus-Streptococcus* group. The genus *Eubacterium* was found to be related to the genus *Clostridium*, the genus *Kurthia* to the lactic acid bacteria, and the genus *Thermoactinomyces* to aerobic, endospore-forming bacilli (Ludwig *et al.*, 1981; Tanner *et al.*, 1981; Stackebrandt & Woese, 1981b). In contrast, members of coryneform taxa, notably *Arthrobacter*, *Brevibacterium*, *Cellulomonas*, *Corynebacterium*, *Curtobacterium* and *Microbacterium*, which rarely, if ever, form a

primary mycelium, were seen to be phylogenetically intermixed with classical actinomycetes (Stackebrandt & Woese, 1981a, b). It is clear from such studies that the possession of branched hyphae can not automatically be used to place a bacterium with actinomycetes. Conversely, the inability of a strain to form branching filaments does not necessarily exclude it from this group of bacteria. The 16S rRNA cataloguing data bear eloquent testimony to the inherent dangers of constructing taxonomies solely on the basis of morphological features (van Niel, 1946).

Direct sequencing of 16S rRNA using the reverse transcriptase method. The 16S rRNA cataloguing method was superseded by the development and application of the reverse trancriptase sequencing method (Qu *et al.*, 1983; Lane *et al.*, 1985). The introduction of the reverse transcriptase method made it possible to determine almost complete 16S rRNA sequences by using primers complementary to the conserved regions in 16S rRNA molecules. The reverse transcriptase method was used to underpin the phylogenetic coherence of several actinomycete taxa, notably the families *Frankiaceae* (Hahn *et al.*, 1989), *Mycobacteriaceae* (Stahl & Urbance, 1990), *Nocardiaceae* (Collins *et al.*, 1988a; Stackebrandt *et al.*, 1988b), *Nocardioidaceae* (Collins & Stackebrandt, 1988; Collins *et al.*, 1989a), *Propionibacteriaceae* (Charfreitag *et al.*, 1988), *Pseudonocardiaceae* (Embley *et al.*, 1988a, b; Warwick *et al.*, 1994), *Streptomycetaceae* (Stackebrandt *et al.*, 1992) and *Streptosporangiaceae* (Kemmerling *et al.*, 1993).

The reliability of nucleotide sequences determined by using the reverse transcriptase method is influenced by strong posttranscriptional base modifications and by the secondary structure of rRNA (Stackebrandt, 1992). However, DNA sequencing methods which involve the polymerase chain reaction (PCR) were subsequently found to give better quality sequence data. The relatively errorfree

polymerase chain reaction(PCR)-based DNA sequencing methods replaced the reverse transcriptase method.

Comparisons between S_{AB} values and corresponding values from complete nucleotide sequences were congruent when relationships were based on complete rRNA sequence data and S_{AB} values greater than 0.5. However, relationships found between bacteria showing S_{AB} values below 0.4 were found to be underestimated (Woese, 1987). Good congruence was found between actinomycete phylogenies derived using the two approaches as S_{AB} values for these organisms were greater than 0.5 (Embley & Stackebrandt, 1994).

16S rRNA sequencing methods based on the polymerase chain reaction. The development and application of the polymerase chain reaction (PCR; Saiki *et al.*, 1988) had an enormous impact on molecular biology, including molecular systematics. Indeed, the application of the PCR to 16S rRNA sequencing coupled with the development of automated nucleotide sequencers, notably those employing non-radioactive labelling (e.g., Applied Biosystems Prism sequencing kits), and computer-assisted data acquisition has helped to revolutionise modern bacterial systematics.

Two procedures are commonly used to determine PCR-amplified 16S rRNA gene (rDNA) sequences. In each case, the first step involves amplification of 16S rDNA using the PCR. The amplified rDNA can either be sequenced directly, following a suitable DNA purification step (Böttger, 1989; Embley, 1991), or cloned into a vector followed by sequencing the resultant recombinant vector (Stackebrandt & Liesack, 1993). The main differences between these two procedures are outlined below:

(*i*) The cloning method involves an extra step, namely cloning.

(*ii*) The cloning method provides better sequencing templates since either circular DNA (for the plasmid vector) or single-strand DNA (for the bacteriophage vector) are more suitable for the Sanger sequencing procedure than the PCR products which are linear and double-stranded.

(*iii*) The direct sequencing method determines the nucleotide sequences of mixtures of all of the rRNA operons present in genomes whereas only one operon is sequenced with the cloning method thereby avoiding the generation of chimeric nucleotides.

However, despite these differences the two procedures are equally suitable for phylogenetic analyses given the limited heterogeneity found between rRNA operons. Detailed procedures for 16S rRNA sequencing are available (Ludwig, 1991; Stackebrandt & Liesack, 1993).

Reliability of 16S rRNA sequencing. It is well known that 16S rRNA sequences determined using the reverse transcriptase method include errors due to the strong secondary structure of the rRNA template (Lane *et al.*, 1988) and the presence of non-dideoxynucleotide-terminated elongation products (DeBorde *et al.*, 1986). It is possible to get a rough estimate of sequencing errors generated using the reverse transcriptase method as corresponding nucleotide sequences derived from the application of the more accurate DNA sequencing procedures are available for several actinomycetes (Table 2-1).

Representatives of the genus *Mycobacterium* have been the subject of three major 16S rRNA sequencing studies (Stahl & Urbance, 1990; Rogall *et al.*, 1990; Pitulle *et al.*, 1992). Stahl and Urbance (1990) used the reverse transcriptase

TABLE 2-1. COMPARISON OF 16S rRNA SEQUENCES OBTAINED BY USING THE REVERSE TRANSCRIPTASE AND DNA SEQUENCING METHODS.

Strain[a]	Reverse transcriptase sequencing method		DNA sequencing method		% Dissimilarity[b]
	Accession number	Reference	Accession number	Reference	
Arthrobacter globiformis	M23411	Woese (unpublished)	X80736	Koch *et al.* (1994)	0 (0/1464)
Corynebacterium renale	M29553	Stahl & Urbance (1990)	X84249	Pascual *et al.* (unpublished)	0 (0/1329)
Gordona bronchialis	X53201	Stackebrandt *et al.* (1988b)	X79287	Klatte *et al.* (1994c)	1.51 (12/794)
Gordona terrae	X53202	Stackebrandt *et al.* (1988b)	X79286	Klatte *et al.* (1994c)	0.3 (4/1320)
Mycobacterium asiaticum	M29556	Stahl & Urbance (1990)	X55604	Pitulle *et al.* (1992)	0.22 (3/1353)
Mycobacterium aurum	M29558	Stahl & Urbance (1990)	X55595	Pitulle *et al.* (1992)	0 (0/1323)
Mycobacterium chitae	M29560	Stahl & Urbance (1990)	X55603	Pitulle *et al.* (1992)	12.75[c] (169/1326)
Mycobacterium flavescens	M29561	Stahl & Urbance (1990)	X52932	Rogall *et al.* (1990)	0.23 (3/1302)
Mycobacterium gordonae	M29563	Stahl & Urbance (1990)	X52923	Rogall *et al.* (1990)	0 (0/1316)
Mycobacterium nonchromogenicum	M29565	Stahl & Urbance (1990)	X52928	Rogall *et al.* (1990)	0 (0/1318)
Mycobacterium terrae	M29568	Stahl & Urbance (1990)	X52925	Rogall *et al.* (1990)	0 (0/1315)
Mycobacterium thermoresistible	M29570	Stahl & Urbance (1990)	X55602	Pitulle *et al.* (1992)	0.45 (6/1343)
Nocardia asteroides DSM43005	X53205	Stackebrandt *et al.* (1988b)	X57949	Rogall *et al.* (1990)	1.65 (22/1332)
Nocardia otitidiscaviarum	M59056	Yang & Woese, (unpublished)	M59056	This study	0.07 (1/1417)
Saccharo-monospora viridis	X54286	Embley *et al.* (1988b)	Z38007	Kim *et al.* (1995)	2.52 (34/1348)
Tsukamurella paurometabola	X53206	Collins *et al.* (1988a)	Z46751	This study	0.69 (9/1305)
Tsukamurella paurometabola NCTC 10741	X53207	Collins *et al.* (1988a)	Z36933	This study	1.28 (16/1251)

[a], Type strains were used unless strain numbers are given; [b], the numbers of differences and total number of nucleotides are given in parentheses; and [c], the sequence reported by Stahl and Urbance (1990) was of a mislabelled strain which was later found to show a 16S rRNA sequence similarity of 99.78% with *Propionibacterium acnes* (accession number M61903).

method, Rogall *et al.* (1990) directly determined PCR-amplified rDNA and Pitulle *et al.* (1992) sequenced cDNA derived from 16S rRNA using the reverse transcriptase technique. Five out of nine 16S rRNA sequences determined by Stahl and Urbance (1990) were identical to those obtained using the DNA sequencing method (Rogall *et al.*, 1990; Pitulle *et al.*, 1992).

Three out of the four remaining common rRNA sequences exhibited low nucleotide sequence dissimilarities which ranged from 0.22 to 0.45% (Table 2-1). If, for the sake of argument, the nucleotide sequences determined using the DNA sequencing methods (Rogall *et al.*, 1990; Pitulle *et al.*, 1992) are taken to be error free then the mean experimental error of the direct sequencing method, based on the reverse transcriptase technique in the study by Stahl and Urbance (1990), was 0.13% with the greatest error set at 0.45%. It is now evident that the strain which Stahl and Urbance (1990) considered to be the type strain of *Mycobacterium chitae* (accession number M29560) was mislabelled since the primary structure of the 16S rRNA of this strain is almost identical to that of the type strain of *Propionibacterium acnes* (99.78% similarity)!

Additional cases of duplicated 16S rRNA sequences can be found in the GenBank/EMBL nucleotide sequence database (Table 2-1). The 16S rRNA sequences of the type strains of *Arthrobacter globiformis* and *Nocardia otitidiscaviarum* determined by Woese (unpublished data) and Yang and Woese (unpublished data) show good congruence, namely 0% and 0.07% error, respectively. Higher error rates are evident when 16S rRNA sequences determined by Collins *et al.* (1988a), Embley *et al.* (1988b) and Stackebrandt *et al.* (1988b) are considered; the error in these latter studies ranges from 0.3 to 2.52% (Table 2-1). The error level of over 1.5%, evident in corresponding studies on *Gordona bronchialis*, *Nocardia asteroides* (DSM43005) and *Saccharomonospora viridis*

strains, is serious, especially if the 97% sequence similarity guideline is taken to define genomic species (Stackebrandt & Goebel, 1994).

The quality of 16S rRNA sequence data deduced from its encoding gene can also be assessed by comparing nucleotide sequences in the public domain, that is, in the GenBank/EMBL database, since many workers have independently sequenced the same strain, typically type strains (Table 2-2). It is evident from the presented information that 16S rDNA sequence discrepancies range from 0% (*Gordona amarae* and *Rhodococcus equi*) to 0.98% (*Corynebacterium kutscheri*). Particularly high levels of disagreement are apparent between nucleotide sequences generated by Takahashi *et al.* (unpublished) and Pascual *et al.* (unpublished). However, it is not clear from these studies which investigators are responsible for the high experimental errors. In general, it can be said that experimental errors in 16S rDNA sequencing studies are relatively low compared with those found using other taxonomic methods, including the numerical taxonomic procedure.

The source of nucleotide sequencing errors may be in the initial PCR and/or in the subsequent DNA sequencing steps. Comparisons of nucleotide sequences of cloned 16S rDNAs from the same organism show that the amplification errors in the polymerase chain reaction are less than 0.02% per 30 to 35 amplification cycles (Liesack *et al.*, 1991; Weisburg *et al.*, 1991). Most of the discrepancies in the PCR step have been attributed to transitional substitutions (Dunning *et al.*, 1988). Since errors caused by PCR are negligible the main cause of experimental error resides in the DNA sequencing step. Several different sequencing procedures are available to determine 16S rDNA sequences but comparative studies on their reliability have yet to be undertaken (Stackebrandt & Liesack, 1993).

TABLE 2-2. SEQUENCING ERRORS FOUND IN ENTRIES TO THE GENBANK/EMBL DATABASE [a].

Strain[b]	Sequence 1		Sequence 2		% Dissimilarity[c]
	Accession number	Reference	Accession number	Reference	
Corynebacterium cystidis	D37914	Takahashi *et al.* (unpublished)	X84252	Pascual *et al.* (unpublished)	0.70 (10/1426)
Corynebacterium glutamicum	X80629	Rainey *et al.* (1995a)	X84257	Pascual *et al.* (unpublished)	0.43 (6/1388)
Corynebacterium kutscheri	D37802	Takahashi *et al.* (unpublished)	X81871	Pascual *et al.* (unpublished)	0.98 (13/1319)
Corynebacterium pilosum	D37915	Takahashi *et al.* (unpublished)	X84246	Pascual *et al.* (unpublished)	0.86 (11/1278)
Gordona amarae	X80635	Klatte *et al.* (1994c)	X80601	Ruimy *et al.* (1994b)	0 (0/1390)
Gordona bronchialis	X79287	Klatte *et al.* (1994c)	X75903	Stubbs & Collins (unpublished)	0.83 (11/1329)
Mycobacterium chlorophenolicum	X79292	Rainey *et al.* (1995a)	X79094	Briglia *et al.* (1994)	0.41 (6/1466)
Rhodococcus equi	X80614	Rainey *et al.* (1995a)	X80603	Ruimy *et al.* (1994b)	0 (0/1395)
Rhodococcus globerulus	X80619	Rainey *et al.* (1995a)	X77779	Asturias *et al.* (1994)	0.23 (3/1323)
Tsukamurella paurometabola	X80628	Rainey *et al.* (1995a)	Z46751	This study	0.27 (4/1474)

[a], The comparable sequences were determined by DNA sequencing methods.

[b], Type strain.

[c], The number of differences and the total number of nucleotides are given in parentheses.

1.4. Species concept and place of 16S rRNA sequencing in bacterial systematics.

The bacterial species concept is still a difficult issue in microbial systematics (Goodfellow & O'Donnell, 1993; Claridge & Boddy, 1994; O'Donnell *et al.*, 1994). Early definitions of bacterial species were mostly based on monothetic groups, described using a subjectively selected set of phenotypic characters. This phenotypic species concept has several limitations, notably the fact that strains which vary in key characters cannot be accommodated. In addition, monothetic classifications often lack uniformity in the sense that different taxonomic criteria are frequently used to delineate species within diverse genera (Goodfellow & O'Donnell, 1993; O'Donnell *et al.*, 1994).

A universal species concept is still awaited in Bacteriology. Nevertheless, it is important to remember that the bacteria species is unique in the sense that it is the only taxonomic rank, apart from infraspecific categories, which corresponds to a taxon that ***'exists'*** in a population sense. It is often useful to distinguish a ***taxospecies***, a group of strains that share a high proportion of similar properties (Sneath, 1989a); from a ***genospecies***, a group of strains capable of genetic exchange (Ravin, 1961); from a ***genomic species***, a group of strains which share high DNA homology values (Wayne *et al.*, 1987). In some circumstances it is convenient to maintain the species name of a taxon so that it can be distinguished from other taxa whatever its validity on other grounds. Such poorly circumscribed species are known as ***nomenspecies*** (Sneath, 1989a).

The subjective nature of the traditional species concept was recognised by Cowan (1978), who defined a species as:

> *"... a group of organisms defined more or less subjectively by criteria, chosen by the taxonomist to show to best advantage and as far as possible put into practice his individual concept of what a species is."*

However, more objective definitions based on well defined and universally applicable criteria have been proposed, such as estimations of genetic relatedness between organisms by measuring the degree of homology between their genomes. This DNA relatedness approach is the basis of the ***phylogenetic species*** concept introduced by Wayne *et al.* (1987):

> *" The phylogenetic definition of a species generally would include strains with approximately 70% or greater DNA-DNA relatedness and with 5°C or less ΔT_m. Both values must be considered."*

DNA-DNA relatedness values are often seen as the gold standard for the circumscription of bacterial species (Wayne *et al.*, 1987; Goodfellow & O'Donnell, 1993). However, DNA relatedness data need to be evaluated in light of results derived from the application of independent taxonomic techniques. Indeed, it is now recommended that novel taxa be described using both genotypic and phenotypic properties (Wayne *et al.*, 1987; Murray *et al.*, 1990). These workers also recommended that distinct genomic species which cannot be separated from one another on the basis of phenotypic properties should not be named until they can be distinguished by a number of phenotypic properties.

The major drawback of DNA-DNA pairing methods involves the cost and effort of securing a complete matrix of relatedness values between all pairs of test strains (Goodfellow & O'Donnell, 1993). The usual practice is to employ only a few organisms as reference strains and to compare all other strains against this restricted set. However, such a process may result in the loss of information on the underlying taxonomic structure (Sneath, 1983; Hartford & Sneath, 1988).

A stable and lasting classification of novel species is possible when test organisms are compared with type strains of all relevant validly described species using the DNA-DNA pairing procedure. This comprehensive approach is desirable

(e.g., Labeda, 1995) but, in practice, comparative studies are usually restricted to an examination of type strains of species established using phenotypic properties (e.g., Kudo *et al.*, 1988; Eguchi *et al.*, 1993). Comparisons with type strains of all validly described species are not practical in the case of taxa, notably the genera *Bacillus*, *Mycobacterium* and *Streptomyces*, which encompass many species.

Data derived from 16S rRNA sequencing studies are increasingly being used to define bacterial species (Stackebrandt, 1992; Ludwig & Schleifer, 1994). Two assumptions underlie this approach, namely that lateral gene transfer has not occurred between 16S rRNA genes and that the amount of evolution or dissimilarity between 16S rRNA sequences of given pairs of organisms is representative of the variation shown by the corresponding genomes. Woese *et al.* (1980) considered that lateral gene transfer between 16S rRNA genes was probably very rare since this gene is responsible for maintaining functional and tertiary structural consistency and is rich in information. The two assumptions form the basis of a reasonable hypothesis though horizontal gene transfer in 16S rDNA may have been recognised between members of the genus *Aeromonas* (Sneath, 1993).

Fox *et al.* (1992) were the first to realise that 16S rRNA molecules of members of closely related species may not always show molecular clock behaviour since they are so conserved. This important observation meant that strains of related species with almost identical 16S rRNA sequences may belong to different genomic species. The relationship between 16S rRNA sequence and DNA relatedness data is illustrated in Figure 2-1 using data derived from studies on representatives of mycolic acid-containing taxa; the raw data used to plot Figure 2-1 are presented in Appendix E.

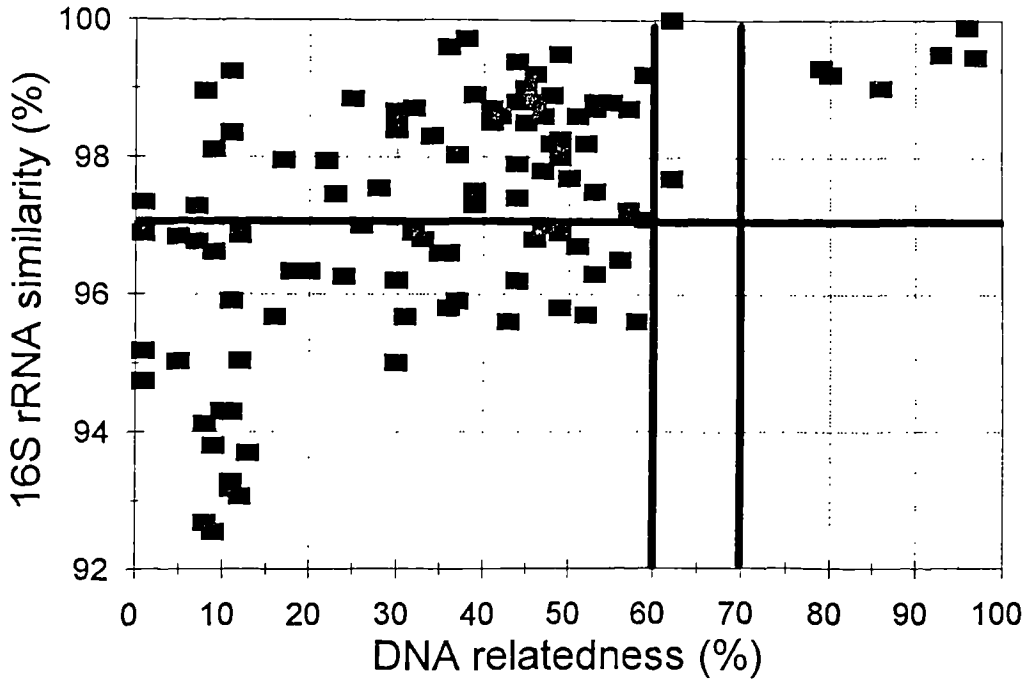


Figure 2-1. Comparison of 16S rRNA sequence similarity and DNA relatedness values. Based on data taken from studies on mycolic acid-containing actinomycetes (the original data and corresponding references are given in Appendix E).

It is evident from Figure 2-1 that a comparison of the 16S rRNA sequence and DNA relatedness data does not demonstrate a linear relationship. However, 16S rRNA sequence similarity values below 97% always correspond to DNA relatedness values below 60% though the reverse is not true. A similar relationship between 16S rRNA sequence and DNA relatedness data was presented by Stackebrandt and Goebel (1994) who argued that organisms assigned to a genomic species could be expected to show a 97% or more similarity value in light of corresponding 16S rRNA sequence data. This cut-off value is plausible given the results shown in Figure 2-1 for mycolic-acid containing actinomycetes.

DNA sequencing studies are more cost-effective and less laborious than DNA-DNA reassociation studies given developments in molecular biology, notably the use of the PCR and the introduction of automatic DNA sequencers (Stackebrandt & Goebel, 1994). However, the two methods should be seen as complementary for the circumscription of bacterial species. 16S rDNA sequencing studies are useful for underpinning taxonomic relationships at species or higher taxonomic ranks whereas DNA relatedness studies are only valid at the species or subspecific level (Goodfellow & O'Donnell, 1993). When closely related organisms are compared genomic species cannot be recognised only by using 16S rDNA sequence data. In such cases DNA-DNA pairing studies need to be carried out.

An useful compromise is to involve both molecular methods to delineate closely related species. Initial 16S rRNA sequencing studies can be undertaken to show the broad relationships between closely related organisms with DNA-DNA pairing studies used to detect the finer taxonomic structure. Deoxyribonucleic acid relatedness values can be determined using two well established procedures, namely by using filter-binding and spectrophotometric techniques (Huss *et al.*, 1983; Johnson, 1991). The spectrophotometric method is preferred as it gives more

reproducible results and eliminates the need for the use of isotopes (Huss *et al.*, 1983; Hartford & Sneath, 1993). The sequential use of the 16S rDNA sequencing and DNA-DNA pairing methods should prove to be of particular value in unravelling the taxonomic structure of genera, such as *Bacillus*, and *Streptomyces*, which encompass many species.

2. Chemosystematics

2.1. Introduction

Chemosystematics or chemotaxonomy is the study of the distribution of chemical components, such as cell wall amino acids, lipids, proteins, isoprenoid quinones and sugars, amongst members of microbial taxa and the use of such information for classification and identification (Goodfellow & O'Donnell, 1994). This definition can be extended to include information derived from whole-organism chemical fingerprinting techniques, such as pyrolytic (pyrolysis gas chromatography, pyrolysis mass spectrometry and pyrolysis tandem mass spectrometry) and spectroscopic methods (infrared and ultraviolet-resonance Raman spectroscopy; Helm *et al.*, 1991; Magee, 1993, 1994; Naumann *et al.*, 1994).

Amino acids, lipids, proteins and sugars are important structural constituents of bacterial cell envelopes. Chemical data derived from analyses of cell components can be used to classify bacteria at different taxonomic ranks according to the pattern of distribution of various chemical markers within and between members of different taxa. Chemotaxonomic analyses of chemical macromolecules, particularly amino acids and peptides (e.g., from peptidoglycans and pseudomureins), lipids (e.g., fatty acids, lipopolysaccharides, mycolic acids and polar lipids), polysaccharides and related polymers (e.g., wall sugars, methanochondroitins, teichoic and teichuronic acids), proteins (e.g., bacteriochlorophylls, cytochromes,

fimbriae, glycoproteins, protein sheaths and whole-organism protein patterns) and enzymes (e.g., hydrolases, lyases, oxidoreductases) and other complex polymeric compounds, such as isoprenoid quinones and sterols, all provide grist to the taxonomic mill. The base composition of DNA is also a chemical property *sensu stricto* though it is usually considered with other information derived from analyses of nucleic acids. Similarly, data derived from enzyme tests are usually dealt with as phenotypic characters (Manafi *et al.*, 1991; Goodfellow & James, 1994).

The taxonomic value of different types of chemical markers in bacterial systematics has been considered in detail (Goodfellow & O'Donnell, 1994; Suzuki *et al.*, 1993). Consequently, this section is restricted to a consideration of chemotaxonomic markers that have been shown to be of particular value for the classification and identification of actinomycetes.

2.2. Specific chemical markers

The most commonly used chemical characters in actinomycete systematics are cellular fatty acids, menaquinones, muramic acid types, phospholipids, whole-organism amino acids and sugars and the base composition of DNA (Goodfellow, 1989a; Williams *et al.*, 1989). Some of the methods used to detect these chemotaxonomic markers can be used to provide quantitative or semi-quantitative data, as in the case of DNA base composition, cellular fatty acid and menaquinone analyses, but other methods yield qualitative data, as in the case of muramic acid, peptidoglycan, phospholipid and whole-organism sugar determinations.

DNA base composition. Actinomycetes form a distinct phyletic branch in the Gram-positive line of descent based on the results of 16S rRNA sequence analyses (Stackebrandt & Woese, 1981b; Woese, 1987; Embley & Stackebrandt, 1994).

Most actinomycetes contain DNA within the range of 50 (*Arcanobacterium* and *Mobiluncus* spp.) to approximately 75 mol% G+C (*Geodermatophilus* spp.). DNA base composition values have proved to be of particular value in the assignment of organisms to genera (Tamaoka, 1994).

Cellular fatty acids. Fatty acids can be defined as carboxylic acid derivatives of long-chain aliphatic molecules. In bacteria, they range in chain length from two (C_2) to over ninety (C_{90}) carbon atoms. Fatty acids in the range C_{10} to C_{24} are of the greatest taxonomic value (Suzuki *et al.*, 1993). Cellular fatty acid composition is usually determined by gas-liquid-chromatography with the resultant data suitable for multivariate statistical analyses, notably using principal component and SIMCA routines (O'Donnell, 1985; Saddler *et al.*, 1987).

Actinomycetes show three major types of fatty acid profile (Table 2-3; Kroppenstedt, 1985). Members of the same genus usually have the same fatty acid type though a number of exceptions exist (e.g., *Corynebacterium bovis* contains 10-methyl-C_{18}). Quantitative analyses of cellular fatty acids are increasingly being used to compare members of closely related species (e.g., Kroppenstedt *et al.*, 1990; Korn-Wendisch *et al.*, 1995).

Mycolic acids are especially long chain 3-hydroxy fatty acids with an alkyl branch at position 2. These molecules are only present in members of the genera *Corynebacterium*, *Dietzia*, *Gordona*, *Mycobacterium*, *Nocardia*, *Rhodococcus* and *Tsukamurella* (Goodfellow, 1992; Rainey *et al.*, 1995c) The variations in chain length and the degree of saturation are especially useful for the classification of members of mycolic acid-containing taxa (Goodfellow & Lechevalier, 1989; Goodfellow, 1992).

TABLE 2-3. FATTY ACID PATTERNS OF ACTINOMYCETES ACCORDING TO THE CLASSIFICATION OF KROPPENSTEDT (1985).

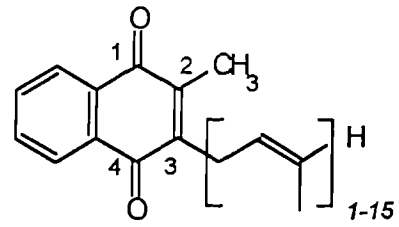
Fatty acid type	Representative taxa	Fatty acid composition*:								
		Straight-chain		*Iso-*	*Iso-*	*Anteiso-*	10-Methyl-		Cyclo-	2-Hydroxy-
		Saturated	Unsaturated	14/16/18	15/17	15/17	17	18	propane	
1a	*Corynebacterium, Dermatophilus*	+++	+++	-	-	-	-	-	-	-
1b	*Dietzia, Gordona, Mycobacterium, Nocardia, Rhodococcus, Tsukamurella*	+++	+++	-	-	-	-	++	-	-
1c	*Actinomyces israelli*	+++	+++	-	-	-	-	-	++	-
2a	*Saccharomonospora*	++	+	+++	+	(+)	-	-	-	+
2b	*Geodermatophilus, Intrasporangium*	(+)	+	++	+++	+	-	-	-	-
2c	*Streptomyces*	+	(v)	+++	+	+++	-	-	-	-
2d	*Ampullariella*	+	+	+++	+++	+++	-	-	-	-
3a	*Nocardioides*	+++	++	+++	(+)	(+)	(+)	+++	-	+
3b	*Micromonospora*	+	+	+++	+++	++	++	(+)	-	-
3c	*Microbispora*	+	+	++	+	+	+++	(+)	-	-
3d	*Nocardiopsis*	+	+	+++	++	+++	(+)	+++	-	-

*, positive; -, negative; v, variable; ( ), less than 5% of the total fatty acids.

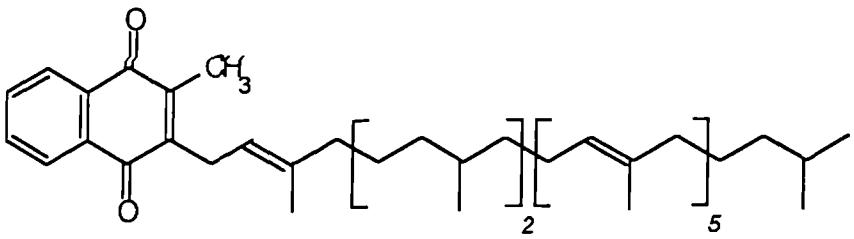
Menaquinones. Several types of isoprenoid quinone have been detected in bacteria (Collins, 1994) but menaquinones (MK) are the most common type found in actinomycete envelopes (Kroppenstedt, 1985; Suzuki *et al.*, 1993; Collins, 1994). These compounds have a chemical structure analogous to that of vitamin K_2 (MK-7; unsaturated menaquinone with seven isoprene units) and are classified according to the number of isoprene units, which can vary from one up to fifteen, and the degree of saturation or hydrogenation. The structure and composition of bacterial menaquinones are determined either semi-quantitatively by mass spectrometry or quantitatively by high-performance liquid chromatography (Kroppenstedt, 1985; Collins, 1994). The typical structures of actinomycete menaquinones are illustrated in Figure 2-2.

The position or point of hydrogenation in isoprenoid side-chains can be very specific and hence of taxonomic value (Collins, 1994). Sophisticated techniques, such as silver-phase high-performance liquid chromatography (Kroppenstedt, 1985) and tandem mass spectrometry (Collins *et al.*, 1988c; Ramsey *et al.*, 1988) are needed to determine the points of hydrogenation in isoprene units. Such studies have provided valuable information for the classification of *Actinomadura* (MK9[II,III,VIII-H_6]), *Microtetraspora* (MK9[II,VIII,IX-H_6]) and *Streptomyces* strains (MK9[II,III,IX-H_6]) [Yamada *et al.*, 1982; Collins *et al.*, 1988c; Kroppenstedt *et al.*, 1990].

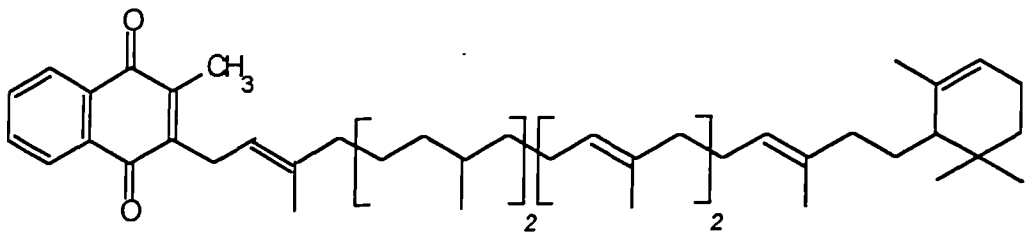
An unique hexahydrogenated menaquinone with eight isoprenoid units in which the end two units were cyclised was discovered in *Nocardia brasiliensis* by Howarth *et al.* (1986); a dihydrogenated isomer was subsequently found as a minor component (Collins *et al.*, 1987). These cyclic menaquinones are characteristic of



(a) Unsaturated menaquinone



(b) MK-9 (*II*, *III*, *IX*-H_6)



(c) MK-8(*II*, *III*-H_4) cyclic

FIGURE 2-2. STRUCTURES OF (a) UNSATURATED MENAQUINONES CHARACTERISTIC OF *TSUKAMURELLA* STRAINS, (b) PARTIALLY SATURATED MENAQUINONES CHARACTERISTIC OF *STREPTOMYCES* STRAINS AND (c) CYCLIC MENAQUINONES CHARACTERISTIC OF *NOCARDIA* STRAINS.

members of the genus *Nocardia sensu stricto* (Kämpfer *et al.*, 1990; Goodfellow 1992) though fully saturated cyclic menaquinones have also been reported for the archaebacterium, *Pyrobaculum organotrophum* (Tindall *et al.*, 1991).

Peptidoglycan. Peptidoglycans provide the basic structure of the bacterial cell wall. Although there is considerable inter-species variation in the detailed structure of peptidoglycan, its chemical architecture remains constant, that is, it consists of β1-4 linked disaccharides of N-acetylglucosamine and N-acetylmuramic acid (Figure 2-3). These glycan chains (up to 100 units) are covalently cross-linked by oligopeptides which connect 3-O-lactoyl groups of muramic acid residues in glycan chains (Hancock, 1994). Structural variation in peptidoglycans mainly occur in the position and type of diamino acid involved in the cross-linkage of the peptide chains at position 3, and on the presence and composition of interpeptide bridges. Peptidoglycans are usually classified according to the system proposed by Schleifer and Kandler (1972; Table 2-4).

Most actinomycetes have the type A peptidoglycan though members of the family *Microbacteriaceae* exhibit the distinctive type B form (Schleifer & Kandler, 1972). 5S and 16S rRNA sequence data have shown that members of the family *Microbacteriaceae* form a distinct phyletic line within the evolutionary radiation encompassed by actinomycetes (Park *et al.*, 1993; Rainey *et al.*, 1994b). Variation in peptidoglycans provides useful information for the classification of actinomycetes at and above the genus level. The muramic acids of actinomycete peptidoglycans can be either N-acetylated (A) or N-glycolated (G) [Uchida & Aida, 1977, 1984]. Most actinomycetes contain N-acetylated muramic acids though members of the families *Micromonosporaceae*, *Mycobacteriaceae* and *Nocardiaceae* are characterised by the presence of N-glycolated muramic acid residues.

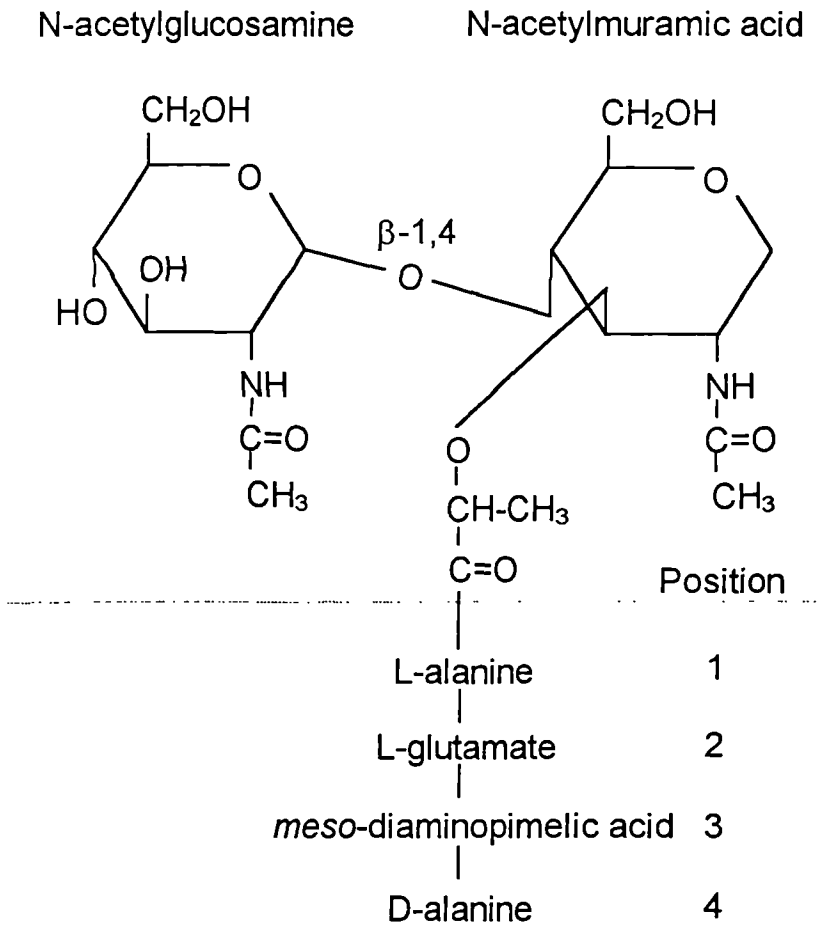


FIGURE 2-3. FRAGMENTS OF THE PRIMARY STRUCTURE OF A TYPICAL PEPTIDOGLYCAN.

TABLE 2-4. CLASSIFICATION OF PEPTIDOGLYCANS *SENSU* SCHLEIFER AND KANDLER (1972)[a].

Position of cross-link	Peptide bridge		Amino acid at position 3
A: Cross-linkage between positions 3 and 4 of two peptide subunits[b]	1. None	α	L-lysine
		β	L-ornithine
		γ	*meso*-diaminopimelic acid
	2. Polymerised units	α	L-lysine
	3. Monocarboxylic L-amino acids or glycine or both	α	L-lysine
		β	L-ornithine
		γ	LL-diaminopimelic acid
	4. Contains a dicarboxylic amino acid	α	L-lysine
		β	L-ornithine
		γ	*meso*-diaminopimelic acid
		δ	L-diaminobutyric acid
	5. Contains a dicarboxylic amino acid and lysine	α	L-lysine
		β	L-ornithine
B: Cross-linkage between positions 2 and 4 of two peptide subunits	1. Contains a L-amino acid	α	L-lysine
		β	L-homoserine
		γ	L-glutamic acid
		δ	L-alanine
	2. Contains a D-amino acid	α	L-ornithine
		β	L-homoserine
		γ	L-diaminobutyric acid

[a], Modified from Rogers *et al.* (1980), Schleifer and Kandler (1972) and Schleifer and Seidl (1985). [b], The variation is marked by a prime in cases where alanine is replaced by glycine at position 1.

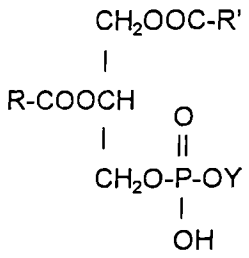
Phospholipids. Phospholipids are the most common polar lipids found in bacterial cytoplasmic membranes (Suzuki *et al.*, 1993). The structural variations found in bacterial phospholipids are illustrated in Table 2-5. The different types of phospholipids are discontinuously distributed in actinomycetes and hence provide useful taxonomic information. Acylated ornithine and polar, amphipatic glycolipids or lysine amides, can be also considered as polar lipids (Minnikin & O'Donnell, 1984).

Lechevalier *et al.* (1977, 1981) classified actinomycetes into five phospholipid groups based on *'semi-quantitative'* analyses of major phospholipid markers found in whole-organism extracts. Phospholipids extracted from actinomycetes by using organic solvent systems (Minnikin *et al.*, 1984) can be separated by two dimensional thin-layer-chromatography (e.g., Embley *et al.*, 1983) and detected using non-specific (5%, w/v, ethanolic molybdophosphoric acid; Suzuki *et al.*, 1993) or specific spray reagents. The latter can be used to detect α-glycols (periodate-Schiff; Shaw, 1968), amino groups (0.2% ninhydrin, w/v, in water-saturated butanol; Consden & Gordon, 1948), choline (Dragendorff reagent; Wagner *et al.*, 1961), lipid phosphates (Dittmer & Lester, 1964) and sugars (α-naphthol; Jacin & Mishkin, 1965). Actinomycetes can be assigned to established phospholipid patterns according to the presence or absence of specific or combinations of specific phospholipid markers (Table 2-6). In general, members of the same actinomycete genus have the same phospholipid type. Phospholipid patterns can be important for the recognition of actinomycete genera (Goodfellow, 1989a; Williams *et al.*, 1989). Proposals for the recognition of *Aeromicrobium* (Tamura & Yokota, 1994) and *Dietzia* (Rainey *et al.*, 1995c) were partly based on polar lipid data.

Sugar composition. Neutral sugars, which are major components of actinomycete cell envelopes, are useful taxonomic markers at the suprageneric level. Sugar composition can be determined by simple paper chromatography (Schaal, 1985) or by using gas-liquid chromatography (Saddler *et al.*, 1991) following full hydrolysis of purified cell walls or whole-organisms (Hancock, 1994). In the latter case, quantitative sugar profiles can be used for multivariate statistical analyses (St-Laurent *et al.*, 1987). The source of the sugars is largely unknown (Hancock, 1994) with only those of mycobacteria having been studied in some detail (Besra & Chatterjee, 1994). In the mycobacterial cell wall, sugars are present in various polysaccharide polymers, including arabinogalactan, lipoarabinomannan, lipomannan and phosphatidylinositol mannosides (Daffé *et al.*, 1990; Chatterjee *et al.*, 1992). Information derived from analyses of sugar composition provides fairly crude data for understanding cell wall structure as the presence of a particular sugar in different strains does not necessarily mean that it is derived from the same macromolecule. More detailed analyses of cell wall polysaccharide polymers, such as so-called linkage analyses, can be used to provide additional information on the structure and function of actinomycete envelopes (Daffé *et al.*, 1993).

Actinomycetes can be assigned to five groups on the basis of the discontinuous distribution of major diagnostic sugars, namely: group **A**, arabinose plus galactose; group **B**, madurose (3-O-methyl-D-galactose); **C**, no diagnostic sugars, **D**, arabinose plus xylose; and **E**, galactose plus rhamnose (Lechevalier & Lechevalier, 1970; Labeda *et al.*, 1984). A few "*rare*" sugars have also been reported to be diagnostic for members of some actinomycete taxa, notably the occurrence of 3-O-methyl-rhamnose in *Catellatospora* (Asano *et al.*, 1989) and tyvelose in *Agromyces* (Maltsev *et al.*, 1992).

TABLE 2-5. MAJOR PHOSPHOLIPIDS FOUND IN ACTINOMYCETES.



R, R' = long-chain alkyl

Overall charge	Polar head group substituent (Y)	Name and abbreviation
1+	Glycerol	Phosphatidylglycerol
2+	Phosphatidylglycerol	Diphosphatidylglycerol
1+	Butane-2,3-diol	Phosphatidylbutanediol
1+	Inositol	Phosphatidylinositol
1+	Acylated mannosylinositols	Phosphatidylinositol mannosides
0	Ethanolamine	Phosphatidylethanolamine
0	Choline	Phosphatidylcholine
0	Methylethanolamine	Phosphatidylmethylethanolamine

Modified from Minnikin and O'Donnell (1984).

TABLE 2-6. CLASSIFICATION OF ACTINOMYCETE PHOSPHOLIPIDS *SENSU* LECHEVALIER *et al.* (1977, 1981).

Type	Characteristics*
I	Nitrogenous phospholipids absent, phosphatidylglycerol variable
II	Only phosphatidylethanolamine
III	Phosphatidylcholine, phosphatidylethanolamine, phosphatidylmethylethanolamine and phosphatidylglycerol variable; phospholipids containing glucosamine absent
IV	Phospholipids containing glucosamine with phosphatidylethanolamine and phosphatidylmethylethanolamine variable
V	Phospholipids containing glucosamine, phosphatidylglycerol with phosphatidylethanolamine variable

*, Most actinomycetes contain phosphatidylinositol (Suzuki *et al.*, 1993); this component is absent in *Dietzia* strains (Rainey *et al.*, 1995c).

Wall chemotypes. The widespread use of simple chemical methods to highlight diagnostic chemical markers proved to be of particular value in actinomycete systematics. The introduction of wall chemotypes by Lechevalier and Lechevalier (1970) provided a much needed practical way of assigning actinomycetes to a number of groups using qualitative chemical data. The system is based on the discontinuous distribution of major diagnostic amino acids and sugars in whole-organism hydrolysates (Table 2-7). Members of most actinomycete genera and families usually have the same wall chemotypes.

3. The order Actinomycetales

3.1. The order *Actinomycetales* Buchanan 1917[AL]

The order *Actinomycetales*, which is the only member of the class *Actinomycetes* Krassilnikov 1949[AL], currently encompasses over ninety validly described genera. The order, first introduced to accommodate members of the family *Actinomycetaceae,* was designed mainly on morphological and pigmentation properties:

> *"Mold-like organisms, not typically water forms, saprophytic or parasitic. Sheath not impregnated with iron, true hyphae with branching often evident, conidia may be developed, but never endospores. Without granules of free sulphur without bacteriopurpurin. Never producing a pseudoplasmodium. Always non-motile."* (Buchanan, 1918).

The order remained morphological concept for many years. In 1973, Gottlieb considered that actinomycetes encompassed '*varied groups of bacteria whose common feature is the formation of hyphae at some stages of development*' but he went on to say that, in some organisms hyphal formation was tenuous and required imagination to believe in it. Subsequently, Gottlieb (1974) defined members of the order *Actinomycetales* as '*bacteria that tend to form branching filaments which in some families developed into a mycelium*'. However, he conceded that filaments

Table 2-7. Classification of actinomycetes based on wall chemotypes *sensu* Lechevalier and Lechevalier (1970).

Wall chemotype	Diagnostic amino acids and sugars
I	LL-Diaminopimelic acid and glycine
II	*meso*-Diaminopimelic acid and/or hydroxy-diaminopimelic acid with glycine
III	*meso*-Diaminopimelic acid and madurose
IV	*meso*-Diaminopimelic acid, arabinose and galactose
V	Lysine and ornithine
VI	Lysine with aspartic acid and galactose variable
VII	Diaminobutyric acid and glycine with lysine variable
VIII	Ornithine
IX*	LL- and *meso*-diaminopimelic acids

*, Wall chemotype proposed for kitasatosporiae (Wellington *et al.*, 1992).

might be short, as in members of the families *Actinomycetaceae* Buchanan 1918[AL] and *Mycobacteriaceae* Chester 1897[AL], and that in members of certain taxa they underwent fragmentation and consequently could only be observed at some stages in the growth cycle.

The relatively simple morphology of mycobacteria partly explains why these organisms were sometimes omitted from classifications of actinomycetes (Waksman, 1961, 1967). Other workers questioned whether actinomycetes formed a natural group preferring to regard them as a convenient but artificial taxon (Sneath, 1970; Prauser, 1970, 1978, 1981; Goodfellow & Cross, 1974). The difficulty of distinguishing between nocardioform actinomycetes and coryneform bacteria was also widely recognised (Williams *et al.*, 1976; Goodfellow & Minnikin, 1981a,b; Locci, 1981).

The morphological concept of an actinomycete has been challenged by information derived from the application of chemical and molecular taxonomic methods. Data from 16S rRNA cataloguing studies showed that most morphological features were poor markers of phylogenetic relationships and that the traditional morphological definition of an actinomycete could not be sustained (Stackebrandt *et al.*, 1980a, b, 1983; Ludwig *et al.*, 1981; Stackebrandt & Woese, 1981a, b). It is perhaps not too surprising that the morphologically simple corynebacteria were found to have a close evolutionary relationship with the more highly differentiated mycobacteria, nocardiae and rhodococci as this grouping is consistent with the results of chemotaxonomic (Minnikin & Goodfellow, 1980, 1981a), comparative immunodiffusion (Lind & Ridell, 1976) and numerical phenetic studies (Goodfellow & Minnikin, 1981b,c; Goodfellow & Wayne, 1982).

The traditional practice of separating the more highly differentiated actinomycetes from the relatively morphologically simple coryneform bacteria no

longer holds as strains of *Actinomyces*, *Oerskovia* and *Promicromonospora* show a closer phylogenetic affinity to members of the genera *Arthrobacter*, *Brevibacterium*, *Cellulomonas*, *Curtobacterium* and *Microbacterium* than to mycelium-form organisms such as *Nocardia* and *Streptomyces* (Goodfellow & Cross, 1984). In addition, members of the mycelium forming genus *Thermoactinomyces* have been reclassified in the family *Bacillaceae* (Park *et al.*, 1993) whereas *Arthrobacter* and *Micrococcus* strains have been considered to be indistinguishable on the basis of 16S rRNA cataloguing data (Stackebrandt & Woese, 1979, 1981a, b; Stackebrandt *et al.*, 1980a). It is evident from these findings that the possession of branched hyphae should not automatically place a strain with the actinomycetes. The order *Actinomycetales* has yet to be formally redefined in light of recent advances in actinomycete systematics. However, the term, "*actinomycetes*" is now used to refer to Gram-positive bacteria with DNA rich in G+C (over 55mol%).

Stackebrandt and Woese (1981b) considered that Gram-positive bacteria could be divided into two major subgroups which corresponded to *'the high G+C* (i.e., *over 55%) actinomycete-type of organisms and the low G+C* (i.e., *below 50%) endospore-forming organisms and their asporongenous relatives'.* The genera *Actinomyces*, *Bifidobacterium*, *Corynebacterium*, *Mycobacterium*, *Propionibacterium* and *Streptomyces* and related taxa were assigned to the high G+C subgroup ("*actinomycetes*"). However, these workers recommended that the order *Actinomycetales* should not encompass bifidobacteria and propionibacteria as these organisms were anaerobic and only loosely associated with aerobic actinomycetes on the basis of 16S rRNA catalogue data. A similar definition of "actinomycetes" and the order *Actinomycetales* based on 16S rRNA catalogue data was given by Stackebrandt and Schleifer (1984).

Goodfellow and Cross (1984) attempted to redefine actinomycetes in phylogenetic terms as '*Gram-positive bacteria with a high G+C content in their DNA (above 55 mol%) which are phylogenetically related from the evidence of 16S rRNA oligonucleotide sequencing and nucleic acid hybridisation studies (thus excluding the genera Bifidobacterium, Kurthia and Propionibacterium)*'. None of the authors mentioned above considered the genera *Bifidobacterium* and *Propionibacterium* as members of the order *Actinomycetales*. Murray (1992) proposed the class *Thallobacteria* to accommodate '*Gram-positive bacteria showing a branch habit, the actinomycetes and related organisms*' but did not give a precise definition of the taxon.

Complete nucleotide sequences of 16S rRNA have revealed that members of the genus *Propionibacterium* are closely related to the genus *Nocardioides.* Consequently, the genus *Propionibacterium* is now considered to belong to the order *Actinomycetales* (Embley & Stackebrandt, 1994). In contrast, the position of the genus *Bifidobacterium* in the phylogenetic tree based on complete 16S rRNA sequence data is similar to its location in the trees derived from 16S rRNA catalogue data (Embley & Stackebrandt, 1994).

Several Gram-positive bacteria with DNA rich in G+C have been excluded from but are considered to be closely related to the actinomycetes. Embley and Stackebrandt (1994) preferred to restrict the term "*actinomycetes*" to the phyletic line that included bifidobacteria and propionibacteria, and introduced the phrase "*actinomycete line of descent*" to encompass *Coriobacterium glomerans* and *Sphaerobacter thermophilus*. A precise definition of actinomycetes based on the G+C content of DNA is complicated by the close relationship found between *Coriobacterium glomerans* and the genus *Atopobium* (Rainey *et al.*, 1994a).

Members of the genus *Atopobium* contain DNA which falls within the range 35 to 46 mol%. A precise definition of the order *Actinomycetales* is still awaited!

3.2. Suprageneric classification of actinomycetes and related taxa

Actinomycetes. The GenBank/EMBL database presently contains more than three hundred 16S rRNA sequences on actinomycetes. Many of these sequences were added to the database after 1993 by investigators who used DNA sequencing procedures which give more accurate data than the reverse transcriptase technique. It is, therefore, now possible to compare suprageneric relationships found between actinomycetes and their neighbors based on 16S rRNA sequence data since members of most actinomycete genera have been sampled (Figure 2-4; Appendix F). A similar phylogenetic tree was generated by Embley and Stackebrandt (1994). The distribution of some morphological and chemical markers against the suprageneric taxa highlighted in the present study is shown in Tables 2-8 and 2-9.

The tree generated by Embley and Stackebrandt (1994) is largely based on unpublished 16S rRNA sequences whereas the one presented here is derived from nucleotide sequences that are in the public domain. Consequently, members of the genera *Actinomadura*, *Micromonospora*, *Microtetraspora*, *Planomonospora* and *Planobispora* are included in this tree though it does not include information on members of the genera *Arthrobacter* (Koch *et al.*, 1994), *Brevibacterium* (Cai & Collins, 1994), *Micrococcus* (Koch *et al.*, 1994), *Nocardia* (this study), *Rhodococcus* (Rainey *et al.*, 1995a) and *Saccharomonospora* (Kim *et al.*, 1995) or on newly described genera, notably *Thermocrispum* (Kom-Wendisch *et al.*, 1995).

The organisms considered by Embley and Stackebrandt (1994) as authentic actinomycetes are shown in Figure 2-4. These workers were also of the view that the phylogenetic relationships found between most of the different actinomycete

Figure 2-4. An abridged phylogenetic tree showing the actinomycete line of descent based on 1371 16S rRNA nucleotide positions on two hundred and ninety-five strains. The evolutionary distances were calculated according to Jukes and Cantor (1969) and the tree constructed by using the neighbor-joining method (Saitou & Nei, 1987) with *Bacillus subtilis* (K00637) as outgroup. A comprehensive tree is given in Appendix F.

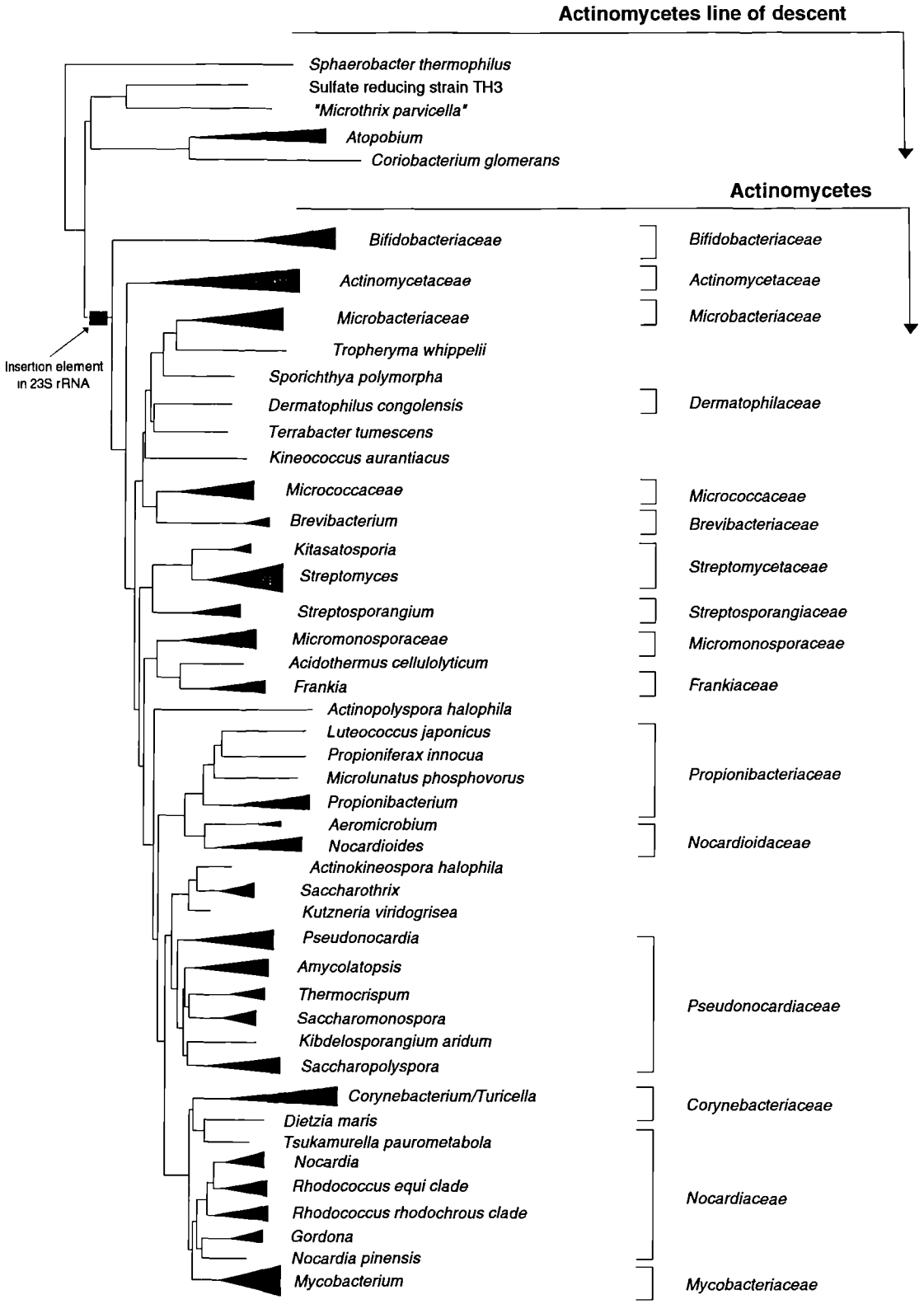

Actinomycetes line of descent
Sphaerobacter thermophilus
Sulfate reducing strain TH3
"Microthrix parvicella"
Atopobium
Coriobacterium glomerans
Actinomycetes
Insertion element in 23S rRNA
Bifidobacteriaceae
Actinomycetaceae
Microbacteriaceae
Tropheryma whippelii
Sporichthya polymorpha
Dermatophilus congolensis
Terrabacter tumescens
Kineococcus aurantiacus
Micrococcaceae
Brevibacterium
Kitasatosporia
Streptomyces
Streptosporangium
Micromonosporaceae
Acidothermus cellulolyticum
Frankia
Actinopolyspora halophila
Luteococcus japonicus
Propioniferax innocua
Microlunatus phosphovorus
Propionibacterium
Aeromicrobium
Nocardioides
Actinokineospora halophila
Saccharothrix
Kutzneria viridogrisea
Pseudonocardia
Amycolatopsis
Thermocrispum
Saccharomonospora
Kibdelosporangium aridum
Saccharopolyspora
Corynebacterium/Turicella
Dietzia maris
Tsukamurella paurometabola
Nocardia
Rhodococcus equi clade
Rhodococcus rhodochrous clade
Gordona
Nocardia pinensis
Mycobacterium
Bifidobacteriaceae
Actinomycetaceae
Microbacteriaceae
Dermatophilaceae
Micrococcaceae
Brevibacteriaceae
Streptomycetaceae
Streptosporangiaceae
Micromonosporaceae
Frankiaceae
Propionibacteriaceae
Nocardioidaceae
Pseudonocardiaceae
Corynebacteriaceae
Nocardiaceae
Mycobacteriaceae

Table 2-8. Distribution of some chemical markers, cultural and morphological properties of actinomycetes and related organisms classified on the basis of 16S rRNA sequence data.

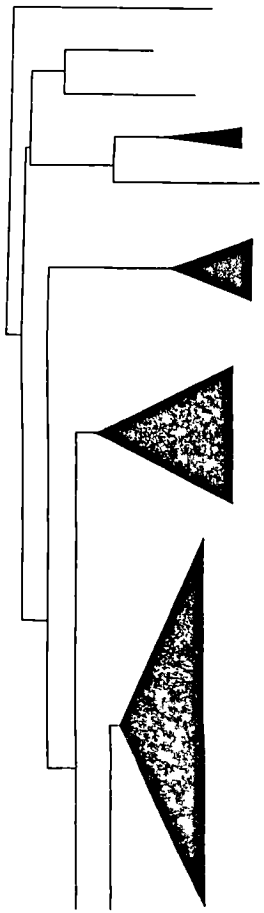
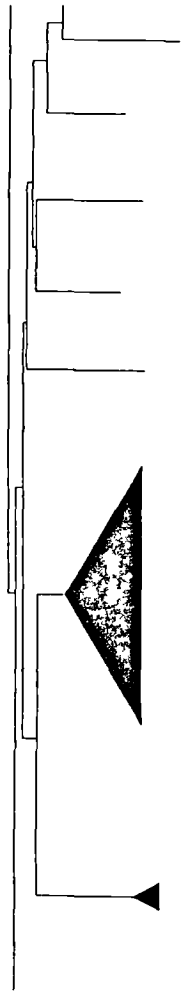
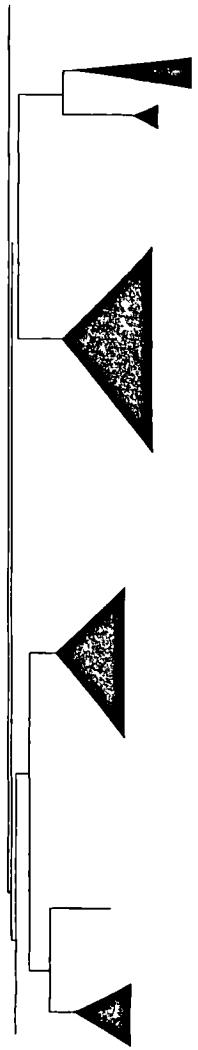
Suprageneric relationships based on 16S rRNA sequence data[a]	Family/genus[b]	Wall chemo-type[c]	Whole-organism sugar pattern[d]	Peptidoglycan[e] and muramic acid[f] types		Fatty acid pattern[g]	Major mena-quinones[h] (MK)	Phospho-lipid type[i]	Mol% G+C of DNA	Relation to oxygen	Morphology
	*Sphaerobacter thermophilus*	ND	ND	A4β	ND	ND	-8	ND	66	aerobe	rods, pleomorphic
	Sulfate reducing strain TH3	ND	ND	ND	ND	ND	ND	ND	ND	aerobe	ND
	"Microthrix parvicella"	ND	ND	ND	ND	ND	ND	ND	ND	aerobe	straight filamentous
	Atopobium	ND	ND	A4β	ND	ND	ND	ND	35-46	anaerobe	rods
	Coriobacterium glomerans	ND	ND	A4α	ND	ND	ND	ND	60-61	anaerobe	ND
	Bifidobacteriaceae										
	Bifidobacterium	VIII	A	A3α,A3β A4α,A4β	ND	ND	ND	ND	57-67	anaerobe	pleomorphic
	Gardnerella vaginalis	ND	ND	ND	ND	ND	ND	ND	42-44	facultative anaerobe	rods, pleomorphic
	Actinomycetaceae										
	Actinomyces	V,VI	–	A4α,A4β	ND	1a,c	-10(H_2,H_4)	PII	57-69	anaerobe	branching rods
	Arcanobacterium haemolyticum	VI	–	A5α	ND	1a	-9(H_4)	ND	48-52	facultative anaerobe	rods, pleomorphic
	Mobiluncus								49-52	anaerobe	curved motile rods
	Microbacteriaceae	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
	Agromyces	VII	–	B2γ	ND	2c	-12	PI	70-72	aerobe	fragmenting hyphae
	Aureobacterium	VIII	–	B2β	ND	2c	-11,-12	PI	67-70	aerobe	rods, pleormophic
	Brevibacterium helvolum	VIII	–	B2γ	ND	ND	ND	ND	ND	ND	ND
	Clavibacter	VII	–	B2γ	ND	2c	-9,-10	PI	68-75	aerobe	short rods (V-shaped)
	"Corynebacterium aquaticum"/ "Corynebacterium mediolanum"	VI	–	B2γ	ND	ND	-10,-11	PI	64-74	ND	ND
	Curtobacterium	VIII	–	B2β		2c	-9	PI	68-75	ND	ND
	Microbacterium	VI	–	B1α,B1β	G	2c	-11,-12	PI	66-70	aerobe	slender rods (V-shape), pleomorphic
	Rathaybacter	VII	ND	B1γ	ND	ND	-10	ND	70-73	aerobe	rods, pleomorphic

TABLE 2-8. CONTINUED.



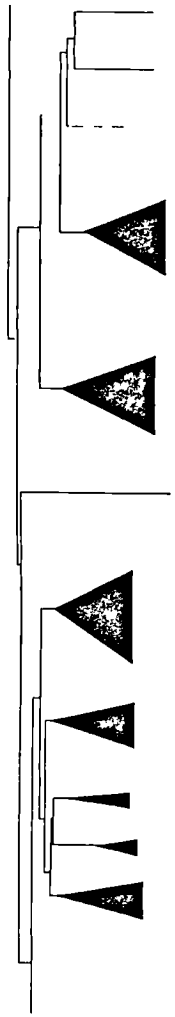
Tropheryma whippelii	ND	ND	ND	ND	ND	ND	ND	ND	ND	uncultured
Sporichthya polymorpha	I	–	A3γ	ND	3a	-9(H_6,H_8)	PI	70	facultative anaerobe	short hyphae
Dermatophilaceae										
Dermatophilus	III	B	A1γ	ND	1a	-8(H_4)	PI	57-59	micro-aerophile	tapering filaments
Terrabacter tumescens	I	–	A3γ	ND	2?	-8(H_4)	PII j	69-72	aerobe	rods and motile cocci
Kineosporia aurantiaca	I	–	A3γ	ND	ND	-9(H_4)	PIII	69	aerobe	short hyphae
Micrococcaceae										
Arthrobacter	VI	–	A3α	ND	2c	-9(H_2)	PI	59-70	aerobe	cocci
Dermabacter	III	ND	ND	ND	ND	ND	ND	ND	facultative anaerobe	short rods
Micrococcus	VI	–	A3α,A4α	ND	2c	-7,-8,-8(H_2),-9(H_2)	PI	65-75	aerobe	cocci
Renibacterium	VI	–	A3α	ND	2c	-9,-10	PI	53-54	aerobe	short rods
Rothia	VI	–	A3α	ND	2c	-7	PI	54-57	facultative anaerobe	cocci, pleomorphic
Stomatococcus	VI	–	A3α	ND	2	ND	PI	56-60	facultative anaerobe	cocci
Jonesia denitrificans^{+}	VI	–	A4α	ND	2b	-9	PV	56-58	aerobe	branched rods, cocci
Promicromonospora^{+}	VI	–	A3α	ND	2b	-9(H_4)	PV	73-74	aerobe	fragmenting hyphae
Brevibacteriaceae										
Brevibacterium	III	C	A1γ	ND	2c	-8(H_2)	PI	60-67	aerobe	short rods

TABLE 2-8. CONTINUED.



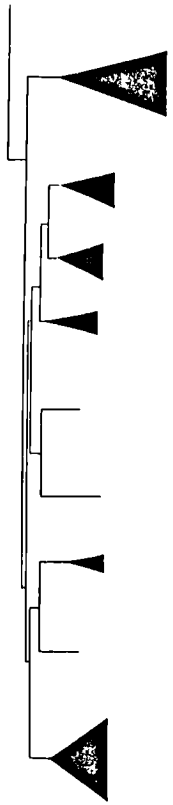
Streptomycetaceae										
Streptomyces	I	–	A3γ	A	2c	-9(H_6,H_8)	PII	69-78	aerobe	branching hyphae
Kitasatosporia	X	C	ND	A	ND	ND	ND	66-73	aerobe	branching hyphae
*Intrasporangium calvum**	I	–	A3γ	ND	1a	-8(ND)	PIV	68	aerobe	branching hyphae
Streptosporangiaceae										
Herbidospora	III	*B*	*ND*	A	*3c*	*-10(H_4,H_2,H_6)*	*PIV*	*69-71*	*aerobe*	*branching hyphae*
Microbispora	III	B	A1γ	A	3c	-9(H_4,H_2),-9	PIV	67-74	aerobe	branching hyphae
Planobispora	III	B	A1γ	ND	3c	-9(H_4,H_2),-9	PIV	70-71	aerobe	branching hyphae
Planomonospora	III	B	A1γ	ND	3a	-9(H_4,H_2),-9	PIV	72	aerobe	filamentous hyphae
Spirillospora	III	B	A1γ	ND	3a	-9(H_4,H_6)	PI,PII	69-71	aerobe	branching hyphae
Streptosporangium	III	B	A1γ	A	3c	-9(H_4,H_2),-9	PIV	69-71	aerobe	branching hyphae
Microtetraspora+	III	B,C	A1γ	A	1?	-9(H_4,H_2)	PIV	66-69	aerobe	branching hyphae
Nocardiopsis dassonvillei+	III	C	A1γ	ND	3d	-10(H_2,H_4)	PIII	64-69	aerobe	branching hyphae
Micromonosporaceae										
Actinoplanes	II	D	A1γ'	G	2d	-9(H_4),-10(H_4)	PII	72-73	aerobe	branching hyphae
Catenuloplanes japonicus	VI	D	A3α	G	2	-9(H_8),-10(H_8)	PIII	71-73	aerobe	branching hyphae
Couchioplanes caeruleus	VI	D	A	G	2c	-9(H_4,H_6)	PII	69-73	aerobe	branching hyphae
Frankia sp. G48 and L27	ND	ND	ND	ND	ND	ND	ND	ND	aerobe	extensive filamentous
Dactylosporangium	II	D	A1γ	G	3b	-9(H_4,H_6,H_8)	PII	71-73	aerobe	branching hyphae
Micromonospora+	II	D	A1γ'	G	3b	-9(H_4),-10(H_4)	PII	71-73	aerobe	branching hyphae
*Pilimelia**	II	D	A1γ	ND	2d	-9(H_2),-9(H_4)	PII	ND	aerobe	branching hyphae
Acidothermus cellulolyticum										
Frankiaceae										
Frankia	III	B,C,D	ND	ND	1	ND	PI	66-71	aerobe	extensive filamentous
Blastococcus+	ND	ND	ND	ND	ND	ND	ND	ND	aerobe	multilocular thallus
Geodermatophilus+	III	C	A1γ	ND	2b	-9(H_4)	PII	73-76	aerobe	multilocular thallus

TABLE 2-8. CONTINUED.



Luteococcus japonicus	I	–	A3γ	A	1?	-9(H_4)	PI	66-68	facultative anaerobe	cocci
Propioniferax innocua	I	–	A3γ	ND	2	-9(H_4)	ND	59-63	facultative anaerobe	rods (clusters, V-form)
Microlunatus phosphovorus	I	–	ND	ND	2	-9(H_4)	ND	68	aerobe	cocci
Propionibacteriaceae										
"Eubacterium combesii"	ND	ND	ND	ND	ND	ND	ND	ND	anaerobe	rods, pleomorphic
Propionibacterium	I	–	A3γ	ND	ND	-9(H_4)	PI^k	53-68	facultative anaerobe	branching hyphae, rods
Nocardioidaceae										
Aeromicrobium	I	–	A3γ	ND	1	-9(H_4)	PII	71-73	aerobe	rods and cocci
Nocardioides	I	–	A3γ	A	3a	-8(H_4)	PI	66-69	aerobe	hyphae and rods
Actinopolyspora halophila	IV	A	A1	ND	2	-9(H_4,H_6)	PIII	64	aerobe	branching hyphae
Actinokineospora riparia	IV									
Kutzneria viridogrisea	III	C	ND	A	3	-9(H_4)	PII	70-71	aerobe	branching hyphae
Saccharothrix	III	C,E	ND	ND	3f	-9(H_4),-10(H_4)	PII	70-73	aerobe	fragmenting hyphae
Pseudonocardiaceae										
Kibdelosporangium	IV	A	A1γ	A	3c	-9(H_4)	PII	66	aerobe	fragmenting hyphae
Pseudonocardia	IV	A	A1γ	ND	2b	-8(H_4)	PIII	68-79	aerobe	branching hyphae
Amycolatopsis	IV	A	A1γ	ND	3f	-9(H_4,H_2)	PII	66-69	aerobe	fragmenting hyphae
Thermocrispum agreste	III	C	ND	ND	3	-9(H_4)	PII	69-73	aerobe	branching hyphae
Saccharomonospora	IV	A	A1γ	ND	2a	-9(H_4),-8(H_4)	PII	66-70	aerobe	branching hyphae
Saccharopolyspora	IV	A	A1γ	ND	2c	-9(H_4,H_2),-10(H_4)	PIII	70-72	aerobe	fragmenting hyphae
*Actinobispora**	IV	A,D	ND	ND	ND	-7(H_2),-9(H_2)	PIV	71	aerobe	branching hyphae
*Actinosynnema**	III	C	ND	A	3f	-9(H_4),-10(H_4)	PII	71-73	aerobe	branching hyphae
*Streptoalloteichus**	III	C	ND	ND	ND	-9(H_6),-10(H_6)	ND	ND	ND	branching hyphae

TABLE 2-8. CONTINUED.



Corynebacteriaceae										
Corynebacterium	IV	A	A1γ	A	1a	-8(H_2),-9(H_2)	PI	51-63	facultative anaerobe	rods
Turicella otitidis	IV	A	ND	ND	1	-10,-11	ND	65-72	ND	rods and V-shaped
Nocardiaceae										
Nocardia	IV	A	A1γ	G	1b	-8(H_4,ω-cycl)	PII	64-72	aerobe	fragmenting hyphae
Rhodococcus equi clade	IV	A	A1γ	G	1b	-8(H_2),-9(H_2)	PII	63-73	aerobe	pleomorphic rods/cocci
Rhodococcus rhodochrous clade	IV	A	A1γ	G	1b	-8(H_2),-9(H_2)	PII	63-73	aerobe	pleomorphic rods/cocci
Tsukamurella	IV	A	A1γ	G	1b	-9	PII	67-68	aerobe	pleomorphic rods/cocci
Dietzia maris	IV	A	A1γ	A	1b	-8(H_2),	PII	73	aerobe	pleomorphic rods/cocci
Gordona	IV	A		G	1b	-9(H_2)	PII	60-66	aerobe	pleomorphic rods/cocci
Nocardia pinensis	IV	A	A1γ	ND	1b	-8(H_4,ω-cycl)	PII	67.5	aerobe	pleomorphic rods/cocci
Mycobacteriaceae										
Mycobacterium	IV	A	A1γ	G	1b	-9(H_2)	PII	62-70	aerobe	bacillary

TABLE 2-8. CONTINUED.

[a], rRNA sequence analyses carried out as described in Appendix F.

[b], Assignment of organisms to taxonomic groups based on 16S rRNA sequence data (Appendix F). Taxa marked with + or * have been assigned to groups on the basis of 16S rRNA cataloguing and chemotaxonomic data, respectively, but have yet to be the subject of 16S rRNA sequence analyses. ND, data not available.

[c], Major constituents in wall chemotypes: I, L-DAP and glycine; II, *meso*-DAP and glycine; III, *meso*-DAP; IV, *meso*-DAP, arabinose and galactose; V, lysine and ornithine; VI, lysine (with variable presence of aspartic acid and galactose); VII, diaminobutyric acid and glycine (lysine variable); VIII, ornithine; IX, LL- and *meso*-DAP. All wall preparations contain major amounts of alanine, glutamic acid, glucosamine and muramic acid (Lechevalier & Lechevalier, 1970, 1980).

[d], Whole organism sugar patterns of actinomycetes containing *meso*-DAP: A, arabinose and galactose; B, madurose (3-O-methyl-D-galactose); C, no diagnostic sugars; D, arabinose and xylose; –, not applicable (Lechevalier *et al.*, 1971).

[e], Peptidoglycan classification (Schleifer & Kandler, 1972; Schleifer & Seidl, 1985) is described in detail in Table 2-4. Type A, cross-linkage between positions 3 and 4 of two peptide subunits: 1, direct cross-linkage (no interpeptide bridge); 2, cross-linkage by polymerised peptide bridge; 3, cross-linkage by interpeptide bridges consisting of monocarboxylic L-amino acids or glycine, or both; 4, cross-linkage by interpeptide bridges containing a dicarboxylic amino acid; 5, cross-linkage by interpeptide bridges containing a dicarboxylic amino acid and lysine. Peptidoglycan type B, cross-linkage between positions 2 and 4 of two peptide subunits: 1, interpeptide bridge containing a L-diamino acid; 2, interpeptide bridge containing a D-diamino acid. Small Greek letters mark the diversity of amino acids in position 3 of the peptide subunit (Table 2-4). Prime (') indicates the replacement of alanine in position 1 in type A peptidoglycans by glycine (Schleifer & Kandler, 1972; Schleifer & Stackebrandt, 1983; Schleifer & Seidl, 1985).

[f], Muramic acid types: A, N-acetylmuramic acid; G, N-glycolylmuramic acid (Uchida & Aida, 1977, 1984).

[g], Fatty acid classification after Kroppenstedt (1985). Numbers refer to the type of fatty acid biosynthetic pathway and letters to the types of fatty acids (FA) synthetised. Type 1, pathway generating straight-chain fatty acids, including saturated and unsaturated (FA-type 1a), 10-methyl-branched (FA-type 1b) and cyclopropane fatty acids (FA-type 1c), the latter two being derived from the unsaturated compounds. Type 2, pathway yielding terminally-branched fatty acids (FA-type 2), that is, *iso*- and *anteiso*-branched fatty acids. Type 3 encompasses organisms that have complex branched fatty acid types, that is, both 10-methyl-branched (type 1) and *iso*- and/or *anteiso*-branched (type 2) fatty acids.

[h], MK-9(H_6,H_8), notation for a hexa- or octahydrogenated menaquinone with nine isoprene units.

[i], Characteristic phospholipids: PI, nitrogenous phospholipids absent (with phosphatidylglycerol variable); PII, only phosphatidylethanolamine; PIII, phosphatidylcholine (with phosphatidylethanolamine, phosphatidylmethylethanolamine and phosphatidylglycerol variable, phospholipids containing glucosamine absent); PIV, phospholipids containing glucosamine (with phosphatidylethanolamine and phosphatidylmethylethanolamine variable) and PV, phospholipids containing glucosamine and phosphatidylglycerol (with phosphatidylethanolamine variable); all preparations contain phosphatidylinositol (Lechevalier *et al.*, 1977, 1981).

[j], Also contains diphosphatidylglycerol (Collins *et al.*, 1989).

[k], Characteristic pattern consisting of diphosphatidylglycerol, phosphatidylglycerol and two incompletely characterised glycolipids (O'Donnell *et al.*, 1985).

Table 2-9. Actinomycete taxa of uncertain phylogenetic affiliation*.

Family/genus	Wall chemo type	Whole-organism sugar pattern	Peptidoglycan and muramic acid types		Fatty acid pattern	Major menaquinones (MK)	Phospho lipid type	Mol% G+C of DNA	Relation to oxygen	Morphology
'Actinoalloteichus cyanogriseus'	ND	ND	ND	ND	ND	-9(H_4)	ND	ND	ND	ND
Actinocorallia herbida	III	C	ND	A	1a	-9(H_4,H_6)	PII	73	aerobe	branching hyphae
Brachybacterium	II	C	A4γ	G	2	-7	PI	68-72	aerobe	rods, cocci, V-shaped
Catellatospora	II	D	ND	G	2b	-9(H_4,H_6), -10(H_6,H_8)	PII	71-72	aerobe	branching hyphae
Cellulomonas	VI	–	A4β,A4α	ND	2b	-9(H_4)	PV	71-76	facultative anaerobe	pleomorphic
Deinococcus	V	C	A3α,A3β	ND	2	-8	'complex'	62-71	aerobe	tetrad cocci
Exiguobacterium	VI	ND	A	ND	2	-7	PII	53-56	facultative anaerobe	rods-cocci, motile
Glycomyces	II	D	ND	ND	2c	-9(H_4),-10(H_4)	PI	71-73	aerobe	branching hyphae
Kineococcus aurantiacus	III	A	A1γ	A	2?	-9(H_2)	ND	75	aerobe	cocci, motile
'Microthrix parvicella'	ND	ND	ND	ND	ND	ND	ND	ND	aerobe	straight filamentous
Mycoplana	III	C	ND	ND	ND	ND	ND	64-69	aerobe	pleomorphic, motile rods
Oerskovia	VI	–	A4α	A	2b	-9(H_4)	PV	70-75	facultative anaerobe	fragmenting mycelium, motile
Pelczaria	VI	C	ND	ND	ND	ND	ND	59	aerobe	cocci
Planotetraspora	III	A,D	ND	ND	ND	ND	ND	ND	aerobe	branching hyphae
Rarobacter	VI	–	A4β	A	2c	-9	ND	65-66	facultative anaerobe	pleomorphic, motile rods
Rubrobacter	VI	–	A3α	ND	3	-8	PI	68	aerobe	pleomorphic
'Sarraceniospora'	II	D	ND	G	2a	-9(H_6,H_8)	PII	70-72	ND	ND
'Thermomonosporaceae'										
Actinomadura	III	B	A1γ	A	3a	-9(H_6,H_8,H_4)	PI	66-70	aerobe	branching hyphae
Planopolyspora										
Thermomonospora	III	C	A1γ	ND	3c	-9(H_6,H_8,H_4)	PI	66-70	aerobe	branching hyphae

*Abbreviations and codes are explained in the footnote of Table 2-9.

lineages were uncertain. Bifidobacteria, including *Gardnerella vaginalis*, formed the deepest phylogenetic line of descent and the families *Nocardioidaceae* and *Propionibacteriaceae* were seen to form a clade. Embley and Stackebrandt (1994) also considered that the family *Actinomycetaceae* and the multimembered clade which contains the families *Arthrobacteriaceae*, *Cellulomonadaceae*, *Microbacteriaceae* and *Dermatophilaceae* formed a distinct phyletic lineage.

It is apparent from Figure 2-4 that bifidobacteria, including *Gardnerella vaginalis*, and the families *Nocardioidaceae* plus *Propionibacteriaceae* each form a monophyletic line. However, members of the family *Actinomycetaceae* and the phyletic line that encompasses the families *Arthrobacteriaceae*, *Cellulomonadaceae*, *Microbacteriaceae* and *Dermatophilaceae* were recovered in separate clades (Figure 2-4).

Good correlation exists between most of the suprageneric groups and the distribution of chemical markers (Table 2-8). The family *Microbacteriaceae* encompasses organisms with a type B peptidoglycan (Park *et al.*, 1993; Rainey *et al.*, 1994b), members of the *Corynebacterium-Mycobacterium-Nocardia* phyletic group have a wall chemotype IV and mycolic acids (Ruimy *et al.*, 1994b; Rainey *et al.*, 1995a) and organisms assigned to the family *Pseudonocardiaeae* have a wall chemotype IV without mycolic acids (Embley *et al.*, 1988a,b; Warwick *et al.*, 1994).

It is evident from Table 2-8 that wall chemotypes can be useful in distinguishing families or higher taxonomic ranks whereas menaquinone and phospholipid types are of particular value in separating actinomycete genera. Several actinomycete taxa, notably the families *Actinoplanaceae* Couch 1955[AL], *Brevibacteriaceae* Breed 1953[AL], *Corynebacteriaceae* Lehmann and Neumann 1907[AL], *Frankiaceae* Becking 1970[AL] and *Micromonosporaceae* Krassilnikov 1938[AL] were described on the basis of morphological and phenetic properties. The

descriptions of these taxa need to be re-evaluated in light of chemical and molecular systematic data.

The actinomycete line of decent. Organisms assigned the **'actinomycete line of decent'** by Embley and Stackebrandt (1994) include members of the genus *Atopobium* Collins and Wallbanks 1992, *Coriobacterium glomerans* Haas and König 1988 (Rainey *et al.*, 1994a) and *Sphaerobacter thermophilus* Demharter *et al.* 1989. "*Microthrix parvicella*" is now considered to be a member of the 'actinomycete line of decent' (Blackall *et al.*, 1994).

Members of the genus *Atopobium* are Gram-positive, non-motile, non-sporeforming organisms which are obligate anaerobes. The major fermentation products produced from glucose are acetic, formic and lactic acids; trace amounts of succinic acid may also be formed. The G+C content of DNA from *Atopobium* strains falls with the range 35 to 46 mol% (Collins & Wallbanks, 1992). *Atopobium minutum* lacks the homologous insertion of about 100 nucleotides, that lies between helices 54 and 55 of the 23S rRNA, which has been found in all actinomycetes examined so far (Roller *et al.*, 1992; Embley & Stackebrandt, 1994). Rainey *et al.* (1994a) found a relatively close suprageneric relationship between the genus *Atopobium* and *Coriobacterium glomerans*.

Coriobacterium glomerans was proposed by Haas and König (1988) for a Gram-positive, anaerobic organism that forms long chains of irregular pear-shaped cells with large spherical involutions and has DNA rich in G+C (60 to 61 mol%). The organism was isolated from the intestinal tract of red soldier bugs, *Pyrrhocoris apterus* (Haas & König, 1987). The genus *Atopobium* and *Coriobacterium glomerans* were also found to be related in the present study (Figure 2-4).

A Gram-intermediate, facultatively thermophilic iron-oxidising organism, labelled strain TH3, was isolated from a copper mine spoilage dump (Brierley, 1978)

and sequenced by Lane *et al.* (1992). The resultant 16S rRNA sequence data form the only available taxonomic information on this strain. Blackall *et al.* (1994) found that a "*Microthrix parvicella*" strain isolated from activated sludge foams in Australia showed a loose phylogenetic relationship to strain TH3. 16S rDNA sequences directly amplified from nucleic acids extracted from Australian soil were found to have a close phylogenetic relationship with strain TH3 and "*Microthrix parvicella*" (Liesack & Stackebrandt, 1992; Stackebrandt *et. al.*, 1993; Blackall *et al.*, 1994). The final member of the 'actinomycete line of decent' is *Sphaerobacter thermophilus*, a pleomorphic Gram-positive organism which has a high G+C content (66 mol%); this organism was isolated from heat-treated sewage sludge (Demharter *et al.*, 1989).

It is clear that additional comparative taxonomic studies are needed to resolve the relationships between organisms assigned to the 'actinomycete line of decent'.

4. Family Nocardiaceae and related taxa

4.1. Suprageneric classification of mycolic acid-containing taxa

Actinomycetes with *meso*-diaminopimelic acid, arabinose and galactose in the wall peptidoglycan (wall chemotype IV *sensu* Lechevalier and Lechevalier 1970) fall into two well separated suprageneric groups (Goodfellow, 1992; Figure 2-4). Wall chemotype IV actinomycetes which contain mycolic acids belong to the genera *Corynebacterium*, *Dietzia*, *Gordona*, *Mycobacterium*, *Nocardia*, *Rhodococcus* and *Tsukamurella* (Goodfellow, 1992; Rainey *et al.*, 1995c) and the their mycolateless counterparts to the family *Pseudonocardiaceae* (Embley *et al.*, 1988a,b; McVeigh *et al.*, 1994; Warwick *et al.*, 1994).

Mycolic acid-containing actinomycetes are currently assigned to three suprageneric taxa, namely the families *Corynebacteriaceae* Lehmann and Neumann 1907[AL], *Mycobacteriaceae* Chester 1897[AL] and *Nocardiaceae* Castellani and Chalmers 1919[AL]. The family *Corynebacteriaceae* currently contains the genus *Corynebacterium* Lehmann and Neumann 1896[AL], the family *Mycobacteriaceae* the genus *Mycobacterium* Lehmann and Neumann 1896[AL] and the family *Nocardiaceae* four genera, namely *Gordona* (Tsukamura 1971) Stackebrandt *et al.* 1988[AL], *Nocardia* Trevisan 1889[AL], *Rhodococcus* Zopf 1891[AL] and *Tsukamurella* Collins *et al.* 1988.

The suprageneric positions of some mycolic acid-containing taxa, notably the genera *Dietzia* Rainey *et al.* 1995 and *Tsukamurella* Collins *et al.* 1988, are not clear. The genus *Dietzia* was proposed by Rainey *et al.* (1995c) for organisms which had previously been classified as *Rhodococcus maris* Nesterenko *et al.* 1982. *Turicella otitidis* Funke *et al.* 1994 contains organisms which lack mycolic acids but are morphologically and phylogenetically related to corynebacteria (Figure 2-4; Appendix F). It is interesting that *Turicella otitidis* shows a relatively close relationship with *Corynebacterium amycolatum* Collins *et al.* 1988, a taxon that also encompasses mycolateless organisms (Collins *et al.*, 1988b).

Members of most mycolic acid-containing taxa can readily be distinguished from one another and from related actinomycetes using a combination of chemical and phenotypic properties (Tables 2-10 and 2-11). The taxonomy of the genera *Corynebacterium* and *Mycobacterium* have been reviewed recently (Liebl, 1992; Von Graevenitz & Krech, 1992; Hartmans & De Bont, 1992) and hence will not be considered in detail here.

4.2. The family *Nocardiaceae* Castellani and Chalmers 1919[AL]

The family *Nocardiaceae* is mainly classified on the basis of chemotaxonomic criteria. Goodfellow (1992) recommended that the taxon should be restricted to actinomycetes that show the following characteristics: (*i*) A peptidoglycan composed of N-acetylglucosamine, D-alanine, L-alanine, and D-glutamic acid with *meso*-diaminopimelic acid as the diamino acid and muramic acid in the N-glycolated form (Uchida & Aida, 1979); (*ii*) a polysaccharide fraction of the wall peptidoglycan rich in arabinose and galactose (whole-organism sugar pattern A *sensu* Lechevalier and Lechevalier 1970); (*iii*) a phospholipid pattern consisting of diphosphatidylglycerol, phosphatidylethanolamine (taxonomically significant nitrogenous phospholipid), phosphatidylinositol and phosphatidylinositol mannosides (phospholipid type II *sensu* Lechevalier *et al.*, 1977); (*iv*) a fatty acid profile showing predominant amounts of straight chain and unsaturated fatty acids plus tuberculostearic acid (fatty acid type 1b *sensu* Kroppenstedt 1985); (*v*) mycolic acids with 48 to 78 carbons; and (*vi*) DNA within the range 66 to 74 mol% GC.

The type genus is *Nocardia* Trevisan 1889[AL].

Validly described species currently classified in the family *Nocardiaceae* are shown in Table 2-12.

***The genus* Nocardia *Trevisan 1889*[AL].** Nocardiae are aerobic, catalase-positive actinomycetes which form rudimentary to extensively branched, substrate hyphae that often fragment *in situ*, or on mechanical disruption, into rod-shaped to coccoid, nonmotile elements. Aerial hyphae, at times visible only microscopically, are almost always present. Short-to-long chains of well-to-poorly differentiated conidia may occasionally be found on aerial hyphae and, more rarely, on both aerial and substrate hyphae. In addition to these morphological properties, nocardiae contain

TABLE 2-10. PHENOTYPIC CHARACTERISTICS OF WALL CHEMOTYPE IV ACTINOMYCETES WITH MYCOLIC ACIDS AND SOME RELATED ORGANISMS[a].

Taxon	Cell morphology	Aerial hyphae	Degree of acid-fastness	Time for visible colonies to appear (days)	5-Fluoro-uracil (20 µg/ml)	Lysozyme (50 µg/ml)	Mitomycin C (5 µg/ml)
Corynebacterium	Straight to slightly curved rods which reproduce by snapping division; club-shaped elements may also be formed	Absent	Sometimes weakly acid-fast	1-2	ND	ND	ND
Dietzia	Short rods and cocci	Absent	Not acid-fast	1-3	ND	ND	ND
Gordona	Rods and cocci	Absent	Often partially acid-fast	1-3	+	+	+
Mycobacterium	Slightly curved or straight rods, sometimes branching filaments that fragment into rods and coccoid elements	Usually absent[b]	Usually strongly acid-fast	2-40	ND	-	ND
Nocardia	Substrate mycelium fragments into rods and coccoid elements	Present	Often partially acid-fast	1-5	-	-	-
Nocardia pinensis	Substrate mycelium resembles that of a pine-tree in early stage of growth (24 h); short microscopically visible aerial hyphae formed	Present	Not acid-fast	9-21	+	+	+
Rhodococcus	Rods to extensively branched substrate mycelium that fragments into irregular rods and cocci	Absent	Often partially acid-fast	1-3	+	+	+
Tsukamurella	Straight to slightly curved rods occur singly, in pairs, or in masses	Absent	Weak to strongly acid-fast	1-3	-	-	-
Turicella otitidis	Coccoid forms which occur as single cells or are arranged in V-shaped forms or palisades	Absent	ND	1-2	ND	ND	ND

+, Resistant and -, sensitive.
[a], Data taken from Blackall *et al.* (1989b), Goodfellow (1992), Boiron *et al.* (1993), Funke *et al.* (1994), Holt *et al.* (1994) and Rainey *et al.* (1995c).
[b], *Mycobacterium farcinogenes* and *Mycobacterium xenopi* may occasionally produce aerial hyphae.

TABLE 2-11. CHEMOTAXONOMIC CHARACTERS WHICH CAN BE USED TO DIFFERENTIATE BETWEEN WALL CHEMOTYPE IV ACTINOMYCETES WITH MYCOLIC ACIDS AND BETWEEN THEM AND AMYCOLATE TAXA[a].

Characters	*Corynebacterium*	*Dietzia*	*Gordona*	*Mycobacterium*	*Nocardia*	*Nocardia pinensis*	*Rhodococcus*	*Tsukamurella*	*Turicella*
Muramic acid type in peptidoglycan[b]	Acetyl	Acetyl	Glycolyl	Glycolyl	Glycolyl	ND	Glycolyl	Glycolyl	ND
Mycolic acids:									
Overall size	22-36	34-38	48-66	60-90	44-60	58-64	34-52	64-78	Absent
Number of double bonds	0-2	ND	1-4	1-3	0-3	2-6	0-4	1-6	Absent
Side chain	8-18	ND	16-18	22-26	12-18	16-18	12-18	20-22	Absent
Fatty acid type[c]	1a[d]	1b	1b	1b[e]	1b	1b	1b	1b	1b
Phospholipids type	I	II	II	II	II	II	II	II	ND
Phosphatidylethanolamine	-[a]	+	+	+	+	+	+	+	ND
Phosphatidylinositol and Phosphatidylinositolmannoside	+	-	+	+	+	+	+	+	ND
Predominant menaquinones	MK-8(H_2), -9(H_2)	MK-8(H_2)	MK-9(H_2)	MK-9(H_2)	MK-8(H_4, ω-cycl)	MK-8(H_4, ω-cycl)	MK-8(H_2)	MK-9	MK-10, -11
Mycobactin	ND	-	+	+	+	ND	-	ND	ND
DNA base composition (G+C mol%)	51-67	73	63-69	62-72	64-72	67.5	63-73	67-68	65-72

+, present and -, absent.

[a], Data taken from Blackall *et al.* (1989b), Goodfellow (1992), Boiron *et al.* (1993), Funke *et al.* (1994), Holt *et al.* (1994) and Rainey *et al.* (1995c).

[b], Acyl group detected using the glycolate test (Uchida & Aida, 1979).

[c], Details of fatty acid types can be found in Table 2-3.

[d], *Corynebacterium bovis* contains tuberculostearic acid (Collins *et al.*, 1982).

[e], *Mycobacterium gordonae* lacks substantial amounts of tuberculostearic acid (Minnikin *et al.*, 1985).

Table 2-12. Validly described species classified in the family *Nocardiaceae* Castellani and Chalmers 1919[AL].

Taxon[a]	Name cited in the Approved Lists of Bacterial Names[b]:
Genus: *Gordona* (Tsukamura 1971) Stackebrandt *et al.* 1988[AL]	-
G. aichiensis (Tsukamura 1982) Klatte *et al.* 1994c[VP]	-
G. amarae (Lechevalier and Lechevalier 1974) Klatte *et al.* 1994c[VP]	*Nocardia amarae*
G. bronchialis[TSP] (Tsukamura 1971) Stackebrandt *et al.* 1988[VP]	*Rhodococcus bronchialis*
G. rubropertincta (Hefferan 1904) Stackebrandt *et al.* 1988[VP]	*Rhodococcus rubropertinctus*
G. sputi (Tsukamura 1978) Stackebrandt *et al.* 1988[VP]	-
G. terrae (Tsukamura 1971) Stackebrandt *et al.* 1988[VP]	*Rhodococcus terrae*
Genus: *Nocardia* Trevisan 1889[AL]	*Nocardia*
N. asteroides[TSP] (Eppinger 1891) Blanchard 1896[AL]	*Nocardia asteroides*
N. brasiliensis (Lindenberg 1909) Pinoy 1913[AL]	*Nocardia brasiliensis*
N. brevicatena (Lechevalier *et al.* 1961) Goodfellow and Pirouz 1982[VP]	*Micropolyspora brevicatena*
N. carnea (Rossi Doria 1891) Castellani and Chalmers 1913[AL]	*Nocardia carnea*
N. farcinica Trevisan 1889[AL]	*Nocardia farcinica*
N. nova Tsukamura 1982[VP]	-
N. pinensis Blackall *et al.* 1989[VP]	-
N. otitidiscaviarum Snijders 1924[AL]	*Nocardia otitidiscaviarum*
N. petroleophila Hirsch and Engel 1956[AL]	*Nocardia petroleophila*[b]
N. seriolae Kudo *et al.* 1988[VP]	-
N. transvalensis Pijper and Pullinger 1927[AL]	*Nocardia transvalensis*
N. vaccinii Demaree and Simth 1952[AL]	*Nocardia vaccinii*
Genus: *Rhodococcus* Zopf 1891[AL]	*Rhodococcus*
R. coprophilus Rowbotham and Cross 1977[AL]	*Rhodococcus coprophilus*
R. equi (Magnusson 1923) Goodfellow and Alderson 1977[AL]	*Rhodococcus equi*
R. erythropolis (Gray and Thornton 1928) Goodfellow and Alderson 1977[AL]	*Rhodococcus erythropolis*
R. fascians (Tilford 1936) Goodfellow 1984[VP]	*Corynebacterium fascians*
R. globerulus Goodfellow *et al.* 1982[VP]	-

TABLE 2-12. CONTINED.

R. marinonascens Helmke and Weyland 1984[VP]	-
R. opacus Klatte *et al.* 1994[VP]	-
R. rhodnii Goodfellow and Alderson 1977[AL]	*Rhodococcus rhodnii*
R. rhodochrous[TSP] (Zopf 1891) Tsukamura 1974[AL]	*Rhodococcus rhodochrous*
R. ruber (Kruse 1896) Goodfellow and Alderson 1977[AL]	*Rhodococcus ruber*
R. zopfii Stoecker *et al.* 1994[VP]	-
"Rhodococcus (Nocardia) coeliaca" (Gray & Thornton 1928) Waksman & Henrici 1948[AL]	*Nocardia coeliaca*[c]
"Rhodococcus (Nocardia) corynebacteroides" Serrano *et al.* 1972[AL]	*Nocardia corynebacteroides*[c]
Genus: *Tsukamurella* Collins *et al.* 1988[VP]	-
Tsukamurella paurometabola[TSP] (Steinhaus 1941) Collins *et al.* 1988[VP]	*Corynebacterium paurometabolum*
Tsukamurella wratislaviensis" Goodfellow *et al.* 1991[VP]	-

Abbreviations: [AL], Cited in the Approved Lists of Bacterial Names (Skerman *et al.*, 1980); [VP], valid publication; [TSP], type species.

[a], Listed references are effective publications; [b], *Nocardia petroleophila* is a member of the genus *Pseudonocardia* (Ruimy et al., 1994b; Warwick et al., 1994) ; and [c], *Nocardia coeliaca* and *Nocardia corynebacteroides* are members of the genus *Rhodococcus* (Goodfellow, 1989b). For details see text.

mycolic acids with 40 to 60 carbons and up to three double bonds; the fatty acid esters released on pyrolysis gas chromatography of mycolic acid esters contain 12 to 18 carbon atoms and may be saturated or unsaturated. The predominant menaquinone corresponds to a hexahydrogenated menaquinone with eight isoprenoid units which the end two units cyclised. The major phospholipids are diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylinositol and phosphatidylinositol mannosides. The G+C content of the DNA is 64 to 72 mol%.

The type species is *Nocardia asteroides* (Eppinger 1891) Blanchard 1896[AL].

Nocardiae are widely distributed and abundant in soil. Some strains are opportunistic pathogens for animals including man.

Historically, the genus *Nocardia* was a dumping ground for "nocardioform actinomycetes" (Lechevalier, 1976). Twenty species of *Nocardia* were cited on the Approved Lists of Bacterial Names (Skerman *et al.*, 1980). *Nocardia autotrophica* (Takamiya and Tubaki 1956) Hirsch 1961[AL], *Nocardia cellulans* Metcalf and Brown 1957[AL], *Nocardia hydrocarbonoxydans* Nolof and Hirsch 1962[AL], *Nocardia mediterranei* (Margalith and Beretta 1960) Thiemann *et al.* 1969[AL], *Nocardia orientalis* (Pittenger and Brigham 1956) Pridham and Lyons 1969[AL] and *Nocardia saturnea* Hirsch 1960[AL] were subsequently reclassified as *Pseudonocardia autotrophica* (Takamiya and Tubaki 1956) Warwick *et al.* 1994, *Cellulomonas cellulans* Stackebrandt and Woese 1981, *Pseudonocardia hydrocarbonoxydans* (Nolof and Hirsch 1962) Warwick *et al.* 1994, *Amycolatopsis mediterranei* (Margalith and Beretta 1960) Lechevalier *et al.* 1986, *Amycolatopsis orientalis* (Pittenger and Brigham 1956) Lechevalier *et al.* 1986 and *Pseudonocardia saturnea* (Hirsch 1960) Warwick *et al.* 1994, mainly on the basis of chemotaxonomic and 16S rRNA sequence data.

The taxonomic status of *Nocardia petroleophila* Hirsch and Engel 1956[AL] is a source of confusion. This organism was cited in the Approved Lists of Bacterial Names (Skerman *et al.*, 1980) but was listed as a '*species incertae sedis*' in the last edition of Bergey's Manual of Systematic Bacteriology (Goodfellow & Lechevalier, 1989) as it did not have a wall chemotype IV (Hirsch, 1960). The 16S rRNA sequence of the type strain of *Nocardia petroleophila* (Hirsch laboratory number [IFAM] 78 = DSM43193 = ATCC15777) was determined by Wersing *et al.* (unpublished; accession number X55608; deposited on September 1990) who inadvertently mislabelled the type strain "*Amycolata petrophilea*". However, the proposal of Wersing *et al.* (unpublished) to transfer *Nocardia petroleophila* to the genus *Amycolata* as *Amycolata petroleophila* has not been effectively published.

Ruimy *et al.* (1994b) also determined the 16S rDNA sequence of the type strain of *Nocardia petroleophila* (ATCC15777^T=IFAM 78; accession number X80596). These authors compared their sequence with that of Wersing *et al.* and concluded that:

> *"The type strain* N. petroleophila *ATCC 15777^T ... was clearly closely related to* Amycolata petrophilea *(IFAM 78) with which it formed a monophyletic taxon."*

It is clear that strains ATCC 15777 and IFAM 78 refer to the same strain and that the proposition forwarded by Ruimy *et al.* (1994b) is misplaced. In contrast, Warwick *et al.* (1994) recommended that *Nocardia petroleophila* should be transferred in the genus *Pseudonocardia* as *Pseudonocardia petroleophila* citing the 16S rRNA sequence data of Wersing *et al.* (unpublished; accession number X55608) as the sole line of evidence. It is clear that the taxonomic status of *Nocardia petroleophila* needs to be clarified!

Five of the remaining validly described species of *Nocardia* should probably be reclassified in the genus *Rhodococcus*. Thus , in the latest edition of Bergey's

Manual of Systematic Bacteriology (Goodfellow, 1989b) it was proposed that (*i*) *Nocardia calcarea* Metcalf and Brown 1957[AL] should be reduced to a synonym of *Rhodococcus erythropolis* (Gray and Thornton 1928) Goodfellow and Alderson 1977[AL]; (ii) *Nocardia corynebacteroides* Serrano *et al.* 1972[AL] and *Nocardia globerula* (Gray 1928) Waksman and Henrici 1948[AL] should become synonyms of *Rhodococcus globerulus* Goodfellow *et al.* 1982; (*iii*) *Nocardia restricta* (Turfitt 1944) McClung 1974[AL] should be seen as a synonym of *Rhodococcus equi* (Magnusson 1923) Goodfellow and Alderson 1977[AL]; and *Nocardia coeliaca* (Gray and Thornton 1928) Waksman and Henrici 1948[AL] should be transferred to the genus *Rhodococcus* as *Rhodococcus coeliaca*. However, none of these recommendations have been validly published. It was left to Rainey *et al.* (1995b) to formally propose that *Nocardia calcarea* and *Nocardia restricta* be reduced to synonyms of *Rhodococcus erythropolis* and *Rhodococcus equi*, respectively. The inclusion of *Nocardia corynebacteroides* in the genus *Rhodococcus* is supported by 16S rDNA sequence data which showed that the strain forms an unique lineage within the evolutionary radiation encompassed by the genus *Rhodococcus* (Rainey *et al.*, 1995a).

The inclusion of *Nocardia amarae* Lechevalier and Lechevalier 1974[AL] in the genus *Nocardia* has frequently been questioned since this organism contains dihydrogenated menaquinones with nine isoprene units, releases C_{16} and C_{18} monounsaturated esters on pyrolysis of methyl mycolates, and is unable to grow in lysozyme broth (Goodfellow, 1992). All of these properties are consistent with proposals to transfer *Nocardia amarae* to the genus *Gordona* as *Gordona amarae* on the basis of both chemical and 16S rDNA sequence data (Goodfellow *et al.*, 1994; Klatte *et al.*, 1994c; Ruimy *et al.*, 1994b). The proposal of Klatte *et al.* (1994c) has been validated (IJSB, 1995, page 199).

The proposals outlined above leave the genus *Nocardia* as a homogeneous taxon for the first time in its turbulent taxonomic history. The genus currently contains eleven validly described species, namely *Nocardia asteroides* (Eppinger 1891) Blanchard 1896[AL] ,*Nocardia brasiliensis* (Lindenberg 1909) Pinoy 1913[AL] ,*Nocardia brevicatena* (Lechevalier *et al.* 1961) Goodfellow and Pirouz 1982, *Nocardia carnea* (Rossi Doria 1891) Castellani and Chalmers 1913[AL], *Nocardia farcinica* Trevisan 1889[AL], *Nocardia nova* Tsukamura 1982, *Nocardia pinensis* Blackall *et al.* 1989, *Nocardia otitidiscaviarum* Snijders 1924[AL], *Nocardia seriolae* Kudo *et al.* 1988, *Nocardia transvalensis* Pijper and Pullinger 1927[AL] and *Nocardia vaccinii* Demaree and Simth 1952[AL].

Some nocardiae are opportunistic pathogens which can cause actinomycete mycetoma and nocardiosis (Schaal & Lee, 1992; McNeil & Brown, 1994). The predominant agents of nocardiosis are *Nocardia asteroides*, *Nocardia farcinica* and *Nocardia nova;* organisms assigned to these taxa are difficult to distinguish on the basis of recommended diagnostic phenotypic properties (Gordon *et al.*, 1974; Goodfellow, 1971; Kudo *et al.*, 1988).

The genus* Gordona *(Tsukamura 1971) Stackebrandt et al. 1988. Gordonae are aerobic, catalase positive actinomycetes which form rods and cocci. They have an oxidative type of metabolism and are arylsulphatase negative. Gordonae have an $A1_{\gamma}$ type peptidoglycan and contain mycolic acids with 48 to 66 carbon atoms and 1 to 4 double bonds. The fatty acid esters released on pyrolysis gas chromatography of mycolic esters contain 16 to 18 carbon atoms. Gordonae contain dihydrogenated menaquinones with nine isoprene units as the predominant isoprenologue and diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylinositol and phosphatidylinositol mannosides as major phospholipid. The G+C content of the DNA is within the range 63 to 69 mole %.

The type species is *Gordona bronchialis* (Tsukamura 1971) Stackebrandt *et al.* 1988.

Gordonae are widely distributed and are common in soil. Some strains have been associated with foams in activated sludge of sewage treatment plants; others have been isolated from sputum with pulmonary disease.

The genus *Gordona* has had a somewhat checkered taxonomic history. The taxon was initially proposed by Tsukamura (1971) for slightly acid-fast actinomycetes isolated from soil and sputa of patients suffering from pulmonary diseases. The three founder species of the genus, namely *Gordona bronchialis*, *Gordona rubra* and *Gordona terrae*, were subsequently reclassified in the redescribed genus *Rhodococcus* (Tsukamura, 1974; Goodfellow & Alderson, 1977).

In the last edition of Bergey's Manual of Systematic Bacteriology, rhodococci were assigned to two aggregate groups based primarily on chemical and serological properties (Goodfellow, 1989b). Members of species originally classified in the genus *Gordona* Tsukamura 1971 contained mycolic acids with 48 to 66 carbon atoms and major amounts of dihydrogenated menaquinones with nine isoprene units (MK-9[H_2]). The remaining strains were characterised by shorter mycolic acids (34 to 52 carbon atoms) and dihydrogenated menaquinones with eight isoprene units (MK-8[H_2]; Alshamaony *et al.*, 1976; Collins *et al.*, 1977, 1985). The two aggregate groups were also recognised by their antibiotic sensitivity profiles (Goodfellow & Orchard, 1974), delayed skin reaction on sensitised guinea pigs and polyacrylamide gel electrophoresis of cell extracts (Hyman & Chaparas, 1977).

The discovery that the two rhodococcal groups were phylogenetically distinct led Stackebrandt *et al.* (1988b) to revive the genus *Gordona* Tsukamura 1971 for organisms classified as *Rhodococcus bronchialis*, *Rhodococcus rubropertincta*, *Rhodococcus sputi* and *Rhodococcus terrae*. In addition, Hall and Ratledge (1986)

had found that gordonae contained mycobactins whereas rhodococci were unable to synthesise these compounds under growth conditions where iron was limited. The mycobactins of *Gordona rubropertincta* and *Gordona terrae* were found to be quite similar and readily separated from those of *Gordona bronchialis*.

Nocardia amarae Lechevalier and Lechevalier 1974[AL] and *Rhodococcus aichiensis* Tsukamura 1982 were subsequently found to have properties consistent with their classification in the genus *Gordona*. They were transferred to the genus *Gordona* as *Gordona amarae* and *Gordona aichiensis* mainly on the basis of chemical and 16S rDNA sequence data (Goodfellow *et al.*, 1994; Klatte *et al.*, 1994c; Ruimy *et al.*, 1994b).

The genus *Gordona* currently contains six validly described species (Table 2-12). It forms a homogeneous taxon which can readily be distinguished from members of other mycolic acid-containing taxa (Tables 2-10 and 2-11).

The genus Rhodococcus Zopf 1891[AL]. Rhodococci are aerobic, Gram-positive, catalase positive, partially acid-fast, nonmotile actinomycetes which can exhibit rods and extensively branched substrate hyphae. The growth cycle starts with the coccus or short rod stage, different organisms then show a more or less complex series of morphological stages; cocci may germinate only into short rods, or form filaments with side projections, or show elementary branching, or in the most differentiated forms produce branched hyphae. The next generation of cocci or short rods is produced by fragmentation of rods, filaments, and hyphae. Some strains produce sparse, microscopically visible, aerial hyphae that may be branched or form aerial synnemata which consist of unbranched filaments that coalesce and project upwards (Goodfellow, 1992). Rhodococci are sensitive to lysozyme, arylsulfatase negative, and produce acid from glucose oxidatively.

The cell wall peptidoglycan contains major amounts of *meso*-diaminopimelic acid, arabinose and galactose. The organisms also contain diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylinositol and phosphatidylinositol mannosides as the major polar lipids, dihydrogenated menaquinones with eight isoprenoid units as the predominant isoprenologue, large amount of straight-chain, unsaturated, and tuberculostearic acids, and mycolic acids with 32 to 52 carbon atoms and 0 to 4 double bonds. The fatty acid esters released on pyrolysis gas chromatography of mycolate esters contain 12 to 18 carbon atoms. The G+C content of the DNA is 63 to 73 mol%.

The type species is *Rhodococcus rhodochrous* (Zopf 1891) Tsukamura 1974[AL].

The organism is widely distributed but particularly abundant in soil and herbivore dung. Some strains are pathogenic for animals including man.

Rhodococci have had a long and confused taxonomic pedigree (Cross & Goodfellow, 1973; Bousfield and Goodfellow, 1976; Goodfellow & Wayne, 1982; Goodfellow & Cross, 1984; Goodfellow & Lechevalier, 1989). The epithet *rhodochrous* (Zopf, 1891) was reintroduced by Gordon and Mihm (1957) for actinomycetes that had properties in common with both mycobacteria and nocardiae but which carried a multiplicity of generic and specific names. The taxon was provisionally assigned to the genus *Mycobacterium* but later found to be heterogeneous on the basis of chemical, molecular genetic and numerical phenetic data. The genus *Rhodococcus* was subsequently resurrected and redefined to encompass *rhodochrous* strains (Tsukamura, 1974; Goodfellow & Alderson, 1977).

Ten species of *Rhodococcus* were cited on the Approved Lists of Bacterial Names by Skerman *et al.* (1980). Three of these species, namely *Rhodococcus bronchialis* (Tsukamura 1971) Tsukamura 1974[AL], *Rhodococcus rubropertinctus*

(Hefferan 1904) Goodfellow and Alderson 1979[AL] and *Rhodococcus terrae* (Tsukamura 1971) Tsukamura 1974[AL], were subsequently transferred to the genus *Gordona* as *Gordona bronchialis* (Tsukamura 1971) Stackebrandt *et al.* 1988, *Gordona rubropertincta* (Hefferan 1904) Stackebrandt *et al.* 1988 and *Gordona terrae* (Tsukamura 1971) Stackebrandt *et al.* 1988. Mordarski *et al.* (1980b) found that *Rhodococcus corallinus* (Bergey *et al.* 1923) Goodfellow and Alderson 1979[AL] and *Gordona* (*Rhodococcus) rubropertincta* belonged to the single DNA homology group. Goodfellow (1989b) subsequently considered that *Rhodococcus corallinus* should become a subjective synonym of *Gordona* (*Rhodococcus*) *rubropertincta* but his proposal was never effectively published.

Thirteen new species were subsequently assigned to the genus *Rhodococcus*, namely *Rhodococcus aichiensis* Tsukamura 1982, *Rhodococcus aurantiacus* Tsukamura and Yano 1985, *Rhodococcus chlorophenolicus* Apajalahti *et al.* 1986, *Rhodococcus chubuensis* Tsukamura 1982, *Rhodococcus fascians* Goodfellow 1984, *Rhodococcus globerulus* Goodfellow *et al.* 1982, *Rhodococcus luteus* Nesterenko *et al.* 1982, *Rhodococcus marinonascens* Helmke and Weyland 1984, *Rhodococcus maris* Nesterenko *et al.*, 1982, *Rhodococcus obuensis* Tsukamura 1982, *Rhodococcus opacus* Klatte *et al.* 1994, *Rhodococcus roseus* Tsukamura *et al.* 1991 and *Rhodococcus zopfii* Stoecker *et al.* 1994. *Rhodococcus aichiensis* Tsukamura 1982, *Rhodococcus aurantiacus* Tsukamura and Yano 1985, and *Rhodococcus chlorophenolicus* Apajalahti *et al.* 1986 have been reclassified as *Gordona aichiensis* Klatte *et al.*, 1994, *Tsukamurella paurometabola* Collins *et al.* 1988 and *Mycobacterium chlorophenolicum* Häggblom *et al.* 1994, respectively, and the new genus *Dietzia* Rainey *et al.* 1995 proposed for *Rhodococcus maris* Nesterenko *et al.* 1982.

Several rhodococci described using phenotypic properties were reduced to subjective synonyms of established taxa. *Rhodococcus chubuensis* Tsukamura 1982 and *Rhodococcus obuensis* Tsukamura 1982 were transferred to the genus *Gordona* and reduced to synonyms of *Gordona sputi* (Tsukamura 1978) Stackebrandt *et al.* 1988 (Riegel *et al.*, 1994). In addition, *Rhodococcus luteus* Nesterenko *et al.* 1982 and *Rhodococcus roseus* Tsukamura *et al.* 1991 became synonyms of *Rhodococcus fascians* Goodfellow 1984 (Klatte *et al.*, 1994a), and *Rhodococcus rhodochrous* (Zopf 1891) Tsukamura 1974[AL] (Rainey *et al.*, 1995b), respectively.

The proposals outlined above leave the genus *Rhodococcus* as a distinct taxon which encompasses eleven validly described species (Table 2-12). However, evidence from 16S rRNA sequencing studies indicated that that the genus is polyphyletic (Rainey *et al.*, 1995a; see Figure 2-4 and Appendix F).

The genus* Tsukamurella *Collins et al. 1988. The genus *Tsukamurella* was introduced to accommodate organisms previously classified as *Corynebacterium paurometabolum* and *Rhodococcus aurantiacus*. *Corynebacterium paurometabolum* was proposed by Steinhaus (1941) for bacteria isolated from the mycetome and ovaries of the bedbug (*Cimex lectularis*) but its assignment to the genus *Corynebacterium* was questioned (Jones, 1975; Collins & Jones, 1982). The organism has an A1γ peptidoglycan (Schleifer & Kandler, 1972) but was distinguished from corynebacteria by the presence of long, highly unsaturated mycolic acids (Collins & Jones, 1982). A similar series of unsaturated mycolic acids were detected in *Rhodococcus aurantiacus* (Goodfellow *et al.*, 1978; Tomiyasu & Yano, 1984), the generic status of which was also considered to be equivocal.

Rhodococcus aurantiacus, first described as *Gordona aurantiacus* Tsukamura and Mizuno 1971, was transferred to the genus *Rhodococcus* by

Tsukamura (1985). However, Goodfellow *et al.* (1978) considered that "*aurantiaca*" strains merited generic status as they contained characteristic mycolic acids and unsaturated menaquinones with nine isoprene units (MK-9). *Corynebacterium paurometabolum* also contains the same type of predominant menaquinone (Collins & Jones, 1982). *Corynebacterium paurometabolum* and *Rhodococcus aurantiacus* were subsequently reduced to a single species and reclassified in the new genus *Tsukamurella* on the basis of 16S rRNA sequence and chemical data outlined above (Collins *et al.*, 1988a). An additional species, namely *Tsukamurella wratislaviensis*, was subsequently assigned to the genus by Goodfellow *et al.* (1991).

Phylogenetic studies on the family **Nocardiaceae.** The extensive 16S rRNA sequencing studies that have been carried out on representative mycobacteria have helped to clarify the internal structure of the genus *Mycobacterium* and provide a sound basis for the molecular identification of mycobacteria (Böddinghaus *et al.*, 1990; Stahl & Urbance, 1990; Rogall *et al.*, 1990; Pitulle *et al.*, 1992). In contrast, relatively few representatives of *Gordona*, *Nocardia* and *Rhodococcus* species have been the subject of 16S rRNA sequencing analyses.

In the present investigation, almost complete 16S rRNA gene sequences were determined for the type strains of validly described species of the genera *Nocardia* and *Tsukamurella*. The resultant data were compared with corresponding 16S rRNA sequence data held in the GenBank/EMBL database on representatives of the genera *Corynebacterium*, *Gordona* and *Rhodococcus*. The combined data were used to establish relationships between members of the families *Corynebacteriaceae*, *Mycobacteriaceae* and *Nocardiaceae* with four different phylogenetic inference methods being used to evaluate the suprageneric taxonomy of mycolic acid-containing actinomycetes. The type strains of all validly described

species of the genera *Nocardia* and *Tsukamurella* were examined for the presence of menaquinones.

B. Materials and Methods

1. Test strains

The sources and histories of the tests strains are given in Table 2-13. All of the strains were maintained as glycerol suspensions (20%, v/v) at -20 °C. Strain N1170, the designated type strain of a putatively novel species named "*Nocardia crassostrae*", is a pathogen of Pacific oysters (*Crassostrea gigas*) and has been distinguished from other nocardiae using phenotypic criteria (Friedman & Hedrick, 1991; C. S. Friedman, personal communication).

2. Chemotaxonomy

2.1. Preparation of biomass.

The test strains were grown in modified Sauton's broth (Mordarska *et al.*, 1972; Appendix A) for 10 days at 30 °C, checked for purity, killed by shaking with formalin (1 %, v/v) overnight and harvested by centrifugation. The resultant pellets were washed twice with distilled water and freeze-dried.

2.2. Analysis of menaquinones

Isolation of isoprenoid quinones. The method described by Minnikin *et al.* (1984) was used to extract and purify isoprenoid quinones from the test strains. Dried biomass (*ca.* 50 mg) was placed in a test tube fitted with a Teflon-lined screw cap and 2 ml of aqueous methanol (10 ml of 3% ,w/v aqueous sodium chloride in 100 ml of methanol) and 2 ml of petroleum ether (b.p. 60-80 °C) added. The contents of the tube were mixed for 15 minutes using a tube rotator then centrifuged for 5 minutes at low speed. The upper organic phase, which contained the isoprenoid quinones, was transferred to a small glass vial and dried under nitrogen at room temperature.

TABLE 2-13. TEST STRAINS.

Laboratory number	Species	Source
JC51^T	*Corynebacterium glutamicum*	NCIMB 10025
N317^T	*Nocardia asteroides*	R. E. Gordon, IMRU 727; Garden soil, Thailand. =ATCC 19247
N318^T	*Nocardia brasiliensis*	R. E. Gordon, IMRU 854; J. D. Schneidau Jr., 381; A. Batista, 631. =ATCC 19296
N1201^T	*Nocardia brevicatena*	DSM 43024; A. Seino KCC A-0029; H. Lechevalier, RIA 709; sputum of patient with tuberculosis
N1200^T	*Nocardia carnea*	DSM 43397; R. E. Gordon IMRU 3419
N1170	*"Nocardia crassostrae"*	C. S. Friedman, Bodega Marine Laboratory, California, USA, NB4H; nocardiosis of oyster
N898^T	*Nocardia farcinica*	M. Tsukamura, 23102(R-3318). =ATCC 3318
N1112^T	*Nocardia nova*	JCM 6044; M. Tsukamura 23095; R. E. Gordon R443; I. B. Christison; N. F. Conant 2338
N36^T	*Nocardia otitidiscaviarum*	NCTC 1934 (*Nocardia caviae*); E. P. Snidjers; middle-ear of guinea-pig. =ATCC 14629
N1164^T	*Nocardia pinensis*	IFO 15059; L. L. Blackall, UQM 3036
N1116^T	*Nocardia seriolae*	JCM 3360; K. Hatai, NA 8191; spleen of a yellowtail (*Seriola quinqueradiata*), Nagasaki, Japan
N1202^T	*Nocardia transvalensis*	DSM 43405; R. E. Gordon, IMRU 3426; mycetoma pedis
N1199^T	*Nocardia vaccinii*	DSM 43285; K. Kieslich, Schering 245; ATCC 11092; stem galls on blueberry
JC7^T	*Tsukamurella paurometabola*	DSM 20162 (*Corynebacterium paurometabolum*)
N805^T	*Tsukamurella wratislaviensis*	R. E. Gordon, IMRU 878; D. M. Powelson, J-17. =NCIMB 13082

^{T}, Type strains. Abbreviations: ATCC, American Type Culture Collection, 12301 Parklawn Drive, Rockville, Md., USA; DSM, Deutsche Sammlung von Microorganismen und Zellkulturen, Mascheroder Weg 1b, D-38124, Braunschweig, Germany; IFO, Institute for Fermentation, Osaka, Japan; IMRU, Institute of Microbiology, Rutgers State University, New Brunswick, N.J., USA.; JCM, Japan Collection of Microorganisms, Saitama, Japan; NCIMB, National Collection of Industrial and Marine Bacteria, St. Machar Drive, Aberdeen, Scotland, UK; NCTC, National Collection of Type Cultures, Central Public Health Laboratory, London, UK; and UQM, University of Queensland, Brisbane, Australia.

The preparations were stored in the dark at -20 °C as isoprenoid quinones are susceptible to strong light and high temperatures (Collins, 1994).

Preparative thin-layer-chromatography of isoprenoid quinones. The extracts containing the isoprenoid quinones were resuspended in 50 µl of petroleum ether (b.p. 60-80 °C) and applied as 2 cm bands on plastic-backed silica gel plates (10 cm×10 cm; Merck 5735). The thin-layer-chromatographic plates were developed in petroleum ether/acetone (95/5, v/v) and the single bands containing the menaquinones visualised and located under UV light at 254 nm. A standard menaquinone (MK-4; Sigma) was co-migrated to help identify the position of extracted menaquinones. The latter were detected as dark brown bands on a fluorescent yellow-green background. The bands were scraped from the plastic plates and deposited in 1.5 ml tubes containing 1 ml of diethyl ether. The preparations were mixed thoroughly by vortexing and centrifuged at 13,000 rpm for 5 minutes. The supernatants were transferred to small vials, dried under nitrogen and stored in the dark at -20 °C.

Analysis of isoprenoid quinones by high-performance liquid chromatography. The purified menaquinones were resuspended in 50 µl of *n*-hexane and 10 µl of each sample injected into a HPLC (Pharmacia LKB) fitted with a reverse-phase column (Spherisorb octadecylsilane [ODS] 5 µm; Jones Chromatography Ltd., Mid Glamorgan, Wales, UK). Acetonitrile-isopropanol (75:25, v/v) was used as the mobile phase and the samples were detected at 254 nm. Retention times and peak areas were determined using an integrator (HP3396A; Hewlett Packard Ltd., Nine Mile Ride, Wokingham, Berkshire, England, UK).

3. Sequencing of 16S rRNA gene

3.1. Preparation of biomass

The test strains were grown on glucose-yeast extract agar (Gordon & Mihm, 1962; Appendix A) plates for 7 days at 30 °C. Biomass was scrapped from the surface of the medium by using plastic loops, carefully transferred to sterile 1.5ml test tubes, and either immediately used for DNA purification or kept at -20 °C until required.

3.2. Reagents and basic procedures

Unless otherwise specified, all solutions were prepared from dilutions of stock solutions of the main reagents. Molecular biology grade reagents and enzymes, including antibiotics, enzyme inducers and substrates, and lytic, modifying and restriction enzymes, were obtained from commercial suppliers (Boehringer Mannheim Biochemical, 1994; Sigma Chemical Company, 1994). Stock solutions were prepared according to Sambrook *et al.* (1989). All buffers and solutions were made up using autoclaved distilled and deionised Milli-Q reagent grade water (Millipore [UK] Ltd., Watford, England, UK) and stored in autoclaved glass bottles. Disposable plasticware and glassware were either autoclaved or oven-baked to eliminate possible contamination with nucleases (Sambrook *et al.*, 1989).

The sequencing method was based on the blunt-end cloning procedure (Promega Co. Southampton, England, UK). The steps involved in sequencing the 16S rRNA genes are summarised in Figure 2-5.

Preparation of competent cells. *Escherichia coli* strain JM109 (Promega Co.) was grown in 300 ml of LB broth (Appendix A) in a flask at 37 °C until an OD_{600} value of about 600 was obtained. The flask was then placed in ice-cold water for 5 minutes

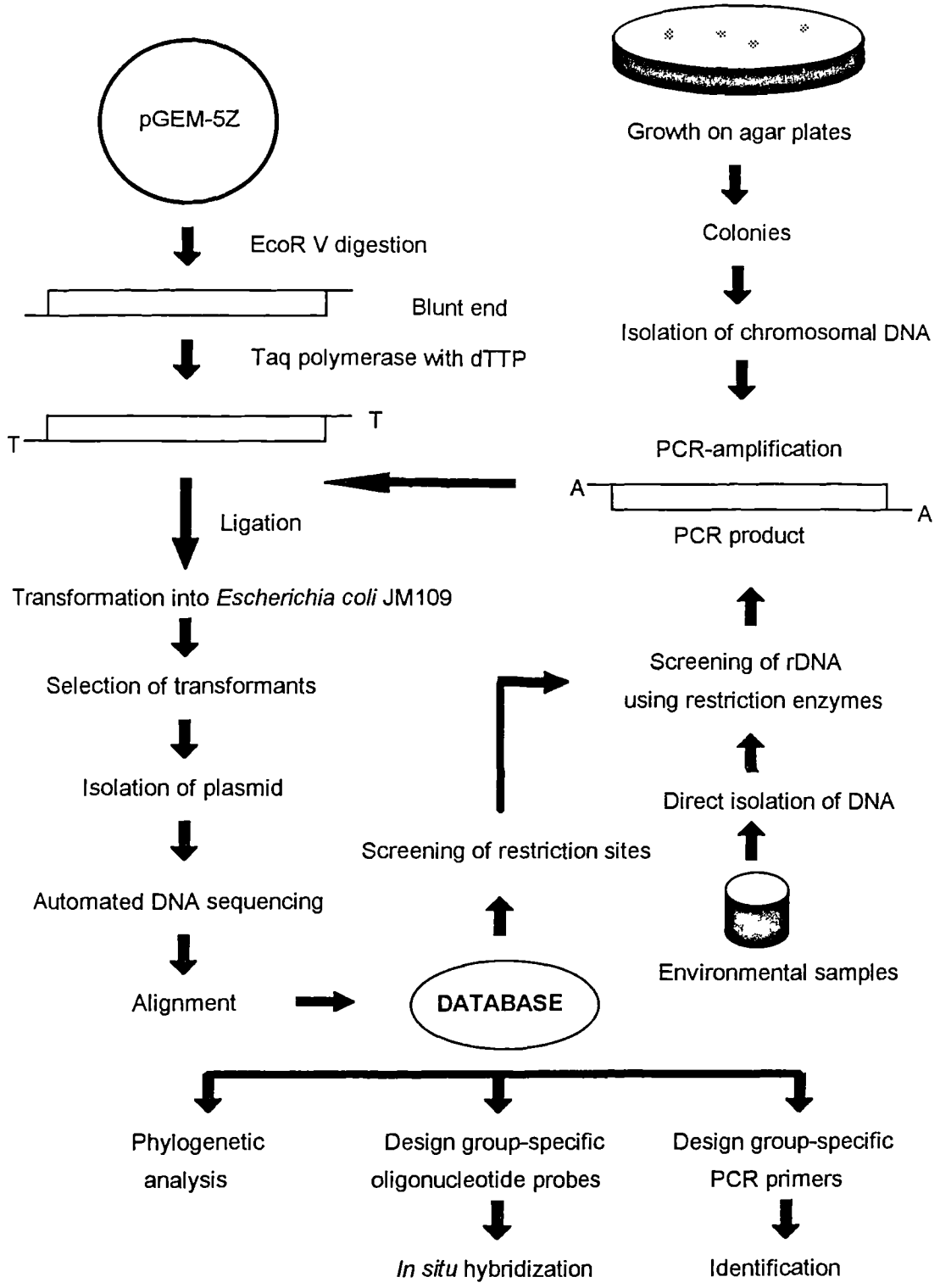


FIGURE 2-5. THE MAJOR STEPS IN SEQUENCING 16S rRNA GENES.

and centrifuged at 12,000×g for 10 minutes at 4 °C. The resultant pellet was resuspended in 100 ml of chilled TFB I buffer (Appendix A) by gentle shaking on ice, reharvested by centrifugation and resuspended in 20 ml of chilled TFB II buffer (Appendix A). The resultant bacterial suspension was aliquoted by 200 µl into chilled sterile 0.8 ml tubes, snap-freezed in a mixture of ethanol and dry-ice and stored at -70 °C.

Preparation of pGEM-T vector. The pGEM-T vector was initially purchased from the manufacturer (Promega Co.) but was later prepared in the laboratory using the following protocol. Approximately 50µg of purified pGEM-5Zf(+) vector (Promega Co.) was digested with 50 units of *EcoRV* restriction enzyme (Boeringer Mannheim Ltd.) at 37 °C overnight. Complete digestion of the plasmids was confirmed by using 0.8 %, w/v agarose gel electrophoresis and the resultant linearised plasmids further purified using an Ultrafree-MC filter unit (0.45 µm; Millipore Ltd., Watford, England, UK) following the manufacturer's protocol.

A reaction mixture (total volume 100 µl) was prepared with the solution containing the linearised pGEM-5 plasmid, *Taq* polymerase buffer (10 %), *Taq* polymerase (2.5 units; Hoefer Scientific Instruments, Newcastle upon Tyne, England, UK) and dTTP (1 mM) in a 0.8 ml tube. One drop of mineral oil was added to the tube which was then incubated at 72°C for 3 hours. The oil was removed carefully from the tube and the resultant linearised plasmid with T-overhangs (pGEM-T vector; Figure 2-5) extracted with phenol-chloroform, precipitated in the presence of 3 volumes of ethanol, washed twice with 70% ethanol and dried under vacuum. The pellet was resuspended in 50µl of distilled water and the amount of vector determined spectrophotometrically at 260 nm (Sambrook *et al.*, 1989). The solution was diluted with water to give a final concentration of 50 ng/ml prior to storage at -20 °C.

3.3. Isolation of chromosomal DNA

Chromosomal DNA was isolated using a method slightly modified from Pitcher *et al.* (1985). A small amount of biomass, that is, approximately the size of a rice grain, was taken from the GYEA plate and placed in a 1.5 ml tube containing 100 µl of lysozyme solution (50 mg/ml in TE buffer; pH 8.0). The solution was homogenised thoroughly using a sterile plastic loop and incubated at 37 °C overnight. Guanidine-Sarcosyl solution (500 µl; Appendix A) was then added and the mixture left at 37°C for 10 minutes. The lysate was cooled in ice for 5 minutes and 250µl of cold 7.5M ammonium acetate (Sambrook *et al.*, 1989) added. The contents of the tube were mixed, kept for 10 minutes in ice when 500µl of chloroform/2-pentanol (24:1, v/v) was added. The resultant solution was thoroughly mixed by hand, centrifuged at 13,000 rpm for 5 minutes and the supernatant transferred to a fresh tube followed by the addition of 0.54 volume of ice-cold isopropanol.

The contents of the tube were mixed, kept at -20 °C for at least 30 minutes and centrifuged at 13,000 rpm for 10 minutes. The pellet was washed twice with 70 % aqueous ethanol and dried under vacuum. It was then resuspended in 90 µl of TE buffer and 10 µl of RNase A solution (10 mg/ml; Sambrook *et al.*, 1989), incubated at 37 °C overnight and the same amount of phenol solution added. The preparation was then thoroughly mixed, centrifuged at 13,000 rpm for 5 minutes, the supernatant transferred to another tube, further extracted using the same amount of chloroform, the DNA precipitated by adding 2.5 volume of cold ethanol followed by centrifugation at 13,000 rpm for 10 minutes. The pellet was washed twice with 70% aqueous ethanol, dried under vacuum, resuspended in 30 to 100µl of Milli Q water and stored at -20 °C.

3.4. Amplification of 16S rRNA genes

Approximately 100 ng of the purified chromosomal DNA was used for polymerase chain reaction (PCR). Two universal primers (27f and 1525r; Table 2-14) were used to amplify almost all of the 16S rDNA. The reaction mixture was prepared as follows:

Taq polymerase buffer x 10	10 µl
dNTP mixture (25mM for each dNTP)	0.8 µl
Template chromosomal DNA	100 ng
Milli Q water	up to 100 µl

Taq polymerase (2.5 units; Hoefer Scientific Instruments) and a drop of mineral oil were then added to the tube and the mixture subjected to the following PCR programme using a thermal cycler (Omnigene; Hybaid Ltd., Middlesex, England, United Kingdom):

Steps		Temperature	Time
Initial extensive denaturation (1 cycle)		94 °C	2 minutes
Main amplification (30 cycles)	Denaturation	94 °C	1 minute
	Annealing	55 °C	1 minute
	Extension	72 °C	3 minutes
Final extension (1 cycle)		72 °C	10 minutes
Cool down (1 cycle)		30 °C	1 minute

The tube was kept at 4 °C and the presence of the 1.5kb long DNA fragment coding for 16S rRNA confirmed using 0.8 %, w/v agarose gel electrophoresis.

TABLE 2-14. OLIGONUCLEOTIDE PRIMERS USED IN POLYMERASE CHAIN REACTION AMPLIFICATION AND SEQUENCING OF 16S rDNA.

Primer	Sequence (5' to 3')[a]	Size	Binding site[b]		Usage[c]		Source
			5'	3'	PCR	Seq	
27f	AGAGTTTGATCMTGGCTCAG	20	8	27	√		Lane (1991)
MG2f	GAACGGGTGAGTAACACGT	19	107	125		√	This study
MG3f	CTACGGGRSGCAGCAC	16	342	357		√	Lane (1991)
MG4f	AATTCCTGGTGTAGCGGT	18	675	692		√	This study
782r	ACCAGGGTATCTAATCCTGT	20	801	782		√	This study
MG5f	AAACTCAAAGGAATTGACGG	20	907	926		√	This study
MG6f	GACGTCAAGTCATCATGCC	19	1190	1208		√	This study
1525r	AAGGAGGTGWTCCARCC	17	1544	1525	√		Lane (1991)

[a]Degeneracies according to IUB code (Appendix B).

[b]Binding site on the 16S rRNA molecule. Numbering according to the *Escherichia coli* system (Brosius *et al.*, 1978; Appendix C).

[c]PCR, primers used in PCR amplification of 16S rDNA; Seq, primers used in dye-deoxy cycle sequencing of cloned 16S rDNA.

3.5. Isolation of 16S rDNA

The resultant PCR products (*ca.* 95 µl) were mixed with 10 µl of loading buffer (Sambrook *et al.*, 1989) and added to a 5 cm-long well on an 0.8 %, w/v agarose gel containing ethidium bromide (0.5 µg/ml; Sambrook *et al.* 1989). Gel electrophoresis was carried out in a horizontal submarine Pharmacia gel electrophoresis apparatus. A Multidrive XL power supply (Pharmacia LKB Biochem Ltd., Science Park, Milton Road, Cambridge, England, UK) was used to apply a voltage of 100V for an hour. The band containing the 16S rDNA was visualised under UV light, isolated using a clean knife and DNA extracted from the gel slice using the Ultrafree-MC filter unit (0.45 µm; Millipore Co. Watford, England, UK) according to the manufacturer's instruction.

3.6. Cloning of 16S rDNA

Ligation of PCR-amplified 16S rDNA into the pGEM-T vector was achieved following the standard method (Sambrook *et al.*, 1989).

A ligation reaction mixture (10 µl) was prepared as follows:

T4 DNA Ligase 10x buffer (Promega Co.)	1 µl
T4 DNA Ligase (Promega Co.)	1 µl
pGEM-T Vector (50 ng/µl)	1 µl
PCR product	x µl (100-300 ng)
Milli Q water	up to 10 µl

The mixture was incubated at 15 °C for 3 hours, heated for 10 minutes at 70 °C, allowed to cool to room temperature and either examined immediately or stored at -20 °C. Two µl of the mixture containing the ligated plasmid was transferred to a fresh 0.8 ml tube and 50 µl of high-efficiency competent *Escherichia coli* JM109 suspension added. The preparation was then flicked gently to mix it, kept in ice for

20 minutes and heat-shock carried out using either a heating block or a PCR thermal cycler at 42 °C for 45 to 50 seconds. The preparation was returned and kept in ice for 2 minutes, transferred to a 1.5 ml tube containing 450 µl of sterile LB broth and incubated at 37 °C for 1 hour. A 50 µl aliquot of transformed *Escherichia coli* cells was dispersed over an LB agar plate supplemented with ampicillin (100 µg/ml), isopropyl-ß-D-thiogalactopyranoside (IPTG) and 5-bromo-4-chloro-3-indolyl-ß-D-galactopyranoside (X-Gal). The plate was incubated overnight and kept at 4 °C for an hour in order to enhance the colour reaction in the colonies. The recombinants, represented by white colonies, were selected using a sterile toothpick. These inocula were used to seed flasks containing 10 ml of fresh LB broth supplemented with ampicillin (50 µg/ml), the inoculated media were incubated at 37 °C overnight.

3.7. Isolation of plasmids

The recombinant *Escherichia coli* cells were harvested after approximately 18 hours by centrifugation at 13,000 rpm for 5 minutes. The resultant pellet was resuspended in 100 µl of ice-cold Mini-Prep Lysis Buffer (Appendix A) and kept at room temperature for 5 minutes. A freshly prepared solution (200 µl) of 0.2N NaOH and SDS (1%) was added to the tube, mixed gently by inversion and incubated for 5 minutes in ice. Potassium acetate solution (150 µl; pH 4.8; Sambrook *et al.*, 1989) was added to the preparation which was mixed gently, incubated in ice for 5 minutes, centrifuged at 13,000 rpm for 10 minutes when the supernatant was carefully transferred to another tube avoiding the addition of white precipitate (chromosomal DNA). Plasmid DNA was extracted by using phenol-chloroform, precipitated by the addition of a 2.5 volume of cold ethanol, centrifuged at 13,000 rpm for 10 minutes, washed twice with 70 % ethanol and dried under vacuum. The dried pellet was redissolved with 90 µl of TE buffer; 10 µl of RNase A solution (10mg/ml) was added to the tube which was incubated at 37 °C overnight, extracted

with phenol-chloroform, precipitated with 3 volumes of ethanol, twice washed with 70 % ethanol, dried and finally resuspended in 30 µl of water. The amount of purified plasmid DNA was determined both spectrophotometrically and by agarose gel electrophoresis.

3.8. Sequencing of 16S rDNA

Cloned 16S rDNA was sequenced by using a PRISM™ DyeDeoxy Terminator cycle sequencing kit (Applied Biosystems, California, USA) and an Applied Biosystems DNA sequencer (model 373A) following the manufacturer's instructions. The primers used for sequencing were M13 forward/reverse universal primers and specially designed primers for actinomycete 16S rDNA (Table 2-14).

3.9. Alignment of 16S rDNA sequences

The partial sequences obtained from the sequencing reactions were recorded on IBM-PC disks and incorporated to the ***AL16S*** program (Chapter I). The partial nucleotide sequences were aligned to get full sequences of the 16S rDNA. The original sequencing gel image was reexamined in cases where ambiguities were seen. Once complete nucleotide sequences were obtained they were aligned against possible phylogenetic neighbours using secondary structural information as implemented in the ***AL16S*** program.

3.10. Data analysis

The aligned 16S rDNA sequences were visually checked to select the homologous alignment positions. Regions which showed ambiguities or were not determined were excluded from further analysis. Most of the sequence analyses discussed in Chapter I were carried out by using the ***AL16S*** program. The ***PHYLIP*** package (Felsenstein, 1993) was used for the neighbor-joining (Satou & Nei, 1987), Fitch and Margoliash (Fitch & Margoliash, 1967) and maximum parsimony methods.

The maximum likelihood method (Felsenstein, 1981a) was also carried out by using the ***fastDNAml*** program (Olsen *et al.*, 1994b). Bootstrap analysis (Felsenstein, 1985) was achieved by using the ***SEQBOOT*** program which forms part of the ***PHYLIP*** package. The resultant tree was rerooted as required by using the ***RETREE*** program in the ***PHYLIP*** package.

Phylogenetic trees were presented both as unrooted radial trees and as rooted dendrograms. In the latter case, unrooted trees were generated and the position of the root identified by adding one or more outgroups. This procedure allows the detection of the internal tree topology without noise effects caused by adding distantly related outgroup organisms (Swofford & Olsen, 1990).

<BLANK>

C. Results

1. Chemotaxonomy

The isoprenoid quinones of the test strains were identified as menaquinones as they co-migrated with the standard (MK-7; Sigma). All of the type strains of the validly described *Nocardia* species contain hexahydrogenated menaquinones with eight isoprene units in which the two end units were cyclised (MK-8[H_4] ω-cycl). The type strain of *Tsukamurella wratislaviensis* contained dihydrogenated menaquinones with eight isoprene units (MK-8[H_2]). Details of the menaquinone profiles of the test strains are given in Table 2-15.

2. 16S rRNA gene sequences

2.1. Sequencing errors

The number of nucleotides determined in the present study together with the accession numbers for the EMBL/GenBank database are given in Table 2-15. Almost complete 16S rDNA sequences of the test strains were determined (1472-1477 nucleotides; positions between 28 and 1524 using *Escherichia coli* numbering system [Brosius *et al.*, 1978; Appendix C]).

Nine out of the fifteen 16S rRNA gene sequences were compared with corresponding nucleotide sequences independently examined by Ruimy *et al.* (1994b) and Rainey *et al.* (1995a) in order to get a notional idea of experimental error. The results of these comparative studies are summarised in Table 2-16. With a single exception, good congruence was found between all of the nucleotide sequence data. The exception involved the type strain of *Nocardia otitidiscaviarum* which had been the subject of previous 16S rRNA sequencing studies (Ruimy *et al.*, 1994b; Rainey *et al.*, 1995a). The nucleotide sequence obtained in the present study was found to be identical to the one examined by Ruimy *et al.* (1994b).

TABLE 2-15. PREDOMINANT MENAQUINONES AND NUCLEOTIDE SEQUENCE ACCESSION NUMBERS OF THE TEST STRAINS.

Laboratory number	Species	Menaquinones	Accession number	Nucleotides determined
JC51^T	*Corynebacterium glutamicum*	Not determined	Z46753	1479
N317^T	*Nocardia asteroides*	MK-8(H_4, ω-cyclic)	Z36934	1472
N318^T	*Nocardia brasiliensis*	MK-8(H_4, ω-cyclic)	Z36935	1474
N1201^T	*Nocardia brevicatena*	MK-8(H_4, ω-cyclic)	Z36928	1474
N1200^T	*Nocardia carnea*	MK-8(H_4, ω-cyclic)	Z36929	1472
N1170	*"Nocardia crassostrae"*	Not determined	Z37989	1473
N898^T	*Nocardia farcinica*	MK-8(H_4, ω-cyclic)	Z36936	1474
N1112^T	*Nocardia nova*	MK-8(H_4, ω-cyclic)	Z36930	1472
N36^T	*Nocardia otitidiscaviarum*	MK-8(H_4, ω-cyclic)	Z46885	1472
N1164^T	*Nocardia pinensis*	MK-8(H_4, ω-cyclic)	Z35435	1477
N1116^T	*Nocardia seriolae*	MK-8(H_4, ω-cyclic)	Z36925	1472
N1202^T	*Nocardia transvalensis*	MK-8(H_4, ω-cyclic)	Z36926	1474
N1199^T	*Nocardia vaccinii*	MK-8(H_4, ω-cyclic)	Z36927	1472
JC7^T	*Tsukamurella paurometabola*	MK-9	Z46751	1474
N805^T	*Tsukamurella wratislaviensis*	MK-8(H_2)	Z37138	1474

^{T}, Type strain.

TABLE 2-16. COMPARISON OF 16S rRNA SEQUENCES OF THE TYPE STRAINS OF *NOCARDIA* SPECIES DETERMINED IN THREE INDEPENDENT STUDIES.

Species	Accession number			Disagreement (%)		
	This study (A)	Ruimy *et al.* 1994 (B)	Rainey *et al.* 1995a (C)	A *vs.* B	A *vs.* C	B *vs.*C
N. asteroides	Z36934	ND[a]	X80606	-	0.14 (2/1463)	-
N. brasiliensis	Z36935	X80591	X80608	0 (0/1393)	0.07 (1/1465)	0.07 (1/1392)
N. brevicatena	Z36928	x80600	ND	0.50 (7/1394)	-	-
N. carnea	Z36929	X80602	X80607	0.22 (3/1392)	0 (0/1463)	0.22 (3/1391)
N. farcinica	Z36936	X80595	X80610	0.14 (2/1392)	0.21 (3/1464)	0.07 (1/1390)
N. nova	Z36930	X80593	ND	0.07 (1/1389)	-	-
N. otitidiscaviarum[b]	Z46885	X80599	X80611	**0 (0/1391)**	**3.82 (56/1465)**	**3.74 (52/1391)**
N. seriolae	Z36925	X80592	ND	0 (0/1392)	-	-
N. transvalensis	Z36926	X80598	X80609	0 (0/1391)	0.07 (1/1466)	0.07 (1/1390)

[a], Not determined.

[b], The type strain of *Nocardia otitidiscaviarum*, examined by Yang and Woese (unpublished; accession number M59056), showed nucleotide sequence dissimilarities of 0.07% (1/1417) with A, 0.07% (1/1384) with B, and 3.88% (55/1417) with C.

but it differed substantially from the nucleotide sequence determined by Rainey *et al.* (1995a). The experimental error found between the nucleotide sequence determined in the present study and that reported by Rainey *et al.* (1995a) for the type strain of *Nocardia otitidiscaviarum* was 3.82%.

All of 16S rDNA sequences generated by Ruimy *et al.* (1994b) differed from the corresponding nucleotide sequences generated both in the present study and by Rainey *et al.* (1995a). Ruimy *et al.* (1994b) found only two *G* residues in the stem region of helix 6 whereas in the two corresponding studies three consecutive *G* residues were detected in helix 6. The presence of three consecutive G residues in helix 6 would be seen to be correct as this fits with the secondary structural model. The computerised sequencing gel image for *Nocardia asteroides* strain N317^T clearly shows the presence of three *G* residues in this region (Figure 2-6).

2.2. Phylogeny of the family *Nocardiaceae* and related actinomycetes

2.2.1. Dataset and alignment

Fifty-three 16S rDNA sequences were aligned manually using the ***AL16S*** program (Chapter I), their accession numbers are given in Appendix G. The 16S rDNA sequences of the *Gordona sputi* and *Rhodococcus opacus* strains showed almost identical primary structures with *Gordona aichiensis* (99.7%) and *Tsukamurella wratislaviensis* (99.6%), respectively and hence were excluded from the final dataset in order to reduce the number of sequences (see Appendix H). Positions which could not be aligned with confidence, or were incomplete (positions between 1-37, 76-94, 187-190, 199-218, 455-477, 1448-1455, and 1485-1542), were deleted from the dataset. The final aligned dataset contained information on 1396 nucleotide positions.

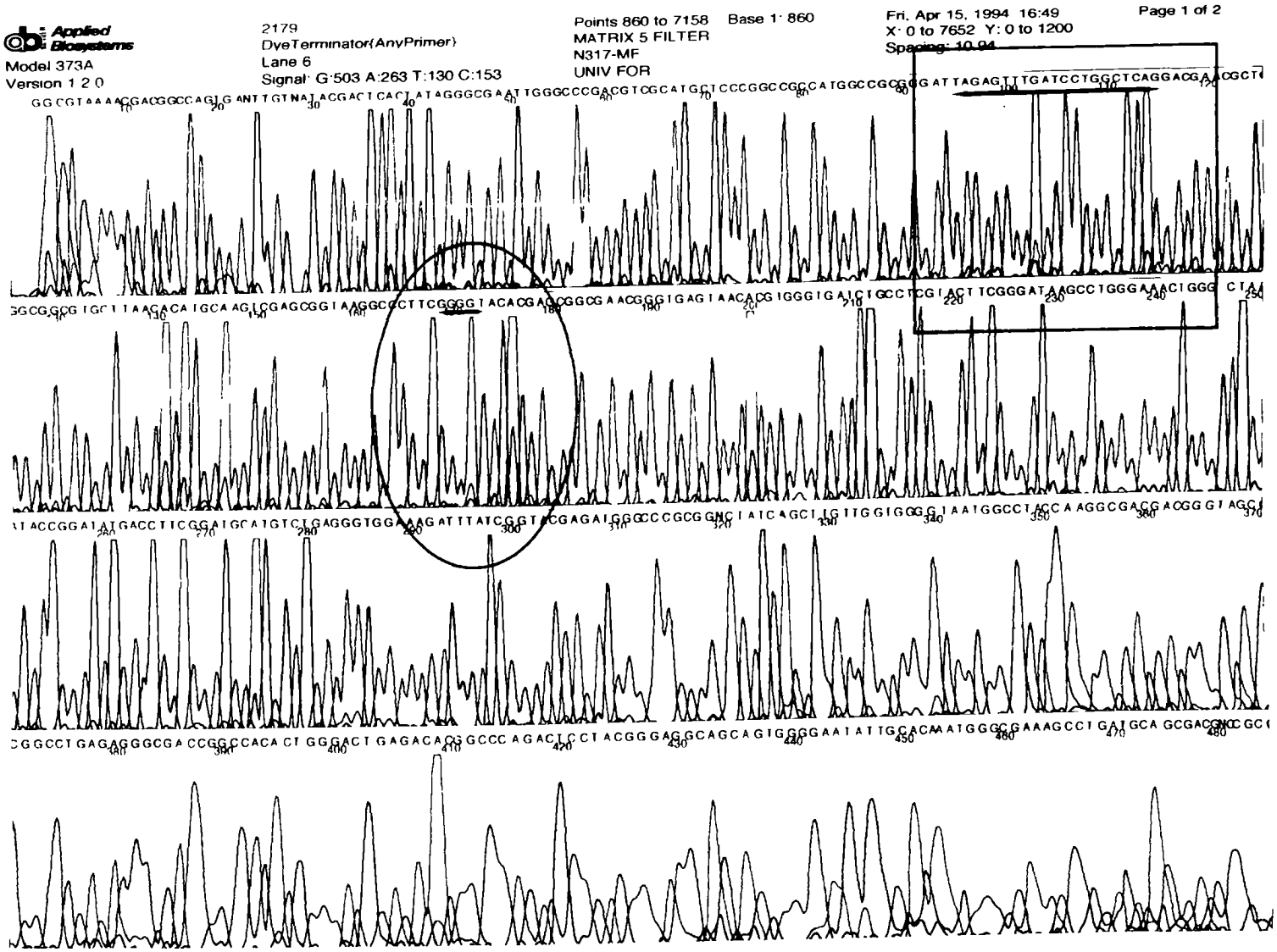


FIGURE 2-6. SEQUENCING GEL OUTPUT GENERATED FOR THE TYPE STRAIN OF *NOCARDIA ASTEROIDES* (N317^T). THE M13 FORWARD PRIMER WAS USED FOR THE SEQUENCING REACTION. THE PCR PRIMER (27F; TABLE 2-14) IS INDICATED IN THE BOX AND THE THREE G-RESIDUES OF HELIX 6 (*SENSU* NEEFS *et al.*, 1993; APPENDIX D) IN THE CIRCLE (SEE TEXT FOR DETAILS).

2.2.2. Evolutionary trees

Evolutionary trees were generated using the neighbor-joining (Saitou & Nei, 1987), least-squares (Fitch & Margoliash, 1967), maximum likelihood (Felsenstein, 1981) and maximum parsimony methods (Eck & Dayhoff, 1966; Kluge & Farris, 1969). The correction of Jukes and Cantor (1969) was applied for the distance matrix methods, that is, for the neighbor-joining and Fitch-Margoliash methods. Bootstrap evaluations were carried out for the neighbor-joining and maximum parsimony methods. It was not practical to perform bootstrap resampling for the Fitch-Margoliash and maximum likelihood methods as such analyses require excessive computing time. For convenience, all of the unrooted trees were depicted in the form of dendrograms. The position of the root of the tree was arbitrarily set between *Dietzia maris* and *Tsukamurella paurometabola*.

The neighbor-joining tree. The unrooted evolutionary tree generated using the neighbor-joining method (Saitou & Nei, 1987) is shown in Figure 2-7 together bootstrap (BP) values based on 1000 resamplings (Felsenstein, 1985). In most cases, the representatives of the validly described genera formed monophyletic groups with high bootstrap values. The five *Gordona* strains were recovered as a tight clade with a BP value of 100. Similarly, the seven *Mycobacterium* strains and the ten *Nocardia* strains formed clades defined by bootstrap values of 100 and 93, respectively. It is particularly interesting that the type strain of *Nocardia pinensis* was loosely associated with the *Gordona* clade showing a BP value of 80. The six *Corynebacterium* strains and the type strain of *Turicella otitidis* formed a monophyletic clade with a BP value of 100. *Dietzia maris* and *Tsukamurella paurometabola* represented separate lineages.

The *Rhodococcus* strains formed several phyletic lines one of which gave rise to the *Nocardia* clade. The type strains of *Rhodococcus coprophilus, Rhodococcus rhodochrous and Rhodococcus ruber* formed a monophyletic group with a BP value of 93. This clade was designated the *Rhodococcus rhodochrous* clade since it contains the type species of the genus, *Rhodococcus rhodochrous*. Similarly, the *Rhodococcus erythropolis* clade encompassed four *Rhodococcus* species, namely *Rhodococcus erythropolis*, *Rhodococcus fascians*, *Rhodococcus globerulus* and *Rhodococcus marinonascens*, and *Tsukamurellae wratislaviensis* albeit with a relatively lower BP value of 66. The phylogenetic affiliations of the remaining three rhodococci, *"Rhodococcus (Nocardia) corynebacteroides"*, *Rhodococcus equi* and *Rhodococcus rhodnii* are not clear.

The Fitch-Margoliash tree. The topology of the unrooted tree (Figure 2-8) generated by using the least squares method of Fitch and Margoliash (1967) showed few differences when compared with the neighbor-joining tree. The branching points of *"Rhodococcus (Nocardia) corynebacteroides"* and *Rhodococcus rhodnii* differed from those seen in the neighbor-joining tree.

The maximum likelihood tree. The product of the maximum likelihood method (Felsenstein, 1981a) is shown in Figure 2-9. The branching patterns of the major clades were more or less the same as those shown in the trees based on the neighbor-joining and Fitch-Margoliash methods. The clades corresponding to the genera *Corynebacterium/Turicella*, *Gordona*, *Mycobacterium* and *Nocardia* were recovered though the branching patterns within the clades were different to those seen in the neighbor-joining and Fitch-Margoliash trees. The relationships of *Dietzia maris*, *Nocardia pinensis*, *Tsukamurella paurometabola* to the major clades were

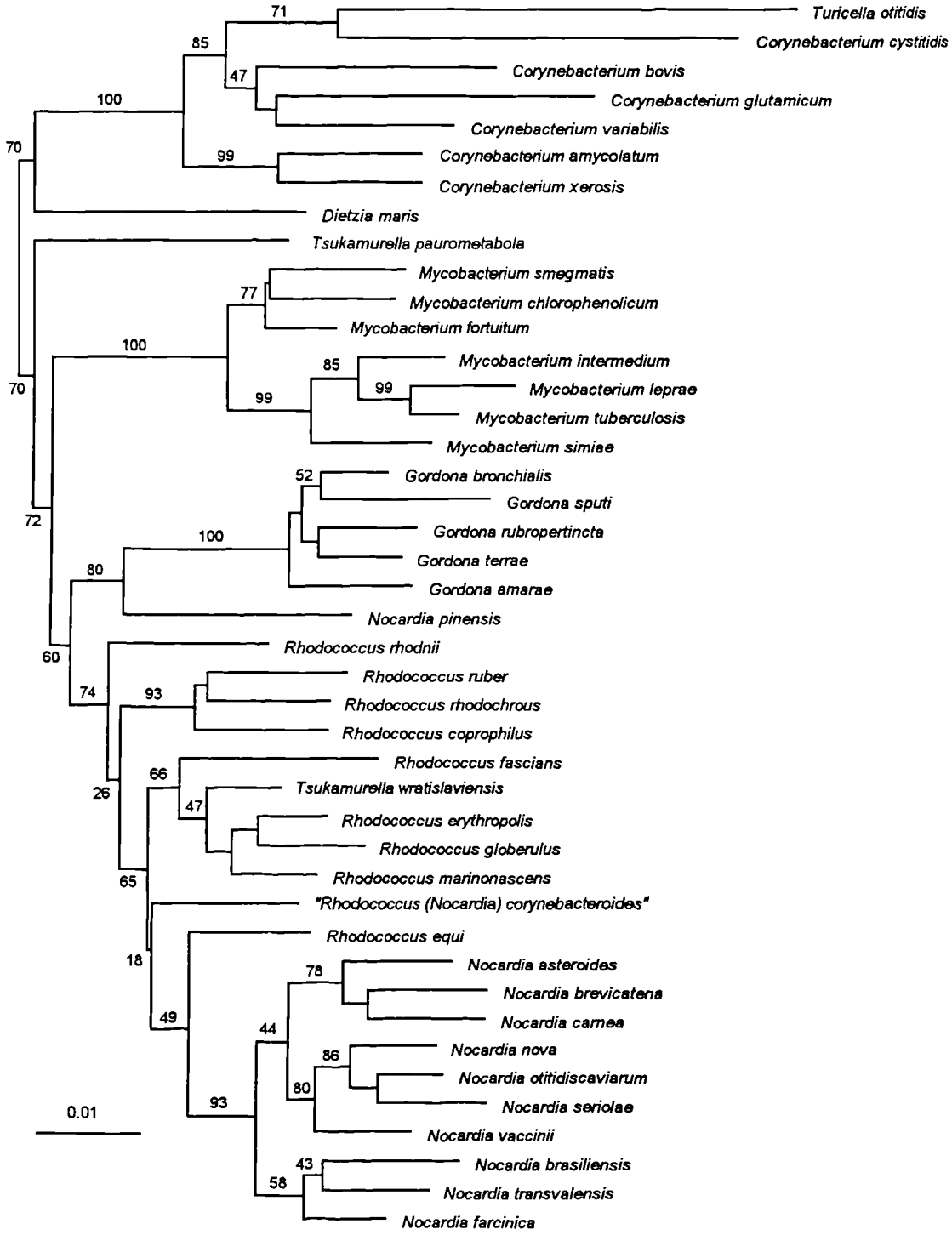


FIGURE 2-7. AN UNROOTED EVOLUTIONARY TREE DERIVED FROM THE 16S rDNA SEQUENCES. THE TREE WAS GENERATED BY USING JUKES AND CANTOR'S CORRECTION (JUKES & CANTOR, 1969) AND THE NEIGHBOR-JOINING METHOD (SAITOU & NEI, 1987).

FIGURE 2-8. AN UNROOTED EVOLUTIONARY TREE DERIVED FROM 16S rRNA SEQUENCES. THE TREE WAS GENERATED BY USING THE JUKES AND CANTOR DISTANCE METHOD (JUKES & CANTOR, 1969) AND THE LEAST SQUARES ALGORITHM (FITCH & MARGOLIASH, 1967).

FIGURE 2-9. AN UNROOTED MAXIMUM LIKELIHOOD TREE DEPICTING RELATIONSHIPS BETWEEN REPRESENTATIVES OF THE MYCOLIC ACID-CONTAINING TAXA.

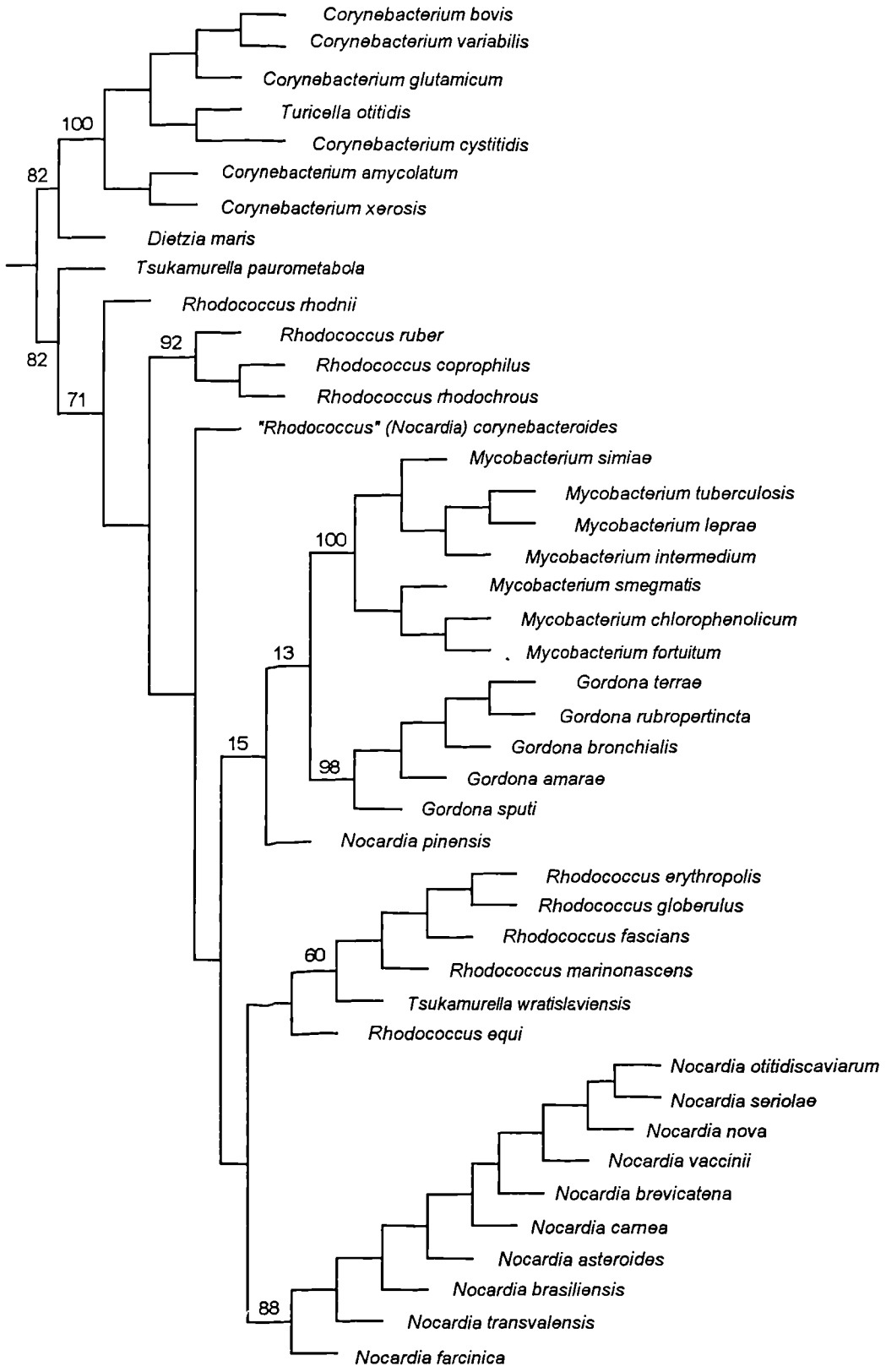


FIGURE 2-10. THE MOST PARSIMONIOUS TREE GENERATED BY USING THE MAXIMUM PARSIMONY CRITERION; THE BRANCH LENGTHS ARE INVALID (FELSENSTEIN, 1993).

also seen in the product of the maximum likelihood analysis. The *Rhodococcus erythropolis* and *Rhodococcus rhodochrous* clades were again evident. However, the branching points of *Rhodococcus equi*, *Rhodococcus rhodnii*, "*Rhodococcus (Nocardia) corynebacteroides*" differed from those found in the other two trees.

Maximum parsimony tree. The unrooted phylogenetic tree based on the maximum parsimony criterion (equivalent to Wagner parsimony; Eck & Dayhoff, 1966; Kluge & Farris, 1969) was generated by using the ***DNAPARS*** program in the ***PHYLIP*** package (Felsenstein, 1993). Five of the trees that were generated were found to be equally most parsimonious; the consensus tree is given in Figure 2-10. It was not possible to derive the lengths of the branches that indicate evolutionary distances hence only the topology found in this analysis was examined.

The major phylogenetic clades corresponding to *Corynebacterium*/*Turicella* (BP value 100), *Gordona* (BP value 98), *Mycobacterium* (BP value 100) and *Nocardia* (BP value 88) were, once again, evident. *Nocardia pinensis* was loosely associated with the phyletic line formed by the *Gordona* and *Mycobacterium* strains. The positions of *Dietzia maris* and *Tsukamurella paurometabola* were the same as in the earlier analyses. The *Rhodococcus erythropolis* (BP value 60) and *Rhodococcus rhodochrous* (BP value 92) clades were again recovered; "*Rhodococcus (Nocardia) corynebacteroides*" and *Rhodococcus rhodnii* formed separate lineages. In general, it can be said that the bootstrap values obtained using the neighbor-joining and maximum parsimony methods are in good agreement.

Placement of the root. A variety of outgroup strains were used with the neighbor-joining method to establish the position of the root in the evolutionary tree; the branching point in the original tree in relation to the outgroup(s) is regarded as the position of the root (Swofford & Olsen, 1990). The phylogenetic tree depicting relationships found between the test strains and the outgroup strains is shown in Figure 2-11. The nearest neighbors to the mycolic-acid containing actinomycetes were *Actinoplanes* strains and members of the family *Pseudonocardiaceae* (Embley & Stackebrandt, 1994); the most distant outgroup was *Bacillus subtilis*. The root positions identified using the various outgroup strains are summarised in Table 2-17.

It is evident from Table 2-17 that the position of the root in the evolutionary tree of the mycolic acid-containing actinomycetes cannot be determined with confidence as there is little congruence between the root positions identified by using the different outgroup(s). Nevertheless, a number of interesting points are evident, *(i)* root position 1, that is, the root position between the *Corynebacterium/Turicella* clade and *Dietzia maris*, was identified using all of the taxonomically distantly related outgroups; *(ii)* several root positions were found using the more closely related outgroups; *(iii)* root positions determined by an individual outgroup organism and by pairs of organisms may differ (e.g., "*Streptomyces coelicolor*" and *Kitasatosporia setae*); *(iv)* different root positions were found using individual members of the same genus (e.g., *Saccharomonospora azurea*, *Saccharomonospora cyanea* and *Saccharomonospora viridis*); and *(v)* the use of multiple outgroups favoured root position 4, that is, the root position between the *Gordona/Nocardia pinensis* and *Mycobacterium* clades.

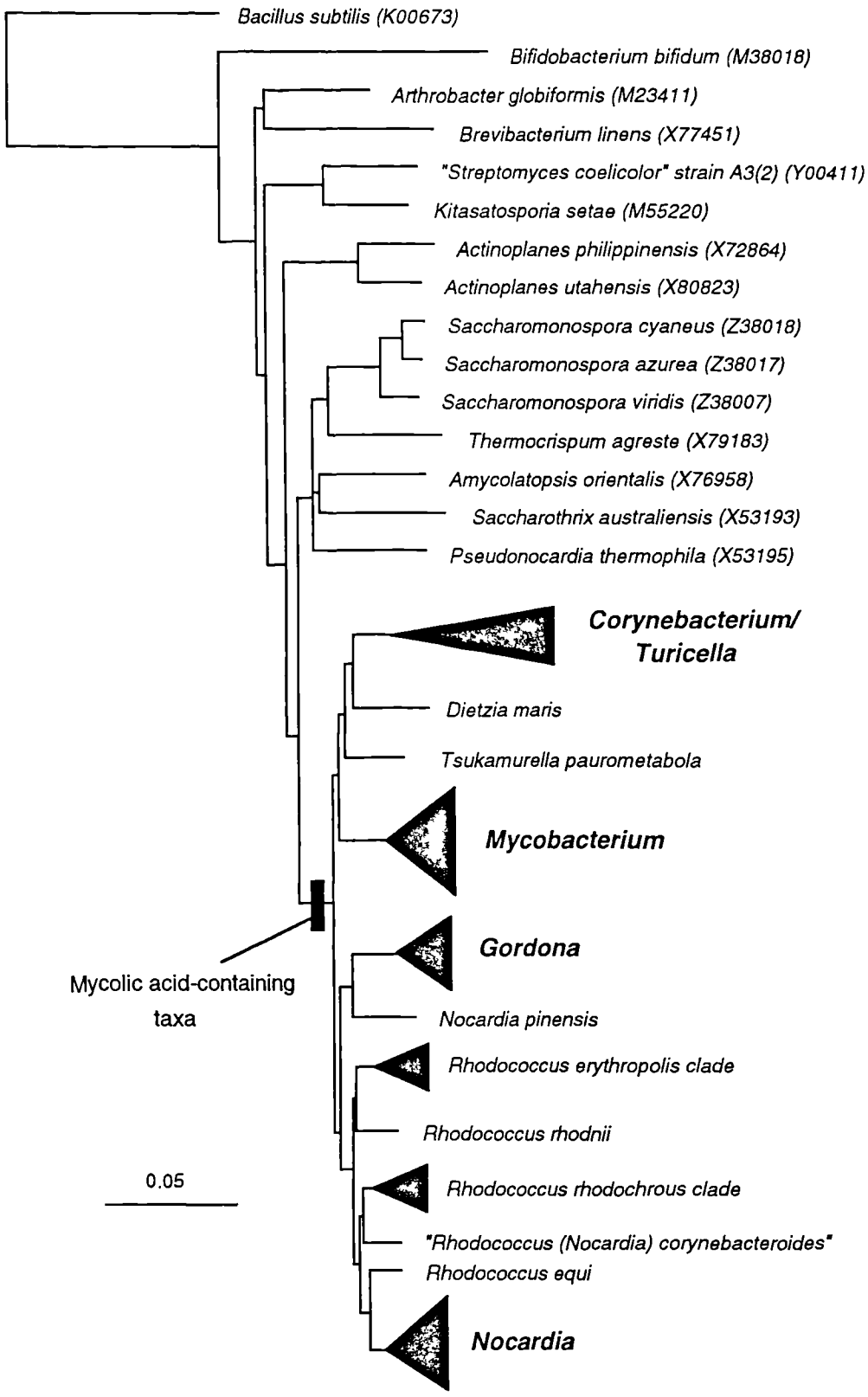
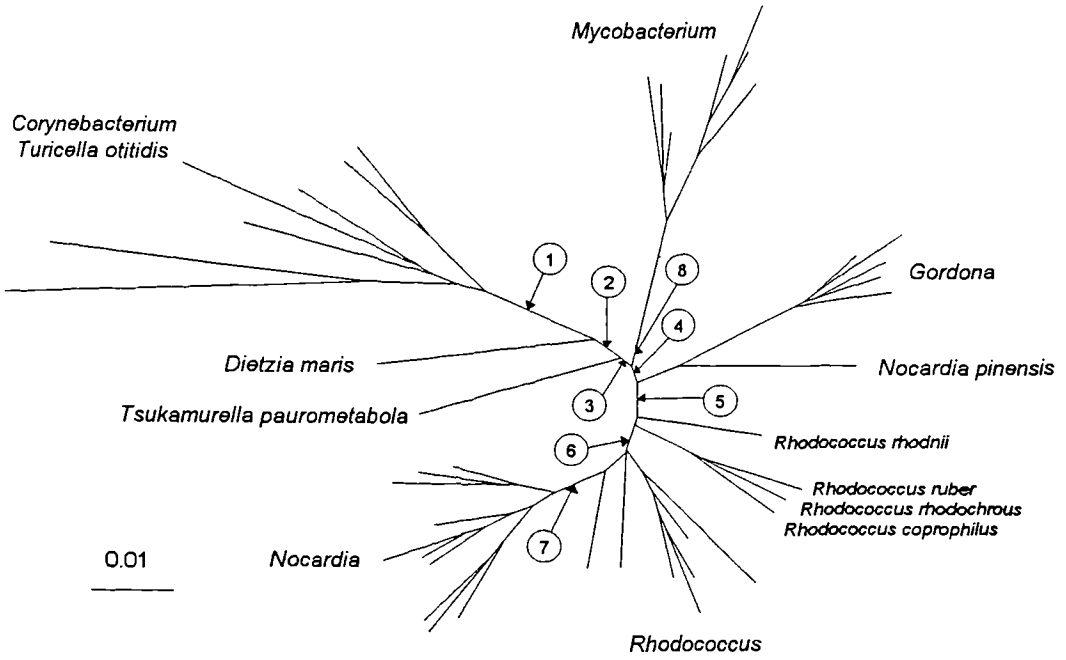


FIGURE 2-11. A PHYLOGENETIC TREE DEPICTING RELATIONSHIPS BETWEEN REPRESENTATIVES OF MYCOLIC ACID-CONTAINING TAXA AND SELECTED OUTGROUP STRAINS. THE ACCESSION NUMBERS OF THE OUTGROUP STRAINS ARE GIVEN IN PARENTHESIS.

TABLE 2-17. EFFECTS OF DIFFERENT OUTGROUP STRAINS ON POSITIONS OF THE ROOT IN PHYLOGENETIC TREES BASED ON MYCOLIC ACID-CONTAINING AND RELATED ACTINOMYCETES.



Position of root	Outgroup(s)*
1	*Arthrobacter globiformis, Bacillus subtilis, Bifidobacterium bifidum, Brevibacterium linens*
2	*Saccharomonospora cyanea*
3	*Saccharomonospora azurea*, (*Saccharomonospora azurea*+ *Saccharomonospora cyanea*+ *Saccharomonospora viridis*)
4	All of the outgroup strains shown in Figure 2-11, *Actinoplanes utahensis, Saccharomonospora viridis, Thermocrispum agreste, (Amycolatopsis orientalis+Saccharomonospora viridis+Pseudonocardia+thermophila +Saccharothrix australiensis+Thermocrispum agreste* [All of the family *Pseudonocardiaceae* and related taxa])
5	*Amycolatopsis orientalis*, ("*Streptomyces coelicolor*"+*Kitasatosporia setae*), (*Actinoplanes philippiensis*+*Actinoplanes utahensis*)
6	"*Streptomyces coelicolor*", *Kitasatosporia setae*
7	*Actinoplanes philippiensis*
8	*Pseudonocardia thermophila, Saccharothrix australiensis*

*, Multiple outgroups are indicated in parentheses.

2.2.3. Comparative nucleotide sequence analyses

16S rRNA sequence similarities. The results obtained using the modules '***Intragroup similarity***' and '***Intergroup similarity***' of the ***AL16S*** program are summarised in Table 2-18. The smallest mean pairwise sequence similarity found within a group was 93.8% for the *Corynebacterium/Turicella* clade; the *Gordona* clade showed the highest intragroup similarity, namely 97.8%. The intragroup similarities for the genera *Mycobacterium*, *Nocardia* and *Rhodococcus* are around 97%. The intergroup similarity values, that is, the mean of the pairwise sequence similarities between members of phyletic lines ranged from 90.9% (*Corynebacterium/Turicella* and *Nocardia*) to 95.4% (*Nocardia* and *Rhodococcus*). *Nocardia pinensis* was most closely related to the *Gordona* clade (95.2%). The closest group to the *Corynebacterium/Turicella* clade was *Dietzia*, a result in good agreement with the phylogenetic relationship depicted in the four unrooted evolutionary trees (Figures 2-7, 2-8, 2-9, 2-10).

Regional sequence variations in the individual phyletic lines. The module '***Regional similarity***' of the ***AL16S*** program (Chapter I) was used to detect the extent of nucleotide sequence variations in different parts of the 16S rRNA molecules of strains in the phyletic lines corresponding to the genera *Corynebacterium/Turicella*, *Gordona*, *Nocardia* and *Rhodococcus*. Variations were calculated as the mean value of pairwise sequence similarities within each of these groups for particular regions of 16S rRNA molecules. It is evident from Figure 2-12 that members of the different phylogenetic groups showed essentially similar patterns of nucleotide sequence variation.

However, considerable differences were found in the region between positions 1 and 100 (*Escherichia coli* numbering system; Appendix C) for the *Corynebacterium/Turicella* and *Nocardia* clades. This variation was mainly

TABLE 2-18. MEAN 16S rRNA SEQUENCE SIMILARITIES FOUND BETWEEN DIFFERENT PHYLETIC LINEAGES (INTERGROUP SIMILARITY).

	Corynebacterium/ Turicella	*Dietzia*	*Gordona*	*Mycobacterium*	*Nocardia*	*Nocardia pinensis*	*Rhodococcus*	*Tsukamurella*
Number of organisms	7	1	6	7	10	1	11	1
Mean intragroup similarity	93.8	-	97.8	97.1	97.0	-	96.7	-
Dietzia	92.7	100.0						
Gordona	91.5	93.5	100.0					
Mycobacterium	91.6	93.3	93.5	100.0				
Nocardia	90.9	92.9	93.9	93.7	100.0			
N. pinensis	91.8	93.2	95.2	94.0	94.8	100.0		
Rhodococcus	92.2	94.6	94.1	93.9	95.4	95.0	100.0	
Tsukamurella	92.2	95.1	94.0	93.9	93.9	93.5	94.9	100.0

*, Calculations based on 1396 nucleotide positions.

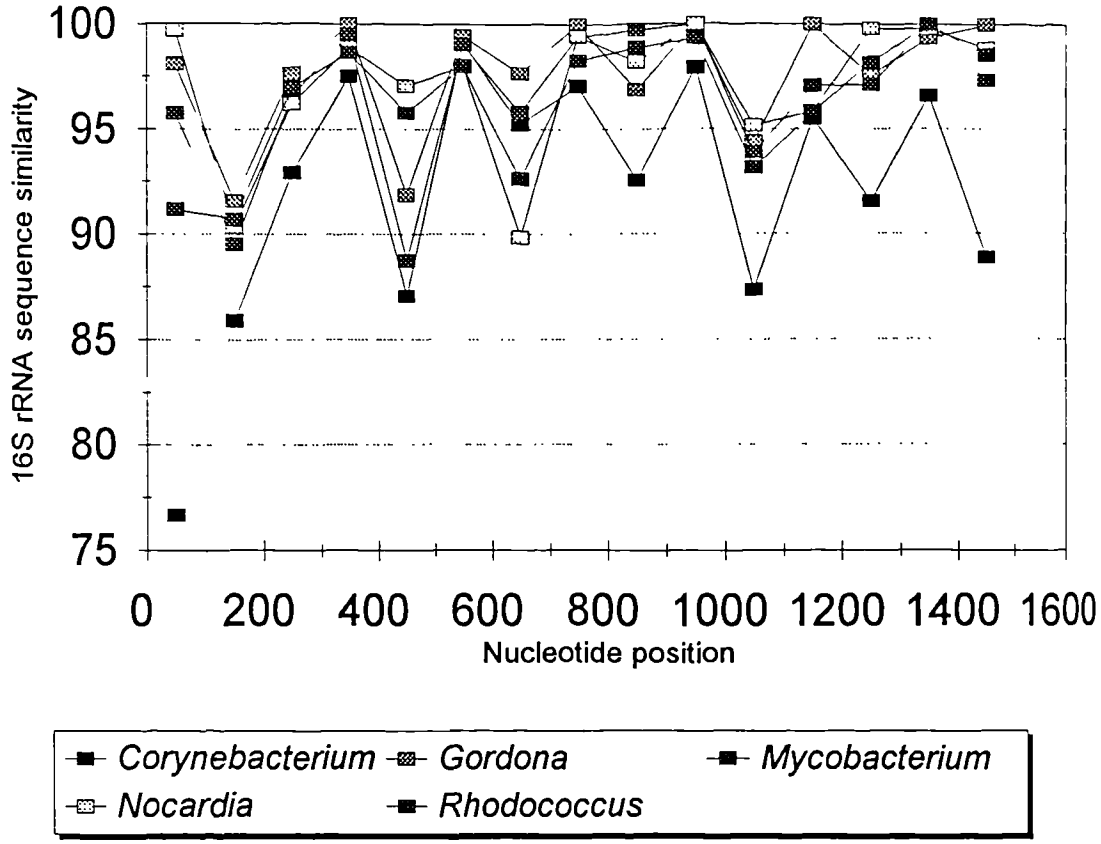


FIGURE 2-12. SUMMARY OF THE ***REGIONAL VARIATION*** ANALYSIS ON FIVE PHYLETIC CLADES. THE MEAN VALUE OF PAIRWISE SEQUENCE SIMILARITIES BETWEEN MEMBERS OF A GIVEN CLADE WAS DETERMINED FOR THE GIVEN REGION OF 16S rRNA MOLECULE BY USING THE '***REGIONAL SIMILARITY***' MODULE OF THE ***AL16S*** PROGRAM (CHAPTER I). SEE TEXT FOR DETAILS.

located in the V1 region (Neefs *et al.*, 1993; Appendix D) where all ten *Nocardia* strains showed identical primary structures, hence an identical secondary structure, whereas members of the *Corynebacterium/Turicella* lineage exhibited a remarkably different primary structure. A similar phenomenon was observed between the *Corynebacterium/Turicella* and *Nocardia* clades in three other regions, namely between positions 401 and 500, 1001 and 1100, and 1201 and 1300. However, members of the *Nocardia* clade showed the greatest nucleotide variation in the region of the macromolecule between positions 601 and 700.

Group signatures. The module '***Group Signature***' of the ***AL16S*** program (Chapter I) was used to determine signature nucleotides that differentiated between the *Corynebacterium/Turicella*, *Dietzia*, *Gordona*, *Mycobacterium*, *Nocardia*, *Nocardia pinensis*, *Rhodococcus* and *Tsukamurella* clades. The results of this analysis are summarised in Table 2-19. This information is analogous to that in frequency matrices derived from numerical phenetic studies.

Some of the well conserved nucleotide positions are unique to particular phyletic clades: *(i)* paired positions 316:337 (T:G) and 408:434 (G:T) for the *Corynebacterium/Turicella* clade; *(ii)* paired positions 418:425 (T:A) and 987:1218 (A:T) for *Dietzia maris*; *(iii)* unpaired position 843 (T) for the *Gordona* line; and *(iv)* paired position 580:761 (T:A) and independent position 1336 (T) for the *Nocardia* clade. Positions specific for the *Corynebacterium/Turicella* and *Dietzia* clades were 771:808 (A:T) and 1059:1198 (T:A). The phylogenetic relatedness of *Nocardia pinensis* to the *Gordona* clade was supported by a distinctive nucleotide signature found in paired positions 1124:1149 (A:T).

TABLE 2-19. GROUP SPECIFIC NUCLEOTIDE SIGNATURES THAT DIFFERENTIATE BETWEEN THE PHYLOGENETIC LINEAGES DETECTED WITHIN REPRESENTATIVES OF THE MYCOLIC ACID-CONTAINING TAXA.

	Coryne-bacterium /Turicella	*Dietzia*	*Gordona*	*Mycoba-cterium*	*Nocardia*	*Nocardia pinensis*	*Rhodo-coccus*	*Tsuka-murella*
No. of sequences/ Positions[a]	7[b]	1[c]	6[d]	56[e]	12[f]	1[g]	12[h]	1[i]
66:103	A:T	A:T	A:T	A:T	G:C	A:T	R*Y	A:T
70:98	W*W	T:A	A:T	W*W	T:A	T:A	T:A	T:A
76:94	A*C	C:G	C:G	Y*R	C:G	C:G	Y*G	C:G
293:304	G:T	G:T	A:T	G:T	R*Y	G:T	R*Y	G:C
307	A	T	T	C	Y	C	Y	C
316:337	T:G	C:G	C:G	C:G	C:G	C:G	C:G	C:G
328	Y	C	C	T	C	C	C	C
407:435	G:C	G:C	A:T	R*Y	M*K	G:T	A:T	A:T
408:434	G:T	G:C	G:C	G:C	G:C	G:C	G:C	G:C
418:425	C:G	T:A	C:G	C:G	C:G	C:G	C:G	C:G
508	T	T	C	C	Y	C	Y	C
580:761	C:G	C:G	C:G	C:G	T:A	C:G	C:G	C:G
586:755	T:G	C:G	C:G	Y*G	C:G	C:G	C:G	C:G
601:637	T:G	T:G	T:G	T:G	N*K	A:T	T:G	T:G
603:635	T:A	T:A	T:A	C:G	Y*R	C:G	Y*R	T:A
610	K	T	T	W	A	A	W	A
611	Y	C	T	B	C	C	C	C
613:627	Y*G	C:G	T:A	S*S	B*V	C:G	H*D	C:G
616:624	G:T	G:C	A:T	R*Y	R*Y	A:T	R*Y	G:C
620	T	T	Y	Y	C	C	C	T
661:744	G:C	A:T	A:T	G:C	B*M	C*C	K*M	G:C
662:743	T:G	C:G	C:G	Y*G	C:G	C:G	C:G	T:G
771:808	A:T	A:T	G:C	G:C	G:C	G:C	G:C	G:C
824:876	C:G	C:G	T:A	W*A	Y*R	C:G	C:G	T:A
825:875	G:C	G:C	A:T	A:T	R*Y	G:C	G:C	A:T
843	C	C	T	C	C	C	C	C
987:1218	G:C	A:T	G:C	G:C	G:C	G:C	G:C	G:C
997:1044	T*R	T:A	T:A	T:G	T:A	T:A	T:A	T:A
998:1043	R*Y	A:T	A:T	G:T	A:T	A:T	R*K	A:T
1001:1039	N*B	A:T	C:G	A:T	C:G	C:G	C:G	A:T
1002:1038	R*Y	G:T	A:T	G:C	R*Y	A:T	G:C	G:T
1003:1037	G*Y	G:C	G*Y	G:C	G:T	G:T	G*Y	G:C
1005	Y	C	C	Y	A	C	H	T
1024	T	T	T	T	C	T	Y	T
1059:1198	T:A	T:A	C:G	C:G	C:G	C:G	C:G	C:G
1120:1153	T:A	T:A	C:G	Y*R	Y*R	C:G	Y*R	T:A
1122:1151	R*Y	A:T	G:C	A:T	R*Y	G:C	R*Y	A:T
1124:1149	G:C	G:C	A:T	G:C	G:C	A:T	G:C	G:C
1133:1141	A:T	A:T	G:C	R*Y	G:C	G:C	R*B	A:T
1134:1140	C:G	C:G	G:C	S*S	S*S	C:G	C:G	C:G
1189	T	T	T	T	C	C	Y	T
1256	T	C	C	C	C	T	C	C
1336	C	C	C	C	T	C	C	C

Table 2-19. Continued.

[a], *Escherichia coli* numbering (Brosius *et al.*, 1978); [b], based on sequence data from *Corynebacterium amycolatum*, *C. bovis* (D38575), *C. cystitidis* (D37914), *C. glutamicum* (Z46753), *C. variabilis* (X53185), *C. xerosis* (M59058) and *Turicella otitidis* (X73976); [c], from *Dietzia maris* (X79290); [d], from *Gordona aichiensis* (X80633), *G. amarae* (X80635), *G. bronchialis* (X79287), *G. rubropertinctus* (X80632), *G. sputi* (X80634) and *G. terrae* (X79286); [e], from *Mycobacterium aichiense* (X55598), *M. asiaticum* (X55604), *M. aurum* (X55595), *M. avium* (X52918, X52934), *M. bovis* (M20940, X55589), *M. celatum* (Z46664, L08170, L08169), *M. chelonae* (M29559, X52921), *M. chitae* (X55603), *M. chlorophenolicum* (X79094), *M. chubuense* (X55596), *M. confluentis* (X63608), *M. cookii* (X53896), *M. diernhoferi* (X55593), *M. fallax* (M29562), *M. farcinogenes* (X55592), *M. flavescens* (X52932), *M. fortuitum* (X52933), *M. gadium* (X55594), *M. gastri* (X52919), *M. geneveuse* (X60070), *M. gilvum* (X55599), *M. gordonae* (X52923), *M. haemophilum* (L24800), *M. hiberniae* (X67096), *M. intermedium* (X67847), *M. intracellulare* (X52927), *M. kansasii* (X15916, M29575), *M. komossense* (X55591), *M. leprae* (X53999), *M. madagascariense* (X55600), *M. malmoense* (X52930), *M. marinum* (X52920), *M. neoaurum* (M29564), *M. nonchromogenicum* (X52928), *M. obuense* (X55597), *M. phlei* (M29566), *M. scrofulaceum* (X52924), *M. senegalense* (M29567), *M. shimoidei* (X82459), *M. simiae* (X52931), *M. smegmatis* (X52922), *Mycobacterium* sp. (M29554), *M. sphagni* (X55590), *M. szulgai* (X52926), *M. terrae* (X52925), *M. thermoresistibile* (X55602), *M. tuberculosis* (X52917), *M. ulcerans* (Z13990), *M. vaccae* (X55601) and *M. xenopi* (X52929); [f], from *Nocardia asteroides* (Z36934, X57949), *N. brasiliensis* (Z36935), *N. brevicatena* (Z36928), *N. carnea* (Z36929), "*N. crassostrae*" (Z37989), *N. farcinica* (Z36936), *N. nova* (Z36930), *N. otitidiscaviarum* (Z46885), *N. seriolae* (Z36925), *N. transvalensis* (Z36926) and *N. vaccinii* (Z36927); [g], from *Nocardia pinensis* (Z35435); [h], *Rhodococcus coprophilus* (X80626), *R. equi* (X80614), *R. erythropolis* (X79289), *R. fascians* (X79186), *R. globerulus* (X80619), *R. marinonascens* (X80617), *R. opacus* (X80630), *R. rhodnii* (X80621), *R. rhodochrous* (X79288), *R. ruber* (X80625), "*Rhodococcus (Nocardia) corynebacteroides*" (X80615) and *Tsukamurella wratislaviensis* (Z37138); [i], from *Tsukamurella paurometabola* (Z46751).

Secondary structure. The unique secondary structure found in helix 18 (Appendix D) in the 16S rRNA sequence of the type strain of *Nocardia pinensis* is shown in Figure 2-13.

2.3. Phylogeny of the genus *Nocardia*

Dataset and alignment. Twelve *Nocardia* strains, which included the type strains of ten out of the eleven validly described species together with *Nocardia asteroides* strain DSM43005 and "*Nocardia crassostrae*" strain N1170, were examined. The 16S rDNA sequence of *Nocardia asteroides* DSM 43005 was generated by Rogall *et al.* (1990). The 16S rDNA sequences of all of the test strains were readily aligned as there were no unalignable positions; the final dataset contained 1448 nucleotide positions. The primary structure of the 16S rDNA sequences of the test strains are given in an aligned form (Appendix H). The remaining organism, the type strain of *Nocardia pinensis*, was excluded from the dataset as it has been shown to be unrelated to typical nocardiae.

Nucleotide sequence similarity. The pairwise similarity values for the 16S rDNA sequences of the twelve nocardiae are given in Table 2-20. The mean similarity recorded for all of the twelve test strains is 97.1 ± 0.7%. The corresponding value for the ten type strains is 97.0 ± 0.6%. The most phylogenetically distant pair of organisms was *Nocardia farcinica* N898^T and *Nocardia seriolae* N1116^T (95.6%), the smallest nucleotide difference, 98.4%, was found between the type strains of *Nocardia nova* and *Nocardia otitidiscaviarum*.

Phylogenetic analyses. The aligned 16S rDNA sequences of the twelve nocardiae were corrected for multiple mutations (Jukes & Cantor, 1969) and unrooted evolutionary trees generated by using the neighbor-joining (Saitou & Nei, 1987) and least squares (Fitch & Margoliash, 1967) methods. The nucleotide sequence data were also examined using the maximum likelihood (Felsenstein, 1981a) and

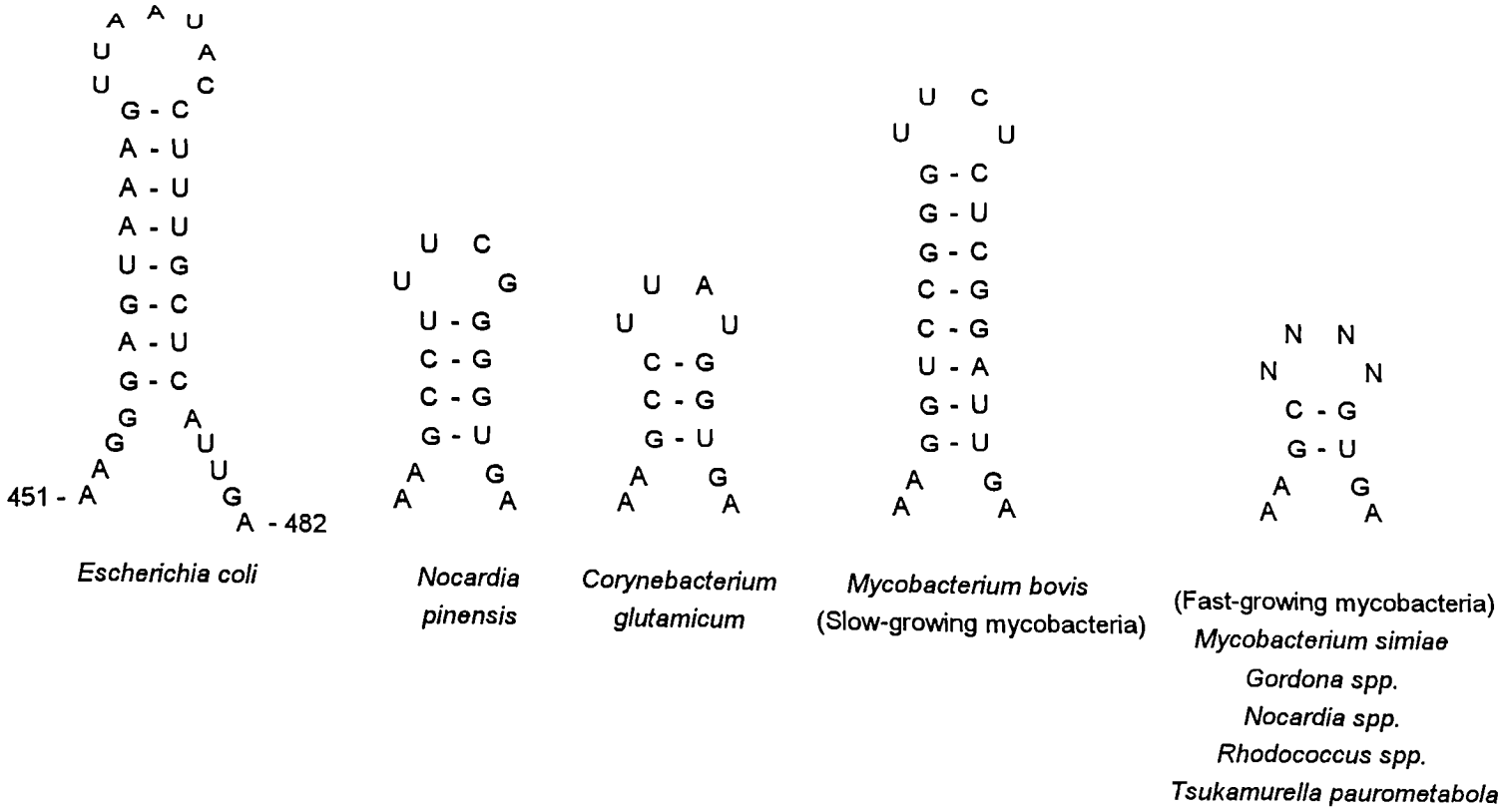


FIGURE 2-13. SECONDARY STRUCTURAL DIFFERENCES IN 16S rRNA SEQUENCES OF *NOCARDIA PINENSIS* N1164^T AND RELATED MYCOLIC ACID-CONTAINING ACTINOMYCETES. HELIX 18 (APPENDIX D) LIES BETWEEN POSITIONS 451 AND 482 OF *ESCHERICHIA COLI* NUMBERING SYSTEM (APPENDIX C).

TABLE 2-20. SIMILARITY VALUES FOR THE 16S rDNA SEQUENCES OF THE REPRESENTATIVE *NOCARDIA* STRAINS. THE NUMBER OF NUCLEOTIDE DIFFERENCES ARE GIVEN IN THE UPPER-RIGHT TRIANGLE.

	Similarity to*:											
Species	*N. asteroides*	*N. asteroide* DSM43005	*N. brasiliensis*	*N. brevicatena*	*N. carnea*	*"N. crassostrae"*	*N. farcinica*	*N. nova*	*N. otitidiscaviarun*	*N. seriolae*	*N. transvalensis*	*N. vaccinii*
N. asteroides	-----	23	34	38	29	47	36	37	46	47	38	42
N. asteroides DSM43005	98.4	-----	30	30	24	37	29	31	33	43	31	42
N. brasiliensis	97.6	97.9	-----	52	44	47	29	48	54	54	34	57
N. brevicatena	97.4	97.9	96.4	-----	32	57	54	43	46	47	47	39
N. carnea	98.0	98.3	97.0	97.8	-----	52	49	48	49	49	45	47
"N. crassostrae"	96.7	97.4	96.8	96.1	96.4	-----	58	35	32	40	43	49
N. farcinica	97.5	98.0	98.0	96.3	96.6	96.0	-----	54	58	63	31	49
N. nova	97.4	97.8	96.7	97.0	96.7	97.6	96.3	-----	23	34	37	27
N. otitidiscaviarum	96.8	97.7	96.3	96.8	96.6	97.8	96.0	98.4	-----	28	45	36
N. seriolae	96.7	97.0	96.3	96.7	96.6	97.2	95.6	97.6	98.1	-----	55	37
N. transvalensis	97.4	97.9	97.7	96.8	96.9	97.0	97.9	97.4	96.9	96.2	-----	43
N. vaccinii	97.1	97.1	96.1	97.3	96.7	96.6	96.6	98.1	97.5	97.4	97.0	-----

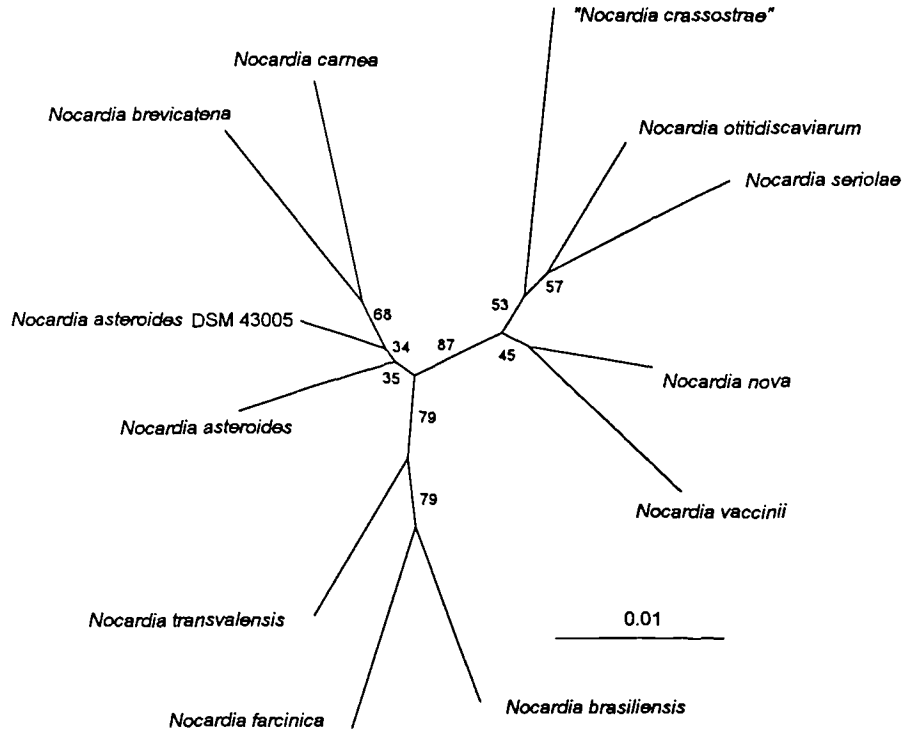
*, Comparisons are based on 1448 positions.

maximum parsimony methods (Eck & Dayhoff, 1966; Kluge & Farris, 1969). Bootstrap analyses were carried out by applying 1000 resamplings for the neighbor-joining, least-squares and maximum parsimony methods. The resultant phylogenetic trees with corresponding bootstrap values are given in Figure 2-14.

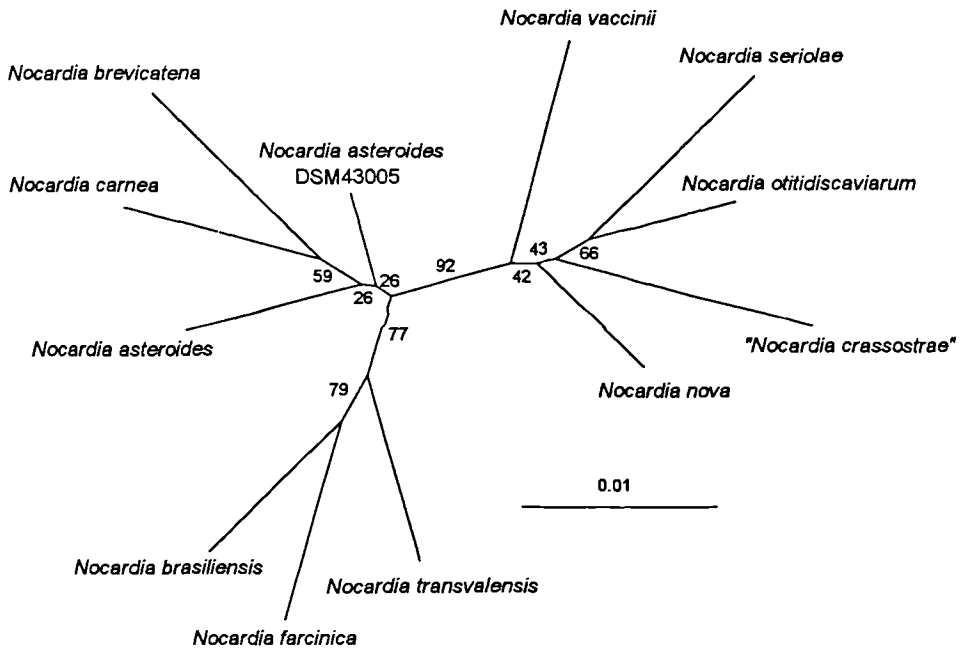
The twelve nocardiae were assigned to two rRNA subgroups in all of the phylogenetic analyses; one subgroup corresponded to *Nocardia asteroides* and related taxa and the other to *Nocardia otitidiscaviarum* and allied species (Figure 2-14). The validity of the two phyletic lines was supported by high bootstrap values for the maximum parsimony, neighbor-joining and Fitch-Margoliash methods, namely 84, 87 and 92, respectively.

The *Nocardia asteroides* subgroup encompassed *Nocardia asteroides* strains N317^T and DSM43005, *Nocardia brevicatena* N1201^T, *Nocardia carnea* N1200^T, *Nocardia brasiliensis* N318^T, *Nocardia farcinica* N898^T and *Nocardia transvalensis* N1202^T. The type strains of *Nocardia brasiliensis, Nocardia farcinica* and *Nocardia transvalensis* formed a subclade in all four trees with relatively low bootstrap values for the maximum parsimony, Fitch-Margoliash and neighbor-joining methods, that is, 74, 77 and 79, respectively. The *Nocardia otitidiscaviarum* subgroup encompassed "*Nocardia crassostrae*" strain N1170, *Nocardia nova* N1112^T, *Nocardia otitidiscaviarum* N36^T, *Nocardia seriolae* N1116^T and *Nocardia vaccinii* N1199^T. These species were recovered as a distinct clade in all four analyses though branching patterns inside the subgroup varied (Figure 2-14).

The separation of the two rRNA groups was largely based on differences found in helix 37-1 (Appendix D), as shown in Figure 2-15. Nucleotide signatures that differentiate the nocardiae are shown in Table 2-21.

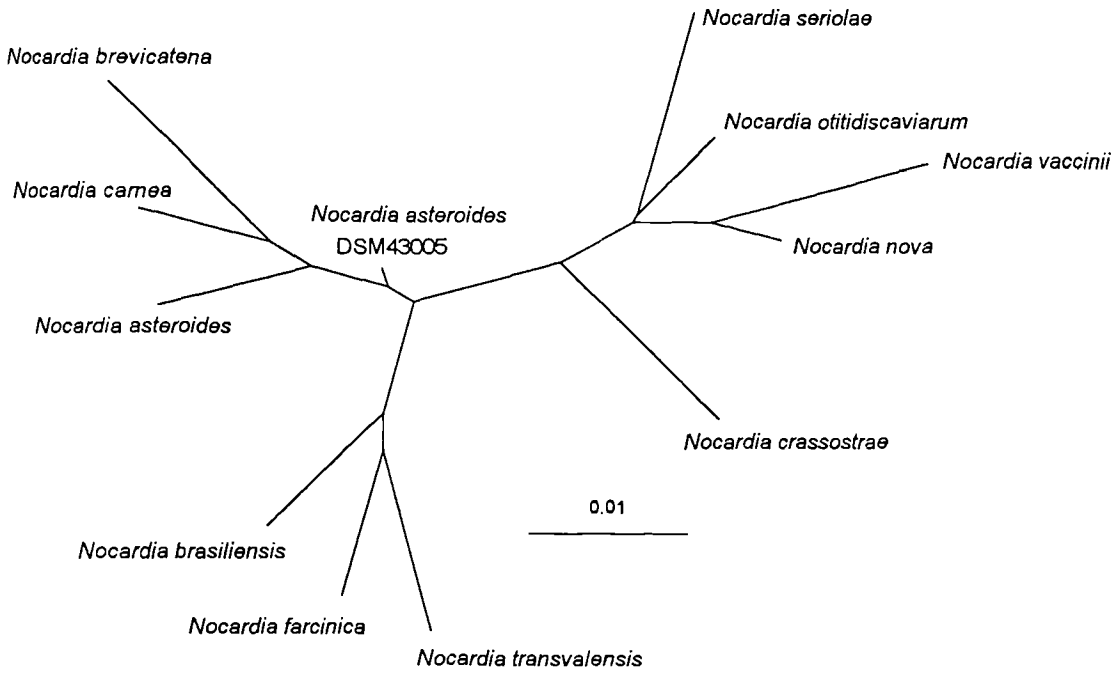


(a) Neighbor-joining tree

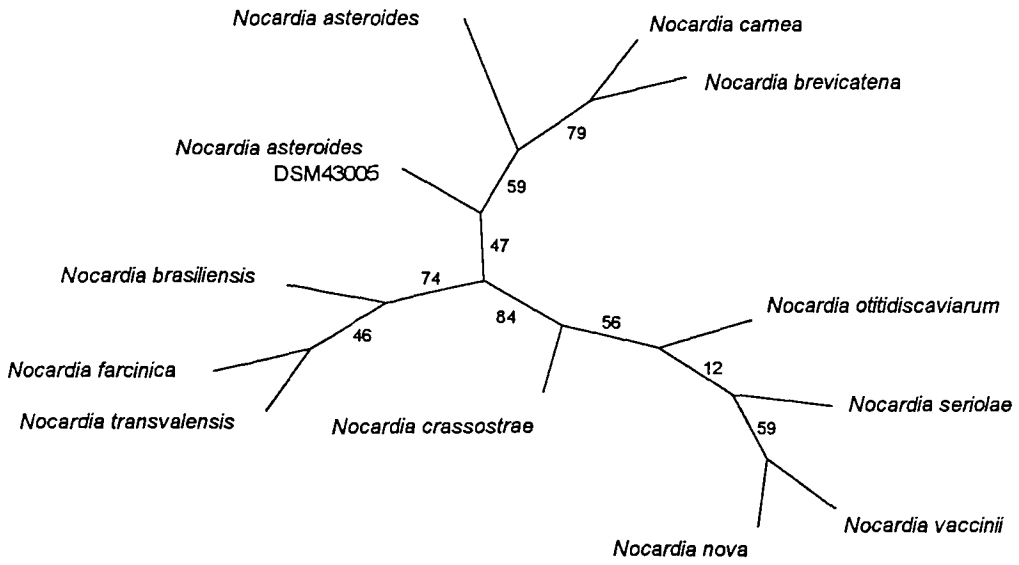


(b) Fitch-Margoliash tree

FIGURE 2-14. PHYLOGENETIC TREES DEPICTING RELATIONSHIPS FOUND BETWEEN REPRESENTATIVE NOCARDIAE. THE TREES WERE BASED ON THE DIFFERENT TREE-MAKING METHODS. THE TEST STRAINS WERE TYPE STRAINS UNLESS OTHERWISE SPECIFIED. BOOTSTRAP VALUES ARE INDICATED AT THE CORRESPONDING BRANCHES.

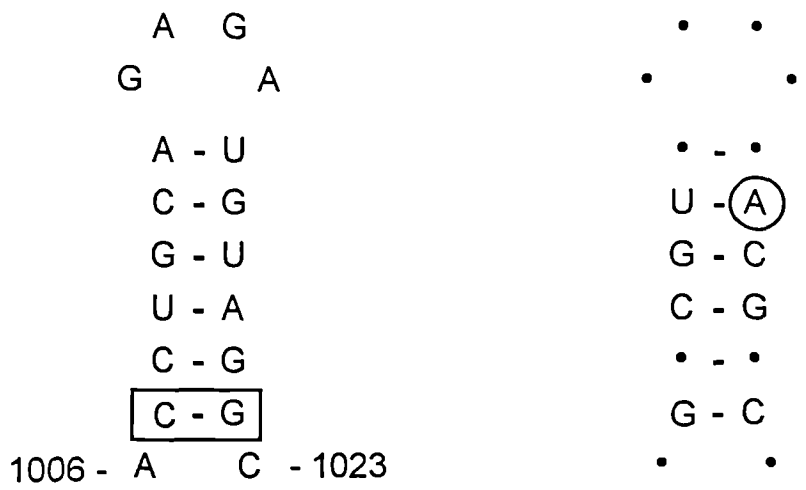


(c) Maximum likelihood tree



(d) Maximum parsimony tree

FIGURE 2-14. CONTINUED.



a. *Nocardia asteroides* rRNA subgroup b. *Nocardia otitidiscaviarum* rRNA subgroup

FIGURE 2-15. OLIGONUCLEOTIDE SIGNATURES THAT DISTINGUISH THE TWO *NOCARDIA rRNA SUBGROUPS*. *THE P37-1 HELIX* (APPENDIX D) LIES BETWEEN *ESCHERICHIA COLI* NUCLEOTIDE POSITIONS 1006 AND 1023 (APPENDIX C); THE PAIR OF NUCLEOTIDES ENCLOSED IN THE BOX IS G-C IN *NOCARDIA TRANSVALENSIS* N1202^T AND THE *A*-RESIDUE IN THE CIRCLE IS REPLACED BY *G* IN "*NOCARDIA CRASSOSTRAE*" N1170.

TABLE 2-21. 16S rDNA SEQUENCE SIGNATURES THAT DIFFERENTIATE BETWEEN THE REPRESENTATIVE *NOCARDIA* STRAINS.

Positions*	*N. asteroides*	*N. asteroides* DSM43005	*N. brasiliensis*	*N. brevicatena*	*N. carnea*	*"N. crassostrae"*	*N. farcinica*	*N. nova*	*N. otitidiscaviarum*	*N. seriolae*	*N. transvalensis*	*N. vaccinii*
137:226	U:A	U:A	U:A	U:A	U:A	U:A	C:G	U:A	U:A	U:A	U:A	U:A
199:218	A:U	A:U	A:U	A:U	A:U	A:U	A:U	A:U	G:C	A:U	A:U	A:U
258:268	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C	A:U	G:C	G:C
293:304	G:C	G:C	A:U	G:C	G:C	A:U	G:C	A:U	G:C	G:C	G:C	G:C
407:435	A:U	A:U	A:U	A:U	A:U	A:U	A:U	A:U	A:U	A:U	C:G	A:U
444:490	A:U	A:U	A:U	A:U	U:A	A:U	A:U	A:U	A:U	A:U	A:U	A:U
445:489	G:C	G:C	G:C	G:C	C:G	G:C	G:C	G:C	G:C	G:C	G:C	G:C
446:488	G:C	G:C	G:C	G:C	C:G	G:C	G:C	G:C	G:C	G:C	G:C	G:C
591:648	C:G	U:A	U:A	U:A	U:A	U:A	U:A	C:G	U:A	U:A	U:A	C:G
603:635	C:G	C:G	C:G	C:G	C:G	C:G	C:G	U:A	C:G	C:G	C:G	C:G
614:626	G:C	G:C	G:C	A:U	G:C	G:C	G:C	G:C	G:C	A:U	G:C	A:U
615:625	G:C	G:C	G:C	C:G	G:C	G:C	G:C	C:G	C:G	C:G	G:C	C:G
616:624	G:C	G:C	G:C	A:U	G:C	G:C	G:C	A:U	A:U	A:U	G:C	A:U
824:876	U:A	U:A	U:A	U:A	U:A	U:A	C:G	U:A	U:A	U:A	U:A	C:G
825:875	A:U	A:U	A:U	A:U	A:U	A:U	G:C	A:U	A:U	A:U	A:U	G:C
1002:1038	G:C	G:C	G:C	G:C	G:C	G:C	G:C	A:U	G:C	G:C	A:U	A:U
1007:1022	C:G	C:G	C:G	C:G	C:G	G:C	C:G	G:C	G:C	G:C	G:C	G:C
1009:1020	U:A	U:A	U:A	U:A	U:A	C:G	U:A	C:G	C:G	C:G	U:A	C:G
1011:1018	C:G	C:G	C:G	C:G	C:G	U:G	C:G	U:A	U:A	U:A	C:G	U:A
1120:1153	U:A	U:A	C:G	U:A	U:A	U:A	C:G	U:A	U:A	U:A	C:G	U:A
1122:1151	A:U	A:U	G:C	A:U	A:U	A:U	G:C	A:U	A:U	A:U	G:C	A:U
1134:1140	C:G	C:G	G:C	C:G	C:G	C:G	C:G	C:G	C:G	C:G	C:G	C:G

*, *Escherichia coli* numbering (Brosius *et al.*, 1978).

<BLANK>

D. Discussion

1. Suprageneric classification of mycolic acid-containing actinomycetes and related strains

It is possible to re-evaluate the taxonomic status of the families *Corynebacteriaceae*, *Mycobacteriaceae* and *Nocardiaceae* given marked improvements in the classification of the genera *Gordona*, *Nocardia*, *Rhodococcus* and *Tsukamurella* following the application of modern taxonomic techniques. It is clear from both the present and earlier studies that members of these families form a monophyletic clade (Embley and Stackebrandt, 1994; Ruimy *et al.*, 1994b; Rainey *et al.*, 1995a; Figure 2-4). In the present study, *Turicella otitidis* was found to be a member of the *Corynebacterium* clade.

Members of the genera *Corynebacterium, Dietzia, Gordona, Mycobacterium, Nocardia, Rhodococcus* and *Tsukamurella* can be divided into two groups on the basis of the discontinuous distribution of certain chemical markers. Corynebacteria and dietziae contain short chain mycolic acids (22-38 carbon atoms) and N-acetylated muramic acid in their peptidoglycans (Goodfellow, 1992; Rainey *et al.*, 1995c) whereas gordonae, mycobacteria, nocardiae, rhodococci and tsukamurellae are characterised by the presence of relatively long chain mycolic acids (34 to 90 carbons atoms) and N-glycolated muramic acid (Goodfellow, 1992). It is , evident from the phylogenetic analyses based on 16S rDNA sequence data that *Corynebacterium amycolatum* and *Turicella otitidis*, which lack mycolic acids, belong to the first group.

The taxonomic status of the two broad chemical groupings was underpinned by the 16S rDNA sequence data though it was not possible to unequivocally identify the position of root in the evolutionary tree. The two groups were recognised in the analyses based on four tree-making methods. Bootstrap values of 70 and 82 were

obtained in the case of the groupings derived from the neighbor-joining and maximum parsimony analyses.

A case can be made for the classification of *Corynebacterium amycolatum* and *Turicella otitidis* in the family *Corynebacteriaceae* based on 16S rRNA sequence and morphological data. It is possible that these organisms contain silent "mycolic acid" genes and hence have lost the ability to synthesise mycolic acids. Such a situation would be analogous to that of certain rhizobia and other nitrogen fixing bacteria which are unable to form effective nodules as they have lost their nitrogenase genes during the course of evolution (Hennecke *et al.*, 1985). It is important that the phylogenetic position of *Corynebacterium amycolatum* and *Turicella otitidis* is clarified by sequencing genes, such as those coding for ATPases and elongation factors, which are independent of the ribosomal RNA genes.

The status of the genus *Turicella* is a matter for debate as organisms in this taxon are closely related to members of the genus *Corynebacterium* on the basis of 16S rDNA sequence and morphological data (Funke *et al.*, 1994). However, unlike corynebacteria, *Turicella otitidis* strains contain unsaturated menaquinones (*i.e.*, MK-10 and MK-11) and lack mycolic acids. These chemical data lend support of the continued recognition of the genus *Turicella.*

The "mycolic acid" phyletic line encompassed representatives of the families *Mycobacteriaceae* and *Nocardiaceae.* It is apparent from the chemotaxonomic and 16S rRNA sequence data that the separation of the families *Mycobacteriaceae* and *Nocardiaceae* is rather artificial. A case can be made for reducing the family *Nocardiaceae* Castellani and Chalmers 1919[AL] to a synonym of the family *Mycobacteriaceae* Chester 1897[AL]. In deed, it is apparent from the present study that mycolic acid-containing actinomycetes can readily be classified into either the families *Corynebacteriaceae* Lehmann and Neumann 1907[AL] or *Mycobacteriaceae*

Chester 1897[AL]. The family *Corynebacteriaceae* provides a niche for the genera *Corynebacterium*, *Dietzia* and *Turicella*, and the family *Mycobacteriaceae* encompasses the genera *Gordona*, *Mycobacterium*, *Nocardia*, *Rhodococcus* and *Tsukamurella*.

2. Phylogeny of Nocardia

The results of the present study together with those from earlier studies show that nocardiae, apart from *Nocardia pinensis*, form a monophyletic group that can be defined by using a combination of phenotypic and chemotaxonomic properties (Goodfellow, 1992; Ruimy *et al.*, 1994b; Rainey *et al.*, 1995a). The detection of hexahydrogenated menaquinones with eight isoprene units in which the two end units were cyclised in the type strains of *Nocardia nova*, *Nocardia seriolae* and *Nocardia vaccinii* is in excellent agreement with corresponding data on representatives of the remaining validly described species of *Nocardia* (Howarth *et al.*, 1986; Collins *et al.*, 1987; Kämpfer *et al.*, 1990).

In general, the 16S rRNA sequences generated in the present study and corresponding nucleotide sequences reported by earlier investigators showed good congruence (Ruimy *et al.*, 1994b; Rainey *et al.*, 1995a). The differences found between corresponding nucleotide sequences ranged from 0 to 0.5%, such differences should not affect the products of phylogenetic analyses, especially the position of organisms in evolutionary trees. However, it is clear both from the present study and from that of Ruimy *et al.* (1994b) that the organism claimed by Rainey *et al.* (1995a) to be a *Nocardia otitidiscaviarum* strain was in fact a contaminant. Consequently, the report by Rainey *et al.* (1995a) that the type strains of *Nocardia farcinica* and *Nocardia otitidiscaviarum* had almost identical primary structures in their 16S rDNA (one nucleotide difference out of 1464 positions) and shared a high level of DNA relatedness (85%) is erroneous. Similar errors have

been uncovered for other organisms, notably *Mycobacterium chitae* (Stahl & Urbance, 1990; Pitulle *et al.*, 1991) and *Rhodoplanes palustris* (Orso *et al.*, 1994; Hiraishi & Ueda, 1994).

A complete 16S rDNA database on nocardiae was generated in the present study. The information in this database should prove to be useful for the molecular identification of unknown nocardiae and for the recognition of novel nocardiae isolated from natural environments. The designation of species-specific and genus-specific oligonucleotide probes and PCR primers can readily be achieved by using specialised software, such as the ***AL16S*** program (Chapter I).

3. Taxonomic status of Nocardia pinensis

The taxonomic status of *Nocardia pinensis* has been controversial from the begining. The organism was assigned to the genus *Nocardia* given a combination of chemical, morphological and physiological properties (Blackall *et al.*, 1989b). However, atypical nocardial features included the slow-growth rate of the organism, mycolic acids that were monounsaturated in the two position and a distinctive antimicrobial sensitivity pattern (see Tables 2-10 and 2-11). Representative strains were considered to be most closely related to *Nocardia amarae*, an organism subsequently transferred to the genus *Gordona* as *Gordona amarae* (Goodfellow *et al.*, 1994; Klatte *et al.*, 1994c; Ruimy *et al.*, 1994b).

In the present study, the type strain of *Nocardia pinensis* was found to be most closely related to the genus *Gordona* in light of the 16S rDNA sequence data. However, this relationship was based on relatively low bootstrap values in the analyses based on the maximum parsimony and neighbor-joining methods. *Nocardia pinensis* strains, like gordonae, have unsaturated mycolic acids and are sensitive to 5-fluorouracil (20 µg/ml), lysozyme (50 µg/ml) and mitomycin C (5 µg/ml) [Blackall *et al.*, 1989b]. However, unlike gordonae, *Nocardia pinensis* strains

produce aerial hyphae and contain hexahydrogenated menaquinones with eight isoprene units in which the end two units are cyclised (MK-8[H_4], ω-cyclic).

It can be concluded that Nocardia pinensis has a combination of chemical and morphological properties that clearly distinguishes it from gordonae. In addition, *Nocardia pinensis* can be separated from members of both the genera *Gordona* and *Nocardia* given the unique secondary structure found in helix 18 (Figure 2-13). It seems likely that further comparative taxonomic studies will show that *Nocardia pinensis* merits the generic status in the family *Nocardiaceae.*

4. Taxonomic status of Tsukamurella wratislaviensis

It is clear from the results of the present study that *Tsukamurella wratislaviensis* Goodfellow *et al.* 1991 should not be retained in the genus *Tsukamurella.* The assignment of this organism to the genus *Rhodococcus* is supported by the both chemical and 16S rDNA sequence data (Ridell *et al.*, 1985). It is evident from the present study that *Tsukamurella wratislaviensis* contains dihydrogenated menaquinone with eight isoprene units (MK-8[H_2]) not major amounts of unsaturated menaquinone with nine isoprene units (MK-9) as reported by Goodfellow *et al.* (1991).

The primary structures of the 16S rRNAs from the type strains of *Tsukamurella wratislaviensis* and *Rhodococcus opacus* Klatte *et al.* 1994 are almost identical (99.59% similarity; 6 nucleotide differences out of 1470 positions; see Appendix H). In addition, both organisms have mycolic acids with 48 to 54 carbon atoms (Ridell *et al.*, 1985; Klatte *et al.*, 1994b) but have yet to be the subject of comparative taxonomic studies. Representatives of *Tsukamurella wratislaviensis* and *Rhodococcus opacus* need to be the subject of comparative phenotypic and

DNA-DNA pairing studies in order to determine whether or not each of these taxa merit species status.

CHAPTER III.
ISOLATION AND CHARACTERISATION OF ACTINOMYCETES ASSOCIATED WITH ACTIVATED SLUDGE FOAMING

A. Introduction

1. Activated sludge treatment

The activated sludge process is the most widely used secondary waste water treatment process in the world (Jenkins *et al.*, 1993). The efficiency of the process relies on the performance of two major treatment units, the aeration basin (biological reactor) and the solids separation device (secondary clarifier) which is usually a gravity sedimentation basin. The aeration basin is designed to promote the removal and biotransformation of both soluble and particulate pollutants by a mixed and variable consortium of micro- and macro-organisms, the **'activated sludge'**. Similarly, the secondary clarifier is designed to allow activated sludge solids to separate by flocculation and gravity sedimentation from treated wastewater. The outcome of these processes is a clarified (low suspended solids, low turbidity) overflow, the secondary effluent, and a thickened underflow which forms the returned activated sludge.

It is important that both the aeration basin and the secondary clarifier work satisfactorily. To this end factors that affect biological oxidation and the separation of solids are important in determining the overall efficiency of the process. A typical scheme for the treatment of activated sludge is shown in Figure 3-1. The most common problems associated with the activated sludge separation process are summarised in Table 3-1.

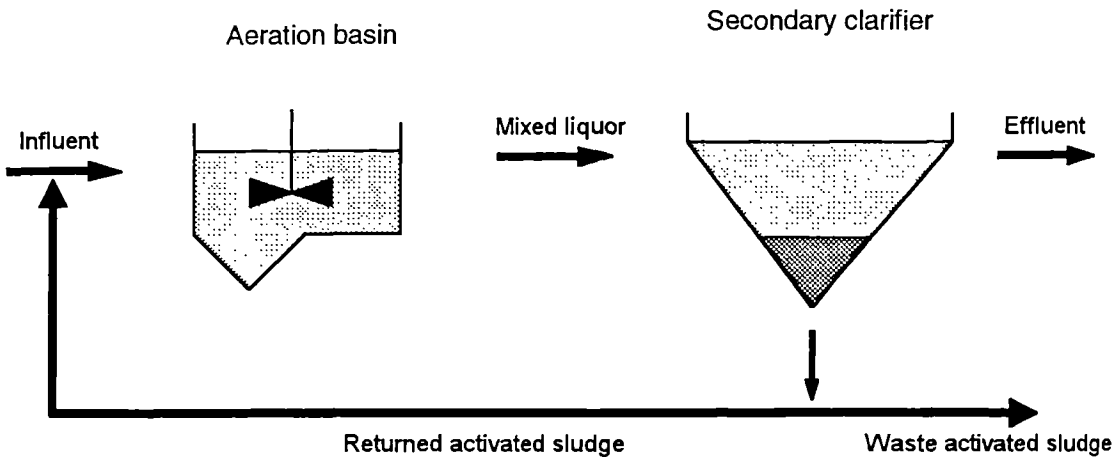


FIGURE 3-1. SCHEMATIC REPRESENTATION OF A TYPICAL ACTIVATED SLUDGE TREATMENT PLANT. MODIFIED FROM JENKINS *et al.* (1993).

TABLE 3-1. CAUSES AND EFFECTS OF ACTIVATED SLUDGE SEPARATION PROBLEMS *.

Abnormality	Cause	Effect
Dispersed growth	Microorganisms do not form flocs but are dispersed forming only small clumps or single cells.	Turbid effluent. No zone settling of the activated sludge.
Slime viscous bulking (non-filamentous bulking)	Microorganisms are present in large amounts of extracellular slime.	Reduced settling and compaction rates.
Pin or pinpoint floc	Small, compact, weak and roughly spherical flocs are formed.	The smaller flocs settle slowly. Low sludge volume index and a turbid effluent.
Filamentous bulking	Filamentous organisms extend from flocs to the bulk solution and interfere with compaction, settling, thickening and concentration of activated sludge.	Very clear supernatant. In severe cases the sludge blanket overflows the secondary clarifier.
Blanket rising	Denitrification in secondary clarifier leads to the release of poorly soluble nitrogen gas which attaches to activated sludge flocs propelling them to the secondary clarifier surface.	A scum of activated sludge forms on the surface of the secondary clarifier.
Foaming/scum formation	Caused by *(i)* non-degradable surfactants and by *(ii)* the presence of filamentous organisms such as "*Microthrix parvicella*" and mycolic acid-containing actinomycetes.	Foams float large amounts of activated sludge solids to the surface of treatment units. Foams accumulate and can putrefy.

*, Modified from Jenkins *et al.* (1993).

Most activated sludge solids separation problems can be attributed to the nature of the activated sludge flocs. The latter are made up of two basic elements, a biological component which consists of a variety of bacteria, fungi, protozoa and some metazoa, and a non-biological component composed of inorganic and organic matter particles. The core of the floc appears to be a number of heterotrophic bacteria that belong to genera such as *Achromobacter*, *Alcaligenes*, *Arthrobacter*, *Citromonas*, *Flavobacterium*, *Pseudomonas* and *Zoogloea* (Dias & Bhat, 1964; Pike, 1972; Tabor, 1976; Jenkins *et al.*, 1993). It has been suggested that microbial extracellular polymers, typically polysaccharides, play an important role in the bioflocculation of activated sludge (Tago & Aiba, 1977; Jenkins *et al.*, 1993).

It is evident from visual observation and physical measurements that there are two levels of structure in activated sludge flocs which are known as the **'micro-structure'** and the **'macro-structure'** (Sezgin *et al.*, 1978). The micro-structure is a product of microbial adhesion, aggregation, bioflocculation and the action of extracellular polymers. This process is the basis for floc formation because without the ability of microorganisms to stick to one another large microbial aggregates would not be formed. Flocs that show this micro-structure are usually small (up to about 75 μm in dimension), spherical and compact and are responsible for turbid supernatants in activated sludge. The larger compact flocs settle rapidly though smaller aggregates sheared off from them settle slowly thereby creating turbid supernatants.

The macro-structure of activated sludge flocs can be attributed to filamentous microorganisms that form a network(backbone)-like structure within the floc onto which floc-forming bacteria cling (Sezgin *et al.*, 1978). When activated sludge contains filamentous organisms, large floc sizes are possible as the backbone of filamentous organisms provides the floc with the strength needed to

hold it together in the turbulent environment of the aeration basin. The filamentous network influences the shape of the floc as the latter grows in the same direction as the filamentous organisms. Most of the activated sludge solids separation problems outlined in Table 3-1 can be interpreted in terms of failures in either the micro-structure or macro-structure of activated sludge flocs.

2. Bulking and Foaming

Bulking. This is a macrostructure failure due to the presence of large numbers of filamentous microorganisms in the activated sludge. These organisms interfere with the compaction and settling of activated sludge either by producing a very diffuse floc structure or by growing in profusion beyond the confines of the activated sludge floc into the bulk medium thereby causing bridging between flocs. The most common casual agents of activated sludge bulking are *"Microthrix parvicella"*, *"Nostocoida limicola"*, *Thiothrix* spp., *"Haliscomenobacter hydrossis"* and unidentified organisms (*e.g.*, types 0041, 0092, 021N, 0675, 0803, 0961, 1701 and 1851; Jenkins *et al.*, 1993; Hudson *et al.*, 1994).

Foaming. The formation of scum or foam in the activated sludge process is also caused by filamentous bacteria, notably *"Microthrix parvicella"* and mycolic acid-containing actinomycetes (Soddell & Seviour, 1990; Blackall *et al.*, 1994; Eikelboom, 1994). These bacteria have hydrophobic cell surfaces and when present in sufficient numbers in activated sludge they render flocs hydrophobic and hence amenable to the attachment of air-bubbles. The air bubble-floc aggregate is less dense than water and hence floats to the surface of the sludge. The hydrophobic flocs tend to stay at the surface of the sludge where they accumulate to form a thick, chocolate-brown coloured foam or scum. Activated sludge foams cause a number of problems, namely extra-house keeping on the part of the operator; blockage of scum removal systems; reduction of oxygen transfer at the surface of mechanically

aerated basins; carriage and dispersal of pathogens in wind blown scum; drying of scum with resultant cleaning and possibly odour problems, and reduction of effluent quality through an increase in effluent suspended solids and Biological Oxygen Demand if the scum reaches the final effluent (Soddell & Seviour, 1990).

3. Actinomycetes associated with activated sludge foaming

The first serious attempt to isolate and identify actinomycetes associated with activated sludge foaming was made by Lechevalier *et al.* (1976). A variety of actinomycetes were isolated from twenty-one sewage-treatment plants, located in nine states of the USA, using Czapek's agar supplemented with yeast extract (YCZ; Higgins & Lechevalier, 1969) and glycerol agar (Gordon & Smith, 1953). The isolated strains were characterised on the basis of phenotypic properties. The predominant species were *Gordona* (*Nocardia*) *amarae*, which was isolated from fifteen plants, and "*Nocardia rhodochrous*" which was recovered from seven plants. In addition, a number of isolates were identified as *Nocardia asteroides* and *Nocardia otitidiscaviarum* ("*N. caviae*") and others were assigned to the genera *Actinomadura, Micromonospora* and *Streptomyces*.

Lechevalier and his colleagues concluded that foaming caused by actinomycetes was evident in widely different areas in the USA though thick foams were not invariably associated with the presence of these organisms. They also noted that actinomycete foams were associated with warm weather, high aeration rates and thick mixed liquors. Filamentous "nocardioform" actinomycetes have also been implicated in foaming in anaerobic digesters (Van Niekerk *et al.*, 1987) but this is unlikely as such organisms are obligate aerobes.

Lemmer and Kroppenstedt (1984) studied the occurrence of actinomycetes in scums of fourteen sewage-treatment plants in Berlin, southern Germany and Switzerland. The causal organisms were isolated either on Columbia agar base

(Oxoid) or mineral agar (Ratzke, 1965) supplemented with paraffin and thiamin, and representative strains identified by using a combination of chemical and phenotypic properties. Eleven rhodococci, two tsukamurellas and a single strain of *Gordona* (*Nocardia*) *amarae* were isolated but only one species was associated with any one plant. In contrast, Sezgin *et al.* (1988) found that the most common isolate in activated sludge scum samples collected from two sewage treatment plants in the USA was *Gordona* (*Nocardia*) *amarae* (51 strains) followed by *Nocardia asteroides* (24 strains). Other isolates were assigned to the genera *Amycolatopsis*, *Dietzia* (*Rhodococcus maris*), *Micromonospora*, *Mycobacterium*, *Nocardia*, *Oerskovia* and *Rhodococcus*.

A microorganism with a pine-tree like morphology was first reported in activated sludge from an Australian sewage treatment plant (Blackall *et al.*, 1989b). The organism, which was isolated directly from foam by using a micromanipulation technique (Skerman, 1968), was assigned to the genus *Nocardia* as *Nocardia pinensis* mainly on the basis of chemical and morphological properties (Blackall *et al.*, 1989a). *Nocardia pinensis* strains were subsequently isolated from activated-sludge sewage-treatment plants in New South Wales, Queensland and Victoria in Australia (Seviour *et al.*, 1990; Soddell & Seviour, 1994). The isolation of this relatively slow-growing actinomycete can be attributed to the successful use of the micromanipulation technique first applied to an examination of activated sludge samples by Blackall *et al.* (1985).

The limitations of the conventional dilution plating technique for the isolation of actinomycetes from activated sludge samples were discussed by Soddell and Seviour (1994). They argued that only *Nocardia pinensis* could be microscopically distinguished from other "nocardioform" actinomycetes, including *Gordona* (*Nocardia*) *amarae,* and that the dilution plating technique was inadequate for the

isolation of slow-growing organisms such as *Nocardia pinensis*. They also isolated a number of mycobacteria from Australian activated sludge sewage-treatment plants.

It is interesting that *Gordona* (*Nocardia*) *amarae* and *Nocardia pinensis* strains have only been isolated from activated sludge. It also appears from the literature that other actinomycetes containing mycolic acids, notably members of the genera *Gordona*, *Nocardia, Rhodococcus* and *Tsukamurella*, may be responsible for the formation of foams and scums in activated sludge. It is also interesting that *Sphaerobacter thermophilus*, a thermophilic actinomycete, has been isolated from high-temperature treated sewage sludge (Demharter *et al.*, 1989). This organism belongs to the actinomycete subline of descent (see Figure 2-4).

The primary aim of the present investigation was to isolate, characterise and identify actinomycetes associated with extensive foaming at the Stoke Bardolph sewage treatment plant near Nottingham, England, UK.

B. Materials and Methods

1. Test strains

1.1. Sample collection

Foam and scum samples were collected from six different sites, including the aeration basins and secondary clarifiers, at Stoke Bardolph Sewage Works (Figure 3-2). The samples were stored at 4 °C.

1.2. Isolation and maintenance of strains

Each sample was serially diluted down to 10^{-5} using sterile quarter-strength Ringer's solution. Five aliquots (*ca.* 0.1 ml) of each dilution were plated onto glucose yeast extract agar (GYEA; Gordon & Mihm, 1962; Appendix A) supplemented with cyclohexamide (50 µg/ml). The inoculated plates were incubated at 30 °C for 5 days. All of the incubated plates were found to support the growth of orange-reddish actinomycete-like colonies which were reminiscent of *Tsukamurella* strains. One colony derived from each of the sampling sites was transferred onto fresh GYEA plates using sterile tooth-picks. All six isolates were subcultured several times and checked for purity by acid-fast and Gram-staining (Hucker & Conn, 1923). The isolates were designated as N1171 to N1176. The six isolates were maintained on GYEA slopes at 4 °C and as glycerol suspensions (20 %, v/v) at -20 °C (Wellington & Williams, 1978).

1.3. Test strains

The six isolates were studied together with six reference strains of *Tsukamurella paurometabola* (Table 3-2).

(a)

(b)

FIGURE 3-2. STOKE BARDOLPH SEWAGE WORKS: (a) AERATION BASIN SHOWING EXTENSIVE FOAMING AND (b) SECONDARY CLARIFIER.

TABLE 3-2. DESIGNATION AND SOURCE OF THE REFERENCE STRAINS.

Lab. No.	Designation	Source
JC7^T	*Tsukamurella paurometabola*	DSM 20162 (*Corynebacterium paurometabolum*)
M333	*Tsukamurella paurometabola*	R. E. Gordon, IMRU 1283 (*Mycobacterium* sp.); G. Altmann, 4479; human eye
M334	*Tsukamurella paurometabola*	R. E. Gordon, IMRU 1520 (*Mycobacterium album*); C. McDurmont, R456
M337	*Tsukamurella paurometabola*	R. E. Gordon, IMRU 1505 (*Mycobacterium album*); C. McDurmont, R161
M343	*Tsukamurella paurometabola*	R. E. Gordon, IMRU 1312 (*Mycobacterium rhodochrous*); L. K. Georg, 44329-65; iceballs, Hong Kong
N663	*Tsukamurella paurometabola*	NCTC 10741 (*Rhodococcus aurantiacus*); M. Tsukamura, 3462; A. Kruse; sputum

^{T}, Type strain. Abbreviations: DSM, Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany; NCTC, National Collection for Type Cultures, Central Public Health Laboratories, Colindale, London, UK.

2. Chemotaxonomy

2.1. Preparation of biomass

Dried biomass was prepared from the test strains as described in Chapter II.

2.2. Analysis of fatty acids, including mycolic acids

Preparation of methyl esters. Fatty acid and mycolic acid methyl esters were prepared using the whole-organism acid methanolysis technique described by Minnikin *et al.* (1980) and modified by Luquin *et al.* (1989). Small amounts of dried biomass (*ca.* 50 mg) in 8.5 ml test tubes fitted with a Teflon lined screw caps (Aldrich Ltd., The Old Brickyard, New Road, Gillingham, Dorset, England, UK) were treated with a mixture (3 ml) of dry-methanol-toluene-sulphuric acid (30:15:1, v/v/v), and the closed test tubes kept at 80 °C in a hot air oven overnight. After cooling to room temperature, 2 ml of *n*-hexane was added to each preparation and the mixtures shaken and centrifuged for 10 minutes at low speed. The hexane extracts were transferred to clean test tubes and mixed with an equal volume of 0.3M phosphate buffer (42.57g of Na_2HPO_4 and 12.0g of NaOH per litre of distilled water; pH 11 to 12). In each case the upper organic layers, which contained the fatty acid and mycolic acid methyl esters, were transferred to clean test tubes and evaporated to dryness on a heating block at 50 °C under a stream of nitrogen.

Analytical and preparative thin-layer-chromatography. Analytical one-dimensional thin-layer-chromatography (TLC) of mixtures of fatty acid methyl esters (FAMEs) and mycolic acid methyl esters (MAMEs) was performed using pieces of aluminum sheet (10 cm × 10 cm; Merck 5554 silica gel 60 F254). Each dried lipid sample was dissolved in 50 µl of diethyl ether and an aliquot of *ca.* 5 µl applied to an aluminum sheet. The preparation was developed in petroleum ether (b.p. 60-80 °C)/acetone (95/5, v/v) and the positions of the separated components

revealed by spraying with ethanolic molybdophosphoric acid (5 %, w/v) followed by heating at 180 °C for 15 minutes.

Fatty acid and mycolic acid methyl esters were purified on plastic-backed TLC sheets (10 cm×10 cm; Merck 5735) and developed using the solvent system mentioned above. The positions of the FAMEs and MAMEs were visualized by spraying with ethanolic rhodamine (0.1 %, w/v) and examining under UV light. The bands containing the FAMEs and MAMEs were scraped from the TLC plates and each band transferred to a 1.5 ml tube. Diethyl ether (*ca.* 1 ml) was added to these preparations and the contents of the tubes centrifuged at 10,000 rpm for 5 minutes. The supernatants were transferred to small glass vials, dried under a stream of nitrogen and stored at -20 °C.

Gas-liquid chromatographic analysis of fatty acid methyl esters. Purified FAMEs were separated and quantified using a gas chromatograph (Shimadzu Mini-3 Gas Chromatograph, Kyoto, Japan) equipped with a SPB™-1 fused silica capillary column (30 m × 0.25 mm ID; Supelco Ltd., Shire Hill, Saffron Walden, Essex, England, UK). The temperature for both the injector and flame-induced-detector (FID) was kept at 270 °C. The column was programmed to operate from 150 to 250 °C with increases of 4 °C per minute; nitrogen was used as the carrier gas. Peak areas and retention times were recorded using a Trio integrator (Trio Ltd., Trivector, Sandy, Bedfordshire, England, UK) and eluted components identified by using a FAME standard mixture (Supelco CP™ Mix; Supelco Ltd.). Peak identities were confirmed using a Hewlett Packard Gas Chromatograph-Mass Spectrometer (HP5890 and HP5972; Hewlett Packard Ltd., Nine Mile Ride, Wokingham, Berkshire, England, UK).

Mass spectrometry of mycolic acid methyl esters. Purified MAMEs were subject to electron-impact mass spectrometry using a mass spectrometer (Kratos MS 80RF) with an ionisation energy of 70 eV.

Pyrolysis gas chromatography of mycolic acid methyl esters. The pyrolysis products of the MAMEs were analysed using a Hewlett Packard gas chromatograph (HP5890) coupled to a mass selective detector (HP5972). The purified MAMEs were redissolved in 200 µl of hexane and injected into the gas chromatograph using an automated sampler (HP7673). Fatty acid methyl esters derived from pyrolysis were separated using a capillary column (HP-5MS cross-linked 5% Ph Me silicone, 30m x 0.25 mm ID) that was programmed from 150 to 250 °C at 5 °C/minute and maintained for 10 minutes thereafter. The injector was kept at 310 °C and helium used as the carrier gas. Individual peaks were identified according to their mass spectra and retention times. The size and degree of saturation of the side chain in the MAMEs were determined by calculating molecular weights (M^+) from the mass spectra using the following equation:

$$m/z = 60 + (number\ of\ carbons \times 14) - (number\ of\ double\ bonds \times 2)$$

In the case of monounsaturated fatty acid side chains, the molecular ion minus methanol (M^+-MeOH; *m/z*-32) shows a stronger intensity than the molecular ion (M^+).

Precipitation test for mycolic acid methyl esters. The procedure used was that described by Hamid *et al.* (1993). Dried crude MAMEs extracted from the test strains were dissolved in dichloromethane (0.5 ml) and the preparations transferred to 1.5 ml polypropylene microcentifuge tubes (Treff, Dagersheim, Switzerland) and the contents evaporated to dryness under nitrogen at 37 °C. The residues were mixed with acetonitrile (0.1 ml) and toluene (0.2 ml) until they were completely

dissolved. Further 0.2 ml aliquots of acetonitrile were added to these preparations; turbidity indicates the insolubility of MAMEs.

2.3. Analysis of menaquinones

Menaquinone profiles of the test strains were examined by using the method described in Chapter II.

2.4. Analysis for diaminopimelic acid isomers

Whole-organism hydrolysis. Dried biomass (*ca.* 30-40 mg) was hydrolysed with 1 ml of 6N HCl in Bijoux bottles at 100 °C for 20 hours. After cooling, the hydrolysates were filtered (Whatman No. 1 Filter Paper, Whatman Ltd., Maidstone, England, UK) and washed twice with sterile distilled water (1 ml). The combined filtrates were concentrated to dryness using a vacuum pump (Genevac CVP 100/2; Edward's High Vacuum, Sussex, England, UK). The dried extracts were dissolved in sterile water (1 ml), dried again and the process repeated until the smell of hydrochloric acid was lost. The residues were then redissolved in distilled water (0.3 ml) and transferred to Bijoux bottles.

Thin layer chromatography. An aliquot (5 µl) of the resultant sample was applied to a cellulose TLC plate (20 cm×20 cm; Merck 5716). An aqueous mixture (5 µl) of the isomers of α,ε-DAP (Sigma) was used as a standard. The plates were developed in methanol/water/10N HCl/pyridine (80:26.25:3.75:10, v/v/v/v) until the solvent front had reached the top of the plate. The plates were dried in a fume cupboard and spots visualised by spraying with 0.2% (w/v) ninhydrin in acetone followed by heating at 100 °C for five minutes. The DAP isomers, which appeared as blue-violet coloured spots after 2 to 3 minutes, were identified by comparison to the standard. The spots corresponding to the DAP isomers became yellow after 24 hours.

2.5. Analysis of whole-organism sugars

Whole-organism hydrolysis and derivatisation. Freeze-dried biomass (*ca.* 50mg) placed in screw-capped tubes was treated with 1ml of 0.5M hydrochloric acid. The tubes were sealed and autoclaved at 121°C for 20 minutes. After cooling, the resultant acid hydrolysates were filtered (45μm; Millipore) and 0.2 ml of a 12M ammonia solution added to each of the filtrates. After the addition of 0.1ml of a freshly prepared solution of sodium borohydride (100mg/ml in 3M ammonia), each preparation was shaken for 5 minutes and incubated at 37°C for an hour. Glacial acetic acid was then added dropwise until effervescence was no longer visible. The preparations were then cooled to 4°C, kept on ice, and 0.3ml of 1-methylimidazole (Sigma) and 2 ml of acetic anhydride slowly added. The resultant preparations were shaken and allowed to stand for a further 15 minutes when 5 ml of distilled water was added. After cooling to room temperature, 1ml of dichloromethane was added to each preparation and the tubes shaken vigorously for 2 to 3 minutes. The upper layers were discarded and the lower layers evaporated to dryness under nitrogen gas and stored at -80°C until required. The residues were redissolved in dichloromethane (50μl).

Gas chromatography. The alditol acetate preparations (1μl) were analysed using a model Shimadzu gas chromatograph fitted with a flame ionisation detector. Separation was achieved using a 0.25 mm x 30-m SP 2380 (Sulpelco) fused silica capillary column with a film thickness of 0.25 m. The initial column temperature, 230 °C, was maintained for 5 minutes then it was increased to 270°C at 10°C/minute. The final temperature was maintained for 20 minutes. The injector and detector temperatures were set at 260 and 300°C, respectively. The total elution time for each sample was seventeen minutes.

3. Pyrolysis mass spectrometry

3.1. Test strains

The ten strains included in the chemical investigations were studied.

3.2. Growth conditions

The test strains were inoculated onto GYEA (Gordon & Mihm, 1962) plates from cultures grown on the same medium and stored at 4 °C. The inoculated plates were incubated at 30 °C for 3 days. All of the media were prepared from the same batch of ingredients.

3.3. Preparation of samples

Ferro-nickel alloy foils and pyrolysis tubes (Horizon Instruments Ltd., Heathfield, West Sussex, England, UK) were washed in acetone and dried at room-temperature. The clean foils were inserted, using clean forceps, into the pyrolysis tubes so as to extend about 6 mm out of the ends of the tubes. In order to reduce the effect of colonial variation, the inocula applied to the foils were prepared from a mixture of several colonies. Inoculum from each plate was used to seed three foils. In addition, four strains were examined in duplicate to determine experimental reproducibility. The assembled tubes plus foils were dried in an oven at 80 °C for 10 minutes.

3.4. Curie-point pyrolysis mass spectrometry

Curie-point pyrolysis mass spectrometry was carried out using a Horizon RAPyD-400 mass spectrometer (Horizon Instruments Ltd.). The inlet heater was set at 160 °C and the heated tube loader at 120 °C. Curie-point pyrolysis was carried out at 530 °C for 2.4 seconds under vacuum with a temperature rise time of 0.6 of a second. The pyrolysates were ionised by collision with a cross-beam of low energy electrons (20 eV) and the ions separated in a quadrupole mass spectrometer at

scanning intervals of 0.35 of a second. Integrated ion counts, recorded for each sample at unit mass intervals from 51 to 200, were stored on a hard disk.

3.5. Multivariate statistical analysis

The multivariate statistical analyses were achieved using the ***PYMENU*** (Horizon Instruments) and ***GENSTAT*** (Nelder, 1979) programs. In brief, mass spectra were normalised to compensate for variations in sample size and the 100 masses showing the highest characteristicity values examined first by principal component analysis then by canonical variate analysis (PC-CVA). A Mahalanobis distance matrix (Mahalanobis, 1936) generated from the analysis was transformed to a similarity matrix using Gower's coefficient (S_G; Gower, 1971) and a dendrogram generated from these data using the unweighted pair group method with arithmetic averages algorithm (Sneath & Sokal, 1973). Details of the multivariate statistical analyses are given in Chapter IV.

4. Whole-organism protein electrophoresis

4.1. Cultivation of test strains and sample preparation

The ten test strains were inoculated onto GYEA plates (Gordon & Mihm, 1962) and incubated at 30 °C for 3 days. Biomass scraped from the plates was suspended in 600 µl of cold GMT buffer (0.125 M Tris, 20% [v/v] glycerol, 10% [v/v] β-mercaptoethanol, pH 6.8) prior to sonication for one minute in two 30 seconds pulses with a 20 second interval between the pulses. The tubes containing the samples were held in crushed ice throughout the sonication treatment. The sonicated cell suspensions were mixed with 400 µl of a sodium dodecyl sulphate (SDS) solution (10%, w/v) and boiled for 5 minutes at 100 °C in a water bath. The tubes were quickly cooled in ice for 10 minutes then centrifuged for 15 minutes at

13000 rpm. The clear supernatants, which contained soluble denatured whole-organism proteins, were transferred to clean tubes and stored at -20 °C.

4.2. Denaturing polyacrylamide gel electrophoresis

Sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was carried out following the manufacturer's instructions using a Hoeffer SE 600 Vertical Slab Gel Unit (Hoeffer Scientific Instruments, Newcastle-under-Lyme, England, UK). The electrophoretic system consisted of a discontinuous 10% separation polyacrylamide gel and a 4% stacking gel.

Gels were cast and run in a Hoeffer SE 600 Vertical Slab Gel Unit using 1.5 mm spacers and sample combs with 10 wells. The gel casting unit was assembled according to the manufacture's instructions using glass plates thoroughly cleaned with ethanol. The separating gel (polyacrylamide 10% T 2.7% C) was prepared and degassed in a 125 ml side arm vacuum flask according to a recipe (Table 3-3) that omitted the SDS, TEMED (N,N,N',N',tetramethylethylenediamine) and ammonium persulphate; these reagents were added to the degassed flask contents and mixed by gentle swirling. The separating gel solution was aspirated with a 20 ml syringe and gently dispensed into the glass plate assembly unit until the gel was approximately 2.0 cm from the top. A gel former was inserted into the gel assembly unit to prepare the necessary space for the stacking gel. Gels were allowed to polymerise at room temperature for an hour. The gel formers were then carefully removed and the surfaces of the gels rinsed once with approximately 2 ml of the stacking gel solution by rocking the casting stand. The rinse solution was then discarded and the stacking gel poured over the solidified separating gel.

The stacking gel (polyacrylamide 4% T, 2.7% C) was prepared and degassed in a 50 ml side arm flask according to a recipe (Table 3-3) that omitted SDS, TEMED and ammonium persulphate; as before the latter were added to the

TABLE 3-3. COMPOSITION OF SDS-PAGE GELS.

Stock solutions / reagents	Separating gel 10% T, 2.7% C	Stacking gel 4% T, 2.7% C
Monomer solution (30% T, 2.7% C)[a]	20.0 ml	2.66 ml
4 x Running gel buffer[b]	15.0 ml	—
4 x Stacking gel buffer[c]	—	5.0 ml
10% SDS (w/v)	0.6 ml	0.2 ml
Mili-Q water	24.1 ml	12.2 ml
Ammonium persulphate (10%, w/v)	300 µl	100 µl
TEMED[d]	20 µl	10 µl

[a], The monomer solution was prepared by mixing 58.4 g of acrylamide and 1.6 g of bis-acrylamide in a final volume of 200 ml of autoclaved Mili-Q water. The solution was filtered and stored at 4°C in the dark.

[b], 4 x running gel buffer consisted of 1.5 M Tris, pH 8.8 (autoclaved).

[c], 4 x stacking gel buffer consisted of 0.5 M Tris, pH 6.8 (autoclaved)

[d] ,TEMED, N,N,N',N',tetramethylethylenediamine.

degassed solution and mixed by gentle swirling. The resultant mixture was aspirated in a 20 ml syringe and used to rinse the surface of the gels as described earlier. The remaining stacking gel solution was poured into the gel casting unit. A 10-well Teflon comb was introduced into the stacking gel which was allowed to polymerise for an hour at room temperature. The stacking gel comb was then removed and the wells rinsed with distilled water.

The gel casting slabs were attached to the upper buffer tank and the application of samples performed with the gels still in the casting stand. The wells were filled with tank buffer (0.025 M Tris, 0.0192 M glycine, 0.1% [w/v] SDS, pH 8.3) and the thawed protein samples applied using sterilised 20 µl gel loading tips (Sigma Ltd.). The upper buffer tank assembly unit (with gel slabs) was inserted into the lower buffer tank which had been previously filled with cold (10°C) tank buffer. The top reservoir was carefully filled with cold tank buffer and air bubbles from underneath the gel assembly unit were removed using a thin glass rod. The buffer was circulated by using a magnetic stirrer with the cooling unit remaining connected to a cold water supply during the entire run.

Electrophoresis was carried out at a constant current of 30 mA per gel until the indicator dye front had migrated 10 cm (approximately 4 hours). The gels were removed from the gel assembly and stained using Coomassie Blue stain, as described on the manufacturer's protocol (Hoefer Scientific Instruments, 1992). Stained gels were dried on a gel drier at 80 °C under vacuum for approximately an hour. The gels were allowed to cool before removal from the gel drier then stored flat away from strong light.

4.3. Data analysis

The gels were scanned using a flat bed scanner (HP Scanjet Plus; Hewlett Packard) at 300 dots-per-inch (dpi) using the 256 gray-scale mode. Images were

stored on floppy-disks in the non-compressed Tag Image File Format version 5 (TIFF). Computer-assisted analysis was carried out using ***GelCompar*** version 3.1 (Applied Maths BVBA, Risquons-Toutstraat 38, B-8511, Kortrijk, Belgium) as described elsewhere (Vauterin & Vauterin, 1992; Pot *et al.*, 1994). The ***GelCompar*** package was run on an IMB-PC compatible computer (Elonex 486-66).

Digitised gel images were normalised by choosing reference peaks from different gels. The background of the resultant normalised gel images were subtracted by applying the rolling disk algorithm as recommended the by manufacturer. The pairwise calculations of similarities between samples were carried out by using the Pearson product-moment correlation coefficient (*r*, Sneath & Sokal, 1973). To optimise matching between pairs of traces, the program performed a secondary alignment of traces before the computation of similarities (Vauterin *et al.*, 1991). This alignment involved a lateral displacement of one trace up to five points on either side of the initial alignment. The highest *r*-value was retained for the cluster analysis. Clustering was achieved by applying the unweighted pair group method with arithmetic averages algorithm (Sneath & Sokal, 1973).

5. Numerical phenetic classification

5.1. Data acquisition

The six sludge isolates and the six *Tsukamurella paurometabola* strains (Table 3-2) were examined for 96 unit characters. The organisms were inoculated onto GYEA plates (Gordon & Mihm, 1962) from GYEA slopes stored at 4 °C, and incubated for 3 days at 30 °C. Biomass scraped from the GYEA plates was suspended in a Bijoux bottle containing sterile quarter-strength Ringer's solution and approximately 7 to 10 glass beads (Jencon Scientific Ltd., Leighton Buzzard, Bedfordshire, England, UK); the preparations were then shaken to obtain homogeneous inocula.

When possible, the tests were carried out following a procedure that involved the use of an automatic multipoint inoculator (Denley-Tech, Denley Instruments Ltd., Daux Road, Billingshurst, Sussex, England, UK). This apparatus allows the standardised, multiple surface inoculation of 90 mm diameter Petri dishes (Sterilin Ltd., Teddington, Middlesex, England, UK) with up to 20 different strains. In this study plates were inoculated with either five or six strains. The inoculated plates were incubated at 30 °C unless otherwise stated and the results recorded after 7, 14, 21 and 28 days. The final test readings were used to code data for computation. Four strains, namely isolate N1171 and *Tsukamurella paurometabola* JC7^T, M334 and N663, were studied in duplicate to assess test reproducibility.

5.2. Morphology and staining

Cultural and colonial properties of the test strains were determined from GYEA plates (Gordon & Mihm, 1962) incubated at 30 °C for 5 days. After incubation, colonies were examined both with the naked eye and microscopically at a magnification of ×400 using a Nikon Optiphot binocular light microscope (Nikon, Tokyo, Japan) fitted with a long distance working objective.

5.3. Degradation tests

The degradation of hypoxanthine (0.5 %, w/v), tyrosine (0.5 %, w/v) and xanthine (0.5 %, w/v) were detected in GYEA; clearing of the insoluble compounds from under and around areas of growth was scored as a positive result.

5.4. Nutritional tests

The organisms were examined for their ability to use thirty-six compounds as sole sources of carbon for energy and growth. Similarly, nine compounds were examined as sole sources of carbon and nitrogen. The carbon, and carbon plus nitrogen, compounds were prepared as aqueous solutions, sterilised by filtration

using disposable filters (0.45 µm; Acrodisc, Gelman Sciences, 600 South Wagner Road, Ann Arbor, Michigan, USA) and added to the molten basal media (Boiron *et al.*, 1993; Appendix A). When scoring plates, growth on the test medium was compared with that on positive and negative control plates. The positive control plate contained glucose as the sole carbon source or glucose plus yeast extract as the sole carbon and nitrogen source; the negative control plates lacked a carbon source or a carbon plus nitrogen source, respectively. Strains were scored positive if growth on the test plate was greater than that on the negative control plate. *Conversely, negative results* were recorded where growth was less than or equal to that on the negative control plate.

5.5. Resistance to antibiotics

The ability of the test strains to grow in the presence of eleven antibiotics was examined. All but one of the antibiotics were sterilised by filtration of aqueous solutions. The exception, rifampicin, was dissolved with an aliquot of dimethylsulphoxide (DMSO; Sigma Ltd.) and the volume made up to the desired concentration with distilled water prior to filtration. The results were scored after 7 and 14 days. The growth of the test strains was compared with that on GYEA plates (Gordon & Mihm, 1962) lacking any antibiotic, cultures showing resistance were scored as positive.

5.6. Growth in the presence of chemical inhibitors

The test organisms were examined for their ability to grow on GYEA (Gordon & Mihm, 1962) supplemented with crystal violet and on GYEA supplemented with 5-fluorouracil. Aqueous solutions containing crystal violet and 5-fluorouracil were sterilised by autoclaving and filtration, respectively, then added to the molten GYEA medium to give the desired concentrations. The plates were read after 7 and 14

days. A positive result was scored when growth on the test plate was greater than or equal to that on the control plate which lacked a chemical inhibitor.

5.7. Growth at 45 °C

The test strains were examined for their ability to grow on GYEA plates (Gordon & Mihm, 1962; Appendix A) after incubation for 7 day's incubation at 45 °C.

5.8. Data analysis

The raw binary data were typed into the ***X*** program (Chapter I) in a +/- format. Test error for individual tests was calculated according to Sneath and Johnson (1992) using the '***Test Error***' procedure implemented in the ***X*** program. A similarity matrix was calculated and clustering carried out using the procedures '***SIMQUAL***' and '***SAHN***' from the ***NTSYS-pc*** program written by F. J. Rolf (Exeter Software, 100 North Country Road, Building B Setauket, New York 11733, USA). The cophenetic correlation coefficient (Sokal & Rolf, 1962) was calculated using the ***NTSYS*** procedures '***COPH***' and '***MXCOMP***'. Three similarity/dissimilarity coefficients, namely the simple matching (S_{SM}), Jaccard (S_J) and pattern (D_P) coefficients, were used. Since the ***NTSYS*** package does not contain the procedure for the D_P coefficient, the pattern distance matrix was calculated by using the ***X*** program, and clustering achieved by using the ***NTSYS*** package. Principal component analysis was carried out using the ***NTSYS*** procedure '***EIGEN***'.

6. Sequencing of 16S rDNA

6.1. Test strains and procedure

16S rRNA genes of isolate N1171 and *Tsukamurella paurometabola* strains JC7^T, M334 and N663 were sequenced by following the procedure described in detail in Chapter II.

6.2. Data analysis

The primary structures of the resultant nucleotide sequences were aligned manually by using the ***AL16S*** program (Chapter I). The reference sequences were obtained from the ribosomal database project release 4.0 (RDP; Larsen *et al.*, 1993) and from the GenBank/EMBL database. The sequence similarities and evolutionary distances of Jukes and Cantor correction (1969) were calculated by using the ***AL16S*** program which was run on an IBM-PC compatible computer. The phylogenetic tree was constructed using the neighbor-joining method (Saitou & Nei, 1987) employing the program ***NEIGHBOR*** as implemented in the ***PHYLIP*** package (Felsenstein, 1993). The resultant phylogenetic tree was heuristically evaluated by using the bootstrap method (Felsenstein, 1985) which is also included in the ***PHYLIP*** package.

C. Results

1. Selective isolation

The GYEA isolation plates supported the growth of many orange-reddish colonies after 5 days incubation at 30 °C. Six representative strains taken from the isolation plates were found to be pure when samples were stained using the Zeihl-Neelson and Gram stains.

2. Morphology and staining

All six isolates were Gram-positive and weakly acid-alcohol fast. The isolates formed straight and slightly curved rods, which occurred singly, in pairs and in masses, but did not differentiate into substrate or aerial hyphae. The colonies were large (diameter over 5mm), orange to red in colour with irregular edges and an irregular elevation.

3. Chemotaxonomy

Fatty acids. Thin-layer-chromatographic analysis of whole-organism methanolysates of the isolates revealed the presence of single spots (R_f 0.4-0.6) corresponding to mycolic acid methyl esters. It is evident from Table 3-4 that both the isolates and *Tsukamurella paurometabola* strains M334 and N663 showed qualitatively similar cellular fatty acid profiles which consisted mainly of saturated and unsaturated straight chain fatty acids and tuberculostearic acid (fatty acid type 1b *sensu* Kroppenstedt 1985; Table 2-3). In all cases hexadecanoic and monounsaturated octadecanoic acids were the predominant components. Isolate N1176 contained proportionally less tuberculostearic acid than the remaining activated sludge isolates.

TABLE 3-4. COMPOSITION OF CELLULAR FATTY ACIDS OF SLUDGE ISOLATES AND REPRESENTATIVE STRAINS OF *TSUKAMURELLA PAUROMETABOLA*.

Strain	Fatty acid composition (%)								
	$C_{14:0}$	$C_{16:1}$	$C_{16:0}$	$C_{18:1}$	$C_{18:0}$	TSA	$C_{19:1}$	$C_{20:1}$	$C_{20:0}$
T. paurometabola M334	2	4	36	30	11	10		6	tr.
T. paurometabola N663	2	8	31	22	12	21	2	2	
Isolate N1171	3	11	37	28	3	15	2	tr.	
Isolate N1172	12	7	32	25	2	20	tr.	2	
Isolate N1173	13	9	30	31	tr.	16			
Isolate N1174	10	8	30	34	tr.	16			
Isolate N1175	11	7	33	28	2	19			
Isolate N1176	2	7	44	43		4			

Abbreviations: $C_{14:0}$, straight chain tetradecanoic acid; $C_{16:1}$, monounsaturated hexadecanoic acid; $C_{16:0}$, hexadecanoic acid; $C_{18:1}$, monounsaturated octadecanoic acid; $C_{18:0}$, octadecanoic acid; TSA, tuberculostearic acid (10-methyloctadecanoic acid); $C_{19:1}$, monounsaturated nonadecanoic acid; $C_{20:1}$, monounsaturated eicosanoic acid; and $C_{20:0}$, eicosanoic acid. tr., trace (less than 1%).

Mycolic acids. Methyl esters of mycolic acids fragment on mass spectrometry in competing pathways depending on their overall size and structural features (Figure 3-3). The mass spectra of the MAMEs of isolate N1171 showed medium sized peaks at *m/z* 292 and 320 corresponding to $C_{20:1}$ and $C_{22:1}$ unsaturated fatty acids (fragments X; Figure 3-4). Fragments corresponding to aldehydes were seen with *m/z* values 736 ($C_{52:4}$) and C762 ($C_{54:5}$). In the higher mass range, fragments corresponding to anhydromycolates, formed by elimination of a molecule of water, were detected at *m/z* 984 ($C_{68:6}$), 1010 ($C_{70:7}$) and 1038 ($C_{72:7}$).

Fatty acid methyl esters released from MAMEs at high temperatures (over 300 °C) can be detected by gas chromatography. The major species of side chains (long chain ester in Figure 3-3) revealed by gas chromatographic-mass spectrometric analysis of the MAMEs extracted from isolate N1171 were $C_{20:1}$ and $C_{22:1}$ (Figure 3-5). This result is in good agreement with the data derived from the mass spectrometric analysis.

The mycolic acid methyl esters of the isolates were not precipitated when solutions in acetonitrile/toluene (1:2, v/v) were treated with additional acetonitrile.

Isoprenoid quinones. The isoprenoid quinones of the six isolates were identified as menaquinones as they were comigrated with the vitamin K (Sigma). All of the strains contained unsaturated menaquinones with nine isoprenoid units (MK-9) as the major isoprenologue though significant amounts of MK-7 were also found (Figure 3-6).

Isomers of diaminopimelic acid. The *meso-* and LL-diaminopimelic acid isomers in the standard were well separated on the TLC plates. Whole-organism hydrolysates of all six isolates contained *meso-* diaminopimelic acid.

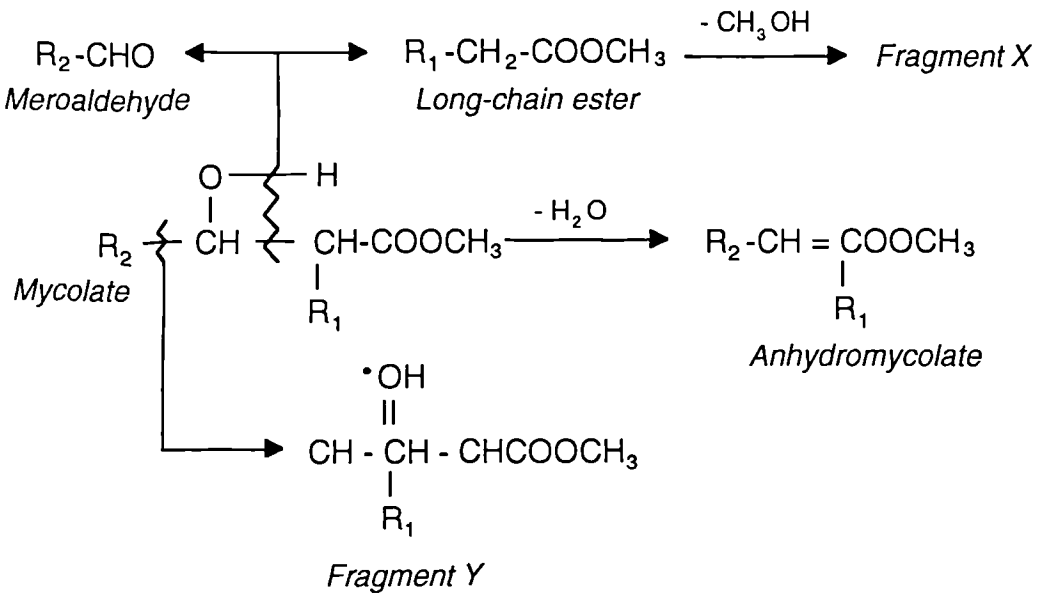


FIGURE 3-3. MASS SPECTRAL FRAGMENTATION PATHWAYS CHARACTERISTIC OF MYCOLIC ACID METHYL ESTERS. THE MOLECULAR MASSES OF ANHYDROMYCOLATES WERE CALCULATED USING THE EQUATION:

$m/z = 86 + [(\text{NUMBER OF CARBONS}-3) \times 14] - (\text{NUMBER OF DOUBLE BONDS} \times 2)$

MODIFIED FROM COLLINS *et al.* (1982).

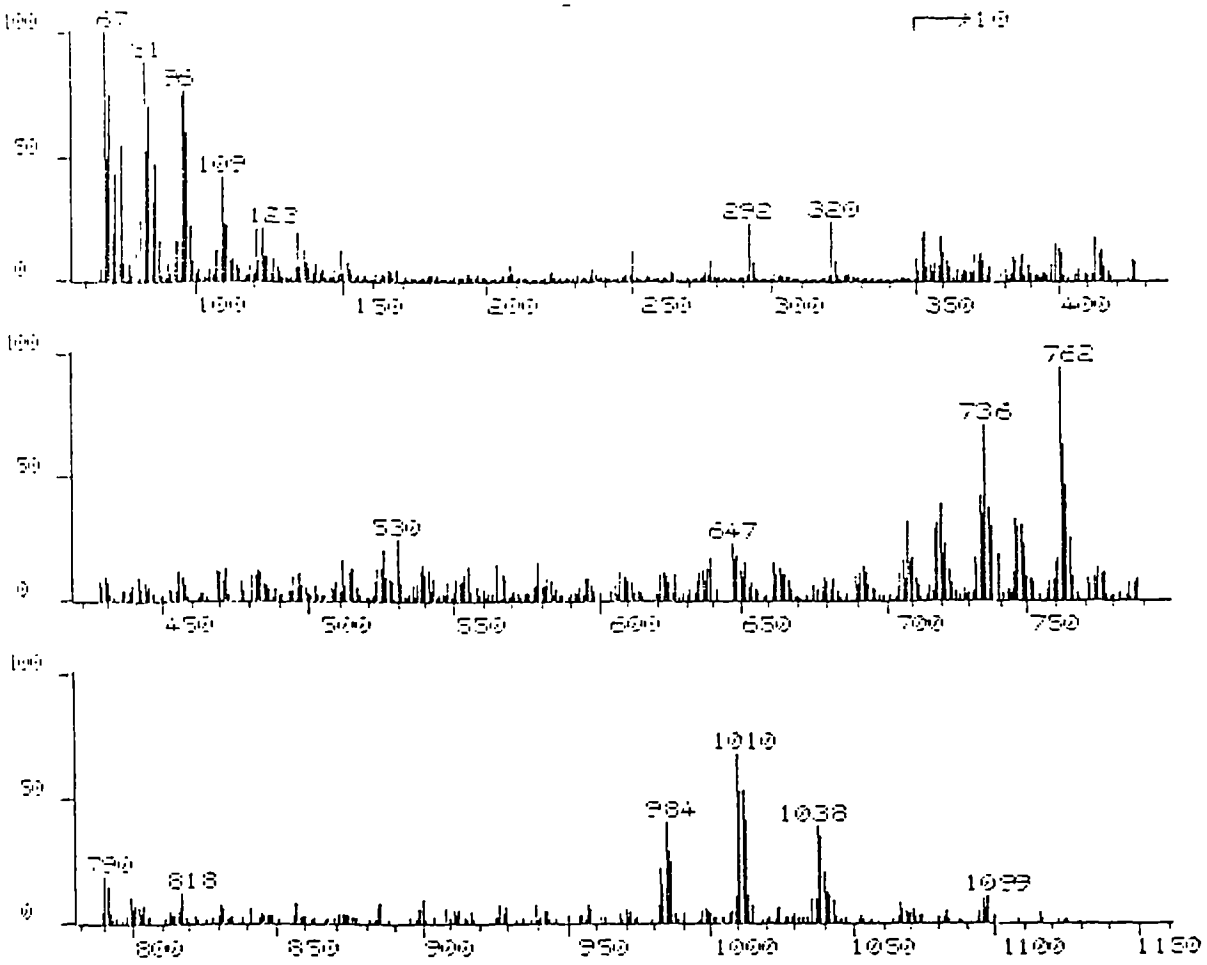
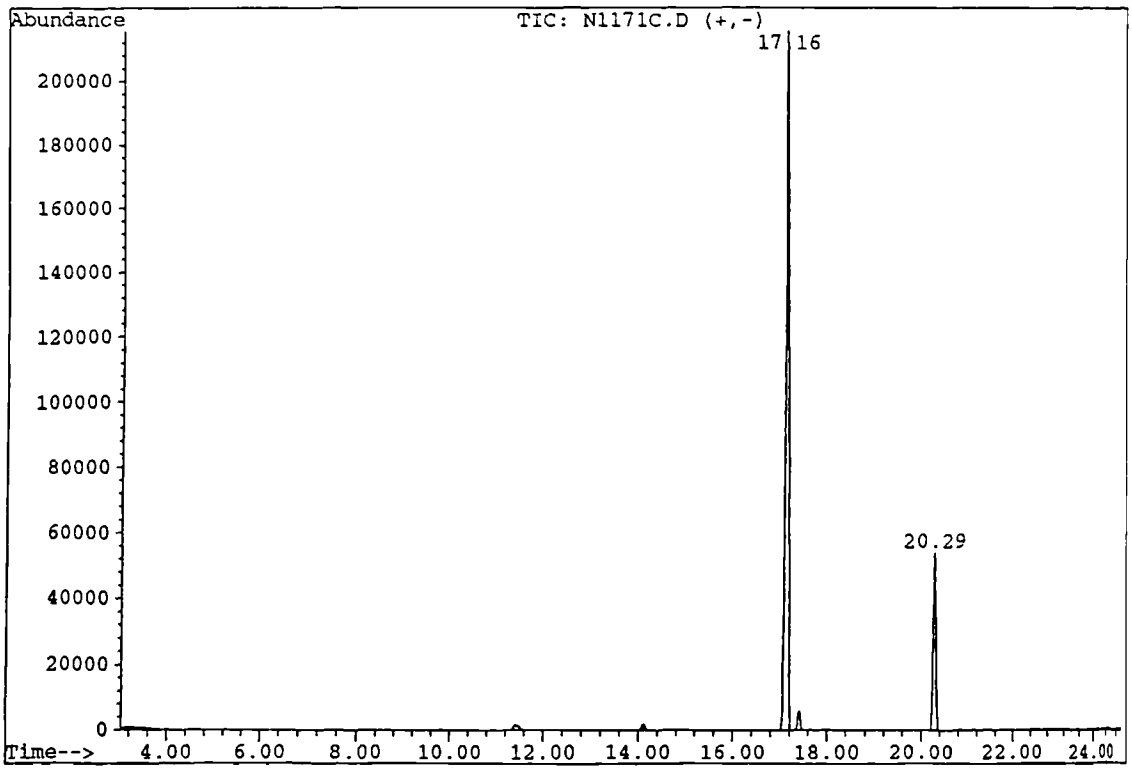
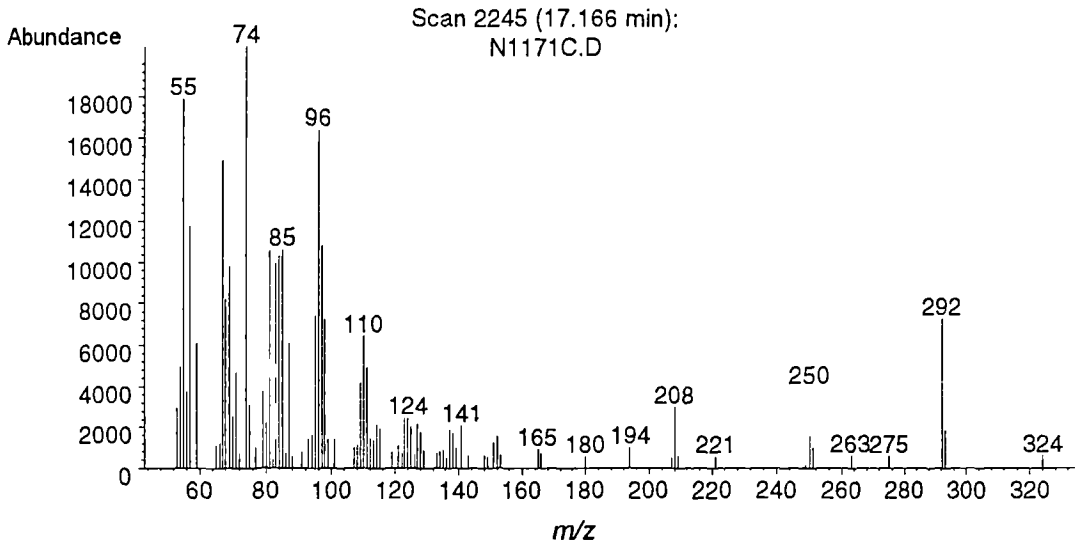


FIGURE 3-4. PARTIAL MASS SPECTRUM OF THE METHYL ESTERS OF MYCOLIC ACIDS EXTRACTED FROM ISOLATE N1171.

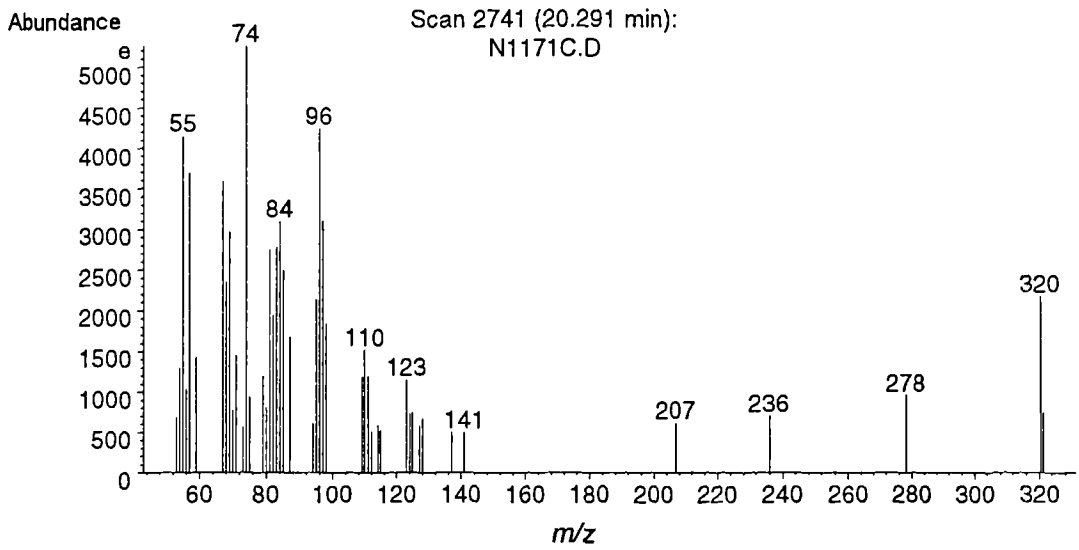


(a) Pyrolysis gas chromatogram

FIGURE 3-5. PYROLYSIS GAS CHROMATOGRAPHIC ANALYSIS OF MYCOLIC ACID METHYL ESTERS EXTRACTED FROM ISOLATE N1171. (a) GAS CHROMATOGRAM SHOWING TWO MAJOR PEAKS THAT CORRESPOND TO THE FATTY ACID METHYL ESTERS $C_{20:1}$ (RETENTION TIME, 17.2 MINUTES) AND $C_{22:1}$ (20.3 MINUTES), (b) CORRESPONDING MASS SPECTRA OF THE PEAKS AT THE RETENTION TIMES OF 17.2 MINUTES, AND (c) 20.3 MINUTES.



(b) $C_{20:1}$ (molecular ion = 324; molecular ion - methanol = 292)



(c) $C_{22:1}$ (molecular ion = 352; molecular ion - methanol = 320)

FIGURE 3-5. CONTINUED.

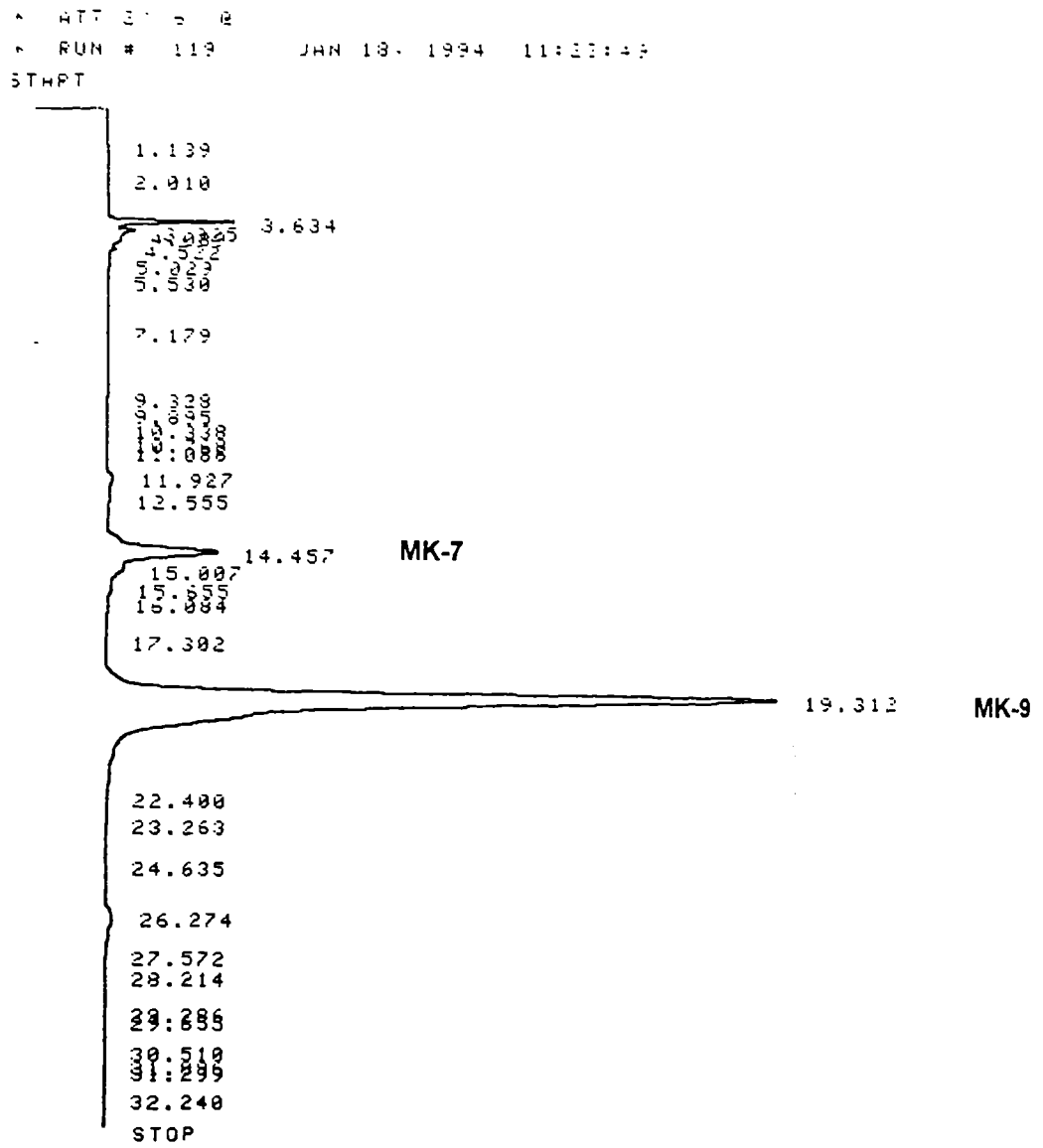


FIGURE 3-6. HIGH-PERFORMANCE LIQUID CHROMATOGRAM SHOWING THE MENAQUINONE PROFILE OF ISOLATE N1171.

Whole-organism sugars. All six isolates contained major amounts of arabinose and galactose in the whole-organism hydrolysates.

4. Numerical phenetic classification

4.1. Final data matrix and calculation of test error

The preliminary data matrix contained information on twelve test strains, including the four duplicated cultures, namely isolate N1171 and *Tsukamurella paurometabola* strains JC7^T, M334 and N663, and ninety-six unit characters. Fifty unit characters were excluded from the raw data set as they gave either all positive or all negative results.

All of the test strains used D(-)-fucose (1 %, w/v), D(+)-galactose (1 %, w/v), D(+)-turanose (1 %, w/v), amyl alcohol (0.1 %, v/v) and sodium pyruvate (0.1 %, w/v) as sole carbon sources and L-asparagine (0.1 %, w/v) as a sole carbon and nitrogen source. They also grew in the presence of crystal violet (0.001%, w/v) and 5-fluorouracil (20 µg/ml), and were resistant to bekanamycin (16, 32 and 64 µg/ml), erythromycin (2, 4 and 8 µg/ml), gentamycin sulphate (16 and 32 µg/ml), kanamycin (4, 8, 16 and 64 µg/ml), neomycin sulphate (4, 8 and 16 µg/ml), oleandomycin phosphate (16 and 32 µg/ml), rifampicin (0.5 and 2 µg/ml) and vancomycin hydrochloride (1, 2 and 4 µg/ml).

In contrast, none of the test strains used adonitol (ribitol; 1%, w/v), L(+)-arabinose (1%, w/v), dulcitol (galactitol; 1%, w/v), *meso*-erythritol (1%, w/v), D(+)-fucose (1%, w/v), D-galacturonic acid (1%, w/v), D(+)-glucosamine (1%, w/v), D-glucuronic acid (1%, w/v), methyl-α-D-mannopyranoside (1 %, w/v), α-L-rhamnose (1 %, w/v), D(+)-xylose (1 %, w/v), *p*-aminosalicylic acid (0.1 %, w/v), methanol (0.1 %, v/v), DL-norleucine (0.2 %, w/v), resorcinol (0.1 %, w/v) or sodium tartrate (0.1 %, w/v) as sole sources of carbon for energy and growth or L-lysine (0.1 %,

w/v) as a sole carbon and nitrogen source. None of the test strains grew at 45 °C or in the presence of novobiocin (64 µg/ml).

The test variance (S_i^2, formula 15; Sneath & Johnson, 1972) and the probability of error (P_i, formula 4; Sneath & Johnson, 1972) between the four duplicated cultures were calculated to determine test error. The duplicated cultures gave the same responses in 84 out of the 96 tests. One test, namely chlortetracycline hydrochloride (4 µg/ml) showed a high test variance, 0.25, and was deleted from the data set. Five tests, namely the ability to grow in the presence of rifampicin (8 µg/ml) and crystal violet (0.01 %, w/v), and L-serine (0.1 %, w/v), L-leucine (0.1 %, w/v) and L-valine (0.1 %, w/v) as sole carbon and nitrogen sources, showed test variances of 0.125. The overall test variance was 0.014 and the average probability of an erroneous test result (p) calculated from the pooled variance (S_i^2) for the tests in the final matrix was 1.41%. The final data matrix contained information on twelve strains and forty-five unit characters.

4.2. Cophenetic correlation coefficient

Cophenetic correlation coefficients were determined for dendrograms based on the Jaccard (S_J), simple matching (S_{SM}) and pattern (D_P) coefficients with both the single linkage and the unweighted pair group method with arithmetic averages (UPGMA) algorithms. In all cases, the highest cophenetic correlation values were obtained with the UPGMA algorithm (Table 3-5). The highest cophenetic correlation value, 0.89797, was found in the S_{SM}, UPGMA analysis. It is also worth noting that the cophenetic correlation values based on the D_P coefficient were significantly smaller than those based on the S_{SM} and S_J coefficients.

TABLE 3-5. COMPARISON OF DENDROGRAMS AND CORRESPONDING SIMILARITY MATRICES USING THE COPHENETIC CORRELATION COEFFICIENT.

Coefficient	Clustering algorithm	Cophenetic correlation
S_J	Single linkage	0.84458
	UPGMA*	0.85661
S_{SM}	Single linkage	0.88797
	UPGMA	0.89797
D_P	Single linkage	0.55207
	UPGMA	0.60412

*, unweighted pair group method with arithmetic averages algorithm.

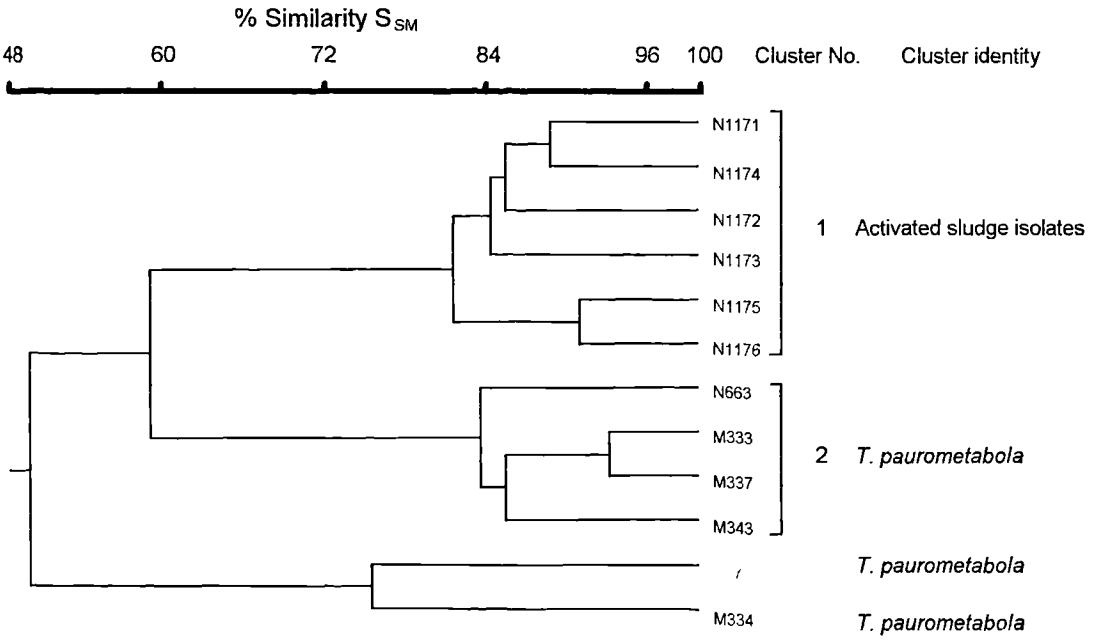
4.3. Numerical phenetic classification based on the S_{SM} coefficient and the unweighted pair group method with arithmetic averages algorithm

The numerical classification based on the S_{SM}, UPGMA analysis is considered in detail because it showed the highest cophenetic correlation value. The twelve test strains were assigned to two multimembered and two single membered clusters defined at the 80% similarity level (S-level; Figure 3-7a). The six activated sludge isolates formed cluster 1 at the 81.7% S-level. Four out of the six *Tsukamurella paurometabola* strains formed cluster 2 at the 83.7% S-level; the two remaining *Tsukamurella paurometabola* organisms, strains JC7^T and M334, formed single membered clusters which were loosely associated at the 75.6% S-level.

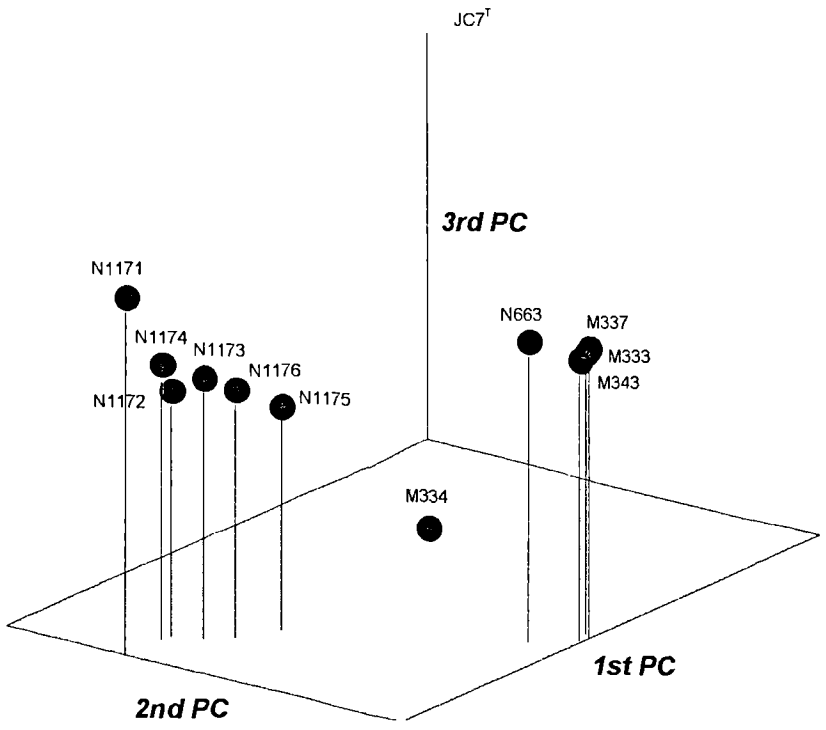
It is evident from the dendrogram that the strains isolated from the activated sludge foams formed a distinct cluster that can readily be distinguished from all of the clusters containing the *Tsukamurella paurometabola* strains. The distribution of positive characters to the four numerically defined clusters is shown in Table 3-6. Characters which can be given presumptive weight to separate the four numerically circumscribed taxa are marked with an asterisk.

4.4. Other numerical phenetic classifications

The two multimembered and two single membered clusters were also recovered in the dendrograms based on the S_{SM}, single linkage (SL), the S_J, UPGMA, and the S_J, SL analyses; the cut-off points used in these analyses were 84%, 68% and 75% S-levels, respectively. In contrast, the six sludge isolates and the six *Tsukamurella paurometabola* strains were assigned to two distinct clusters in both the S_P, UPGMA and the S_P, SL analyses; in each case the clusters were defined at the 0.79 S_P level. However, less confidence can be placed on the results of these two analyses given the very low cophenetic correlation values (Table 3-5).



(a)



(b)

FIGURE 3-7. (a) DENDROGRAM AND (b) ORDINATION PLOT SHOWING RELATIONSHIPS BETWEEN THE SLUDGE ISOLATES AND THE *TSUKAMURELLA PAUROMETABOLA* STRAINS BASED ON 45 PHENOTYPIC CHARACTERS. THE DENDROGRAM WAS GENERATED BY USING THE SIMPLE MATCHING COEFFICIENT AND THE UNWEIGHTED PAIR GROUP METHOD WITH ARITHMETIC AVERAGES ALGORITHM. THE ORDINATION PLOT WAS BASED ON THE RESULTS OF A PRINCIPAL COMPONENT ANALYSIS; THE FIRST THREE AXES ACCOUNTED FOR 69.09% OUT OF THE TOTAL VARIATION. ^{T}, TYPE STRAIN.

TABLE 3-6. DISTRIBUTION OF POSITIVE CHARACTERS TO CLUSTERS DEFINED AT THE 80% SIMILARITY LEVEL IN THE S_{SM}, UPGMA ANALYSIS.

	Cluster 1 (isolates)	Cluster 2 (*T. pauro-metabola* strains)	*Tsukamurella paurometabola*	
Number of organisms/strain number	6	4	JC7^T	M334
Character				
Morphological tests:				
Colony colour, white/creamy*	0	4	1	1
Colony colour, orange/red*	6	0	0	0
Colony edge, entire	0	0	1	0
Colony edge, irregular	6	4	0	1
Colony elevation, convex	0	0	1	0
Colony elevation, irregular	6	4	0	1
Colony size, small (<2mm)	0	0	1	0
Colony size, large (>5mm)	6	4	0	1
Degradation (%, w/v):				
Hypoxanthine (0.5)	6	4	0	0
Tyrosine (0.5)	6	4	0	0
Xanthine (0.4)	6	4	0	0
Resistance to antibiotics (µg/ml):				
Chlortetracycline hydrochloride (2)	1	4	1	1
Chlortetracycline hydrochloride (8)	0	3	0	0
Erythromycin (16)*	0	4	1	1
Gentamicin sulphate (64)	4	4	1	1
Neomycin sulphate (32)	5	4	1	1
Novobiocin (16)	5	4	1	1
Novobiocin (32)	2	4	1	1
Oleandomycin phosphate (64)	5	4	1	1
Penicillin G (16)*	0	4	1	1
Penicillin G (32)*	0	4	1	0
Penicillin G (64)	0	4	1	0
Rifampicin (16)	0	2	1	1
Rifampicin (4)	2	4	1	1
Rifampicin (8)	1	3	1	1
Vancomycin hydrochloride (8)	2	4	1	1

TABLE 3-6. CONTINUED.

Character	Cluster 1 (isolates)	Cluster 2 (*T. pauro-metabola* strains)	*Tsukamurella paurometabola*	
Number of organisms/strain number	6	4	JC7^T	M334
Growth in presence of chemical inhibitor:				
Crystal violet (0.01 %, w/v)	4	3	0	0
Sole carbon and nitrogen source (0.1%, w/v):				
L-Histidine*	0	4	0	0
L-*iso*-Leucine	3	4	1	1
L-Leucine	3	3	1	1
L-Phenylalanine	6	4	0	0
L-Serine	6	4	0	0
Succinamide	0	0	0	1
L-Valine	0	2	0	0
Sole carbon source (1 %, w/v:)				
D(-)-Arabinose	0	2	0	0
D(+)-Arabitol	5	4	0	1
β-Gentibiose	0	1	0	0
meso-Inositol*	6	4	0	1
D(+)-Melezitose	6	4	0	0
Salicin*	0	4	0	1
D-Sorbitol	6	4	0	1
Xylitol	5	4	0	1
Sole carbon source (0.1 %, w/v or v/v):				
Butane-1,3-diol	2	2	0	0
Butane-1,4-diol	0	1	0	0
Butane-2,3-diol	6	4	0	0
Ethanolamine*	0	4	1	1
Propane-1,2-diol	2	4	0	0
Sodium benzoate	1	2	0	0
Tri-sodium citrate	5	4	1	1

*, Characters that can be weighted to differentiate between the four clusters.

4.5. Principal component analysis

It is evident from the three-dimensional ordination plot based on the first three principal component axes (Figure 3-7b) that the test strains fall into two multimembered and two single membered groups that correspond to the four clusters recovered in the S_{SM}, UPGMA classification. The first three axes accounted for 36.43 %, 23.65 % and 9.00 % out of the total variance.

5. Molecular systematics

Almost complete 16S rRNA gene sequences were obtained for sludge isolate N1171 and for *Tsukamurella paurometabola* strains JC7^T, M334 and N663 (1474 to 1476 nucleotides; positions between 28 and 1524 using *Escherichia coli* numbering [Appendix C]; Table 3-7); the corresponding nucleotide sequence accession numbers are given in Table 3-7. The primary structures of the 16S rDNA sequences of the test strains are given in the aligned form in Appendix I.

The 16S rDNA sequence similarities found between the four test strains are shown in Table 3-7. Similarities of over 99% were found between all four nucleotide sequences. *Tsukamurella paurometabola* strains JC7^T and M334 showed the highest similarity (99.59 %). Isolate N1171 was most closely related to *Tsukamurella paurometabola* strain N663 (99.53 %). Dissimilarities due to different secondary structures in helices 6 and 49 (Appendix D) were found in the length of the four 16S rDNA sequences (Figure 3-8). A evolutionary distance matrix generated according to Jukes and Cantor (1969) was used to construct an unrooted phylogenetic tree by applying the neighbor-joining method (Saitou & Nei, 1987; Figure 3-9). The four test

TABLE 3-7. MATRIX OF SIMILARITY VALUES (LOWER PART) AND THE NUMBER OF DIFFERENCES (UPPER PART) IN THE 16S RDNA SEQUENCES OF SLUDGE ISOLATE N1171 AND *TSUKAMURELLA PAUROMETABOLA* STRAINS JC7^T, M334 AND N663.

Strain	Accession number	Numbers of nucleotides determined	Strains			
			JC7^T	N663	M334	N1171
JC7^T	Z46751	1474	———	8/1474	6/1472	9/1472
N663	Z36933	1476	99.46	———	9/1474	7/1474
M334	Z37151	1476	99.59	99.39	———	10/1472
N1171	Z37150	1474	99.39	99.53	99.32	———

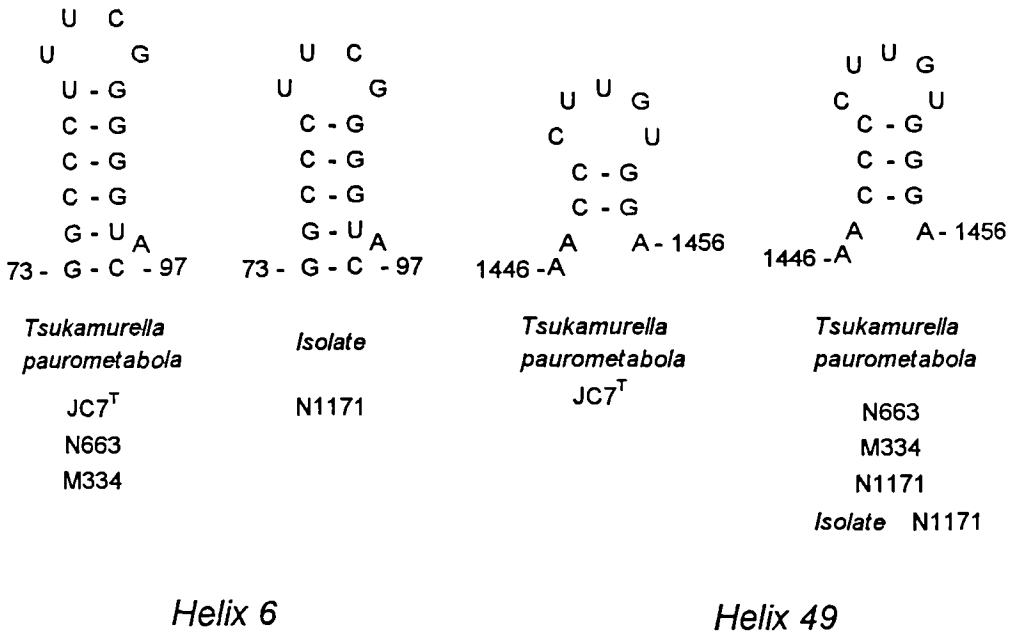


FIGURE 3-8. DIFFERENCES IN THE SECONDARY STRUCTURE OF THE 16S rRNA OF ISOLATE N1171 AND *TSUKAMURELLA PAUROMETABOLA* STRAINS JC7^T, M334 AND N663. HELIX NUMBERING ACCORDING TO NEEFS *et al.* (1993; APPENDIX D).

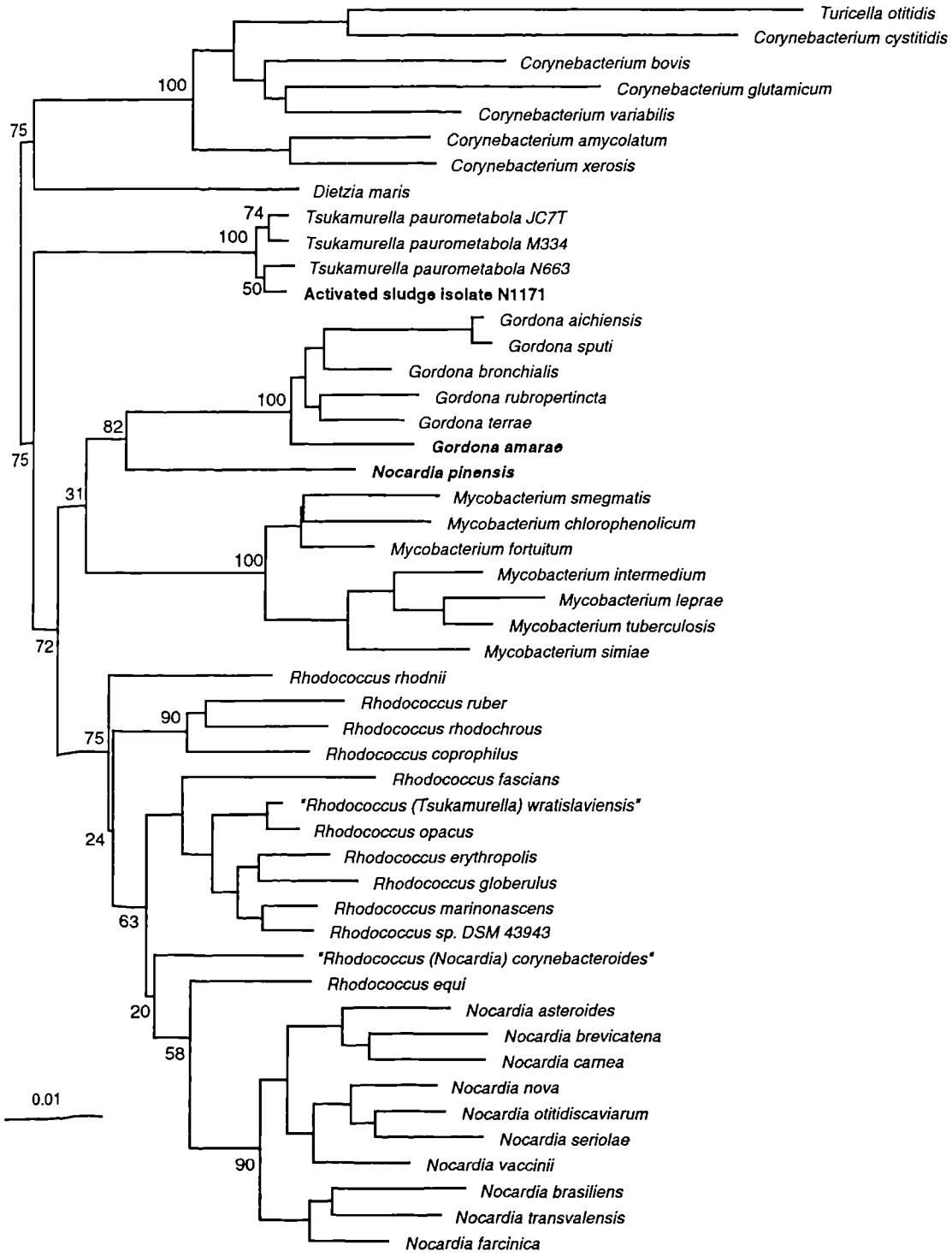


FIGURE 3-9. AN UNROOTED PHYLOGENETIC TREE SHOWING RELATIONSHIPS BETWEEN ISOLATE N1171 AND REPRESENTATIVES OF MYCOLIC ACID CONTAINING TAXA. THE TREE WAS CONSTRUCTED BY USING THE JUKES AND CANTOR DISTANCE (1969) AND THE NEIGHBOR-JOINING METHOD (SAITOU & NEI, 1987). THE NUMBERS AT THE NODES INDICATE THE LEVELS OF BOOTSTRAP SUPPORT BASED ON 1000 RESAMPLINGS. THE SCALE BAR INDICATES 0.01 SUBSTITUTIONS PER NUCLEOTIDE POSITION. ORGANISMS ASSOCIATED WITH FOAMS ARE INDICATED IN BOLD TYPE. CORRESPONDING ACCESSION NUMBERS ARE GIVEN IN APPENDIX G.

strains formed a tight clade, with 100% bootstrap support, that was distinct from clades composed of representatives of mycolic acid containing genera. Two subgroups, namely isolate N1171 plus *Tsukamurella paurometabola* strain N663 and *Tsukamurella paurometabola* strain JC7^T plus M334, were supported by low bootstrap values, that is, 74% and 50%, respectively.

6. Chemical fingerprints

6.1. Curie point pyrolysis mass spectrometry

The six sludge isolates were compared with the six representative strains of *Tsukamurella paurometabola* by using Curie point pyrolysis mass spectrometry. The pyrolysis mass spectra derived from sludge isolate N1171 and *Tsukamurella paurometabola* strains JC7^T, M334, N663 are given in Figure 3-10. It is evident from both the dendrogram (Figure 3-11a) and the three dimensional diagram (Figure 3-11b) that all four of the duplicated cultures, namely JC7^T, M334, N663 and N1171, clustered together.

It is also evident from Figure 3-11 that the test strains formed two multimembered and two single membered groups which corresponded to the numerical phenetic clusters highlighted in the S_{SM}, UPGMA analysis. Cluster 1 encompassed the six sludge isolates and cluster 2 the *Tsukamurella paurometabola* strains M333, M337, M343 and N663. *Tsukamurella paurometabola* strains JC7^T and M334, which were loosely associated at the 70% S-level in the numerical phenetic classification, were recovered separately as outliers. The type strain of *Tsukamurella paurometabola* (JC7^T) was sharply separated from all of the other *Tsukamurella paurometabola* strains.

6.2. Whole-organism protein electrophoresis

Whole-organism protein profiles of the twelve test strains obtained from the SDS-PAGE analysis were analysed by using the ***GelCompar*** package (Vauterin & Vauterin, 1992). The relationships found between the test strains based on the normalised gel images are given in Figure 3-12. The two sets of duplicated cultures, namely sludge isolate N1171 and *Tsukamurella paurometabola* strain JC7^T, each clustered together. All of the sludge isolates were assigned to the same group defined at a correlation coefficient (*r*) level of 67% though the protein profile of sludge isolate N1171 differed from those of the other isolates. *Tsukamurella paurometabola* strains M333, M337, M343 and N663, that is, the members of cluster 2 defined in the numerical phenetic study based on the S_{SM}, UPGMA analysis, formed a fairly heterogeneous group defined at a *r*-level of 54% Two remaining organisms, *Tsukamurella paurometabola* strains JC7^T and M334, joined together at the 54% *r*-level.

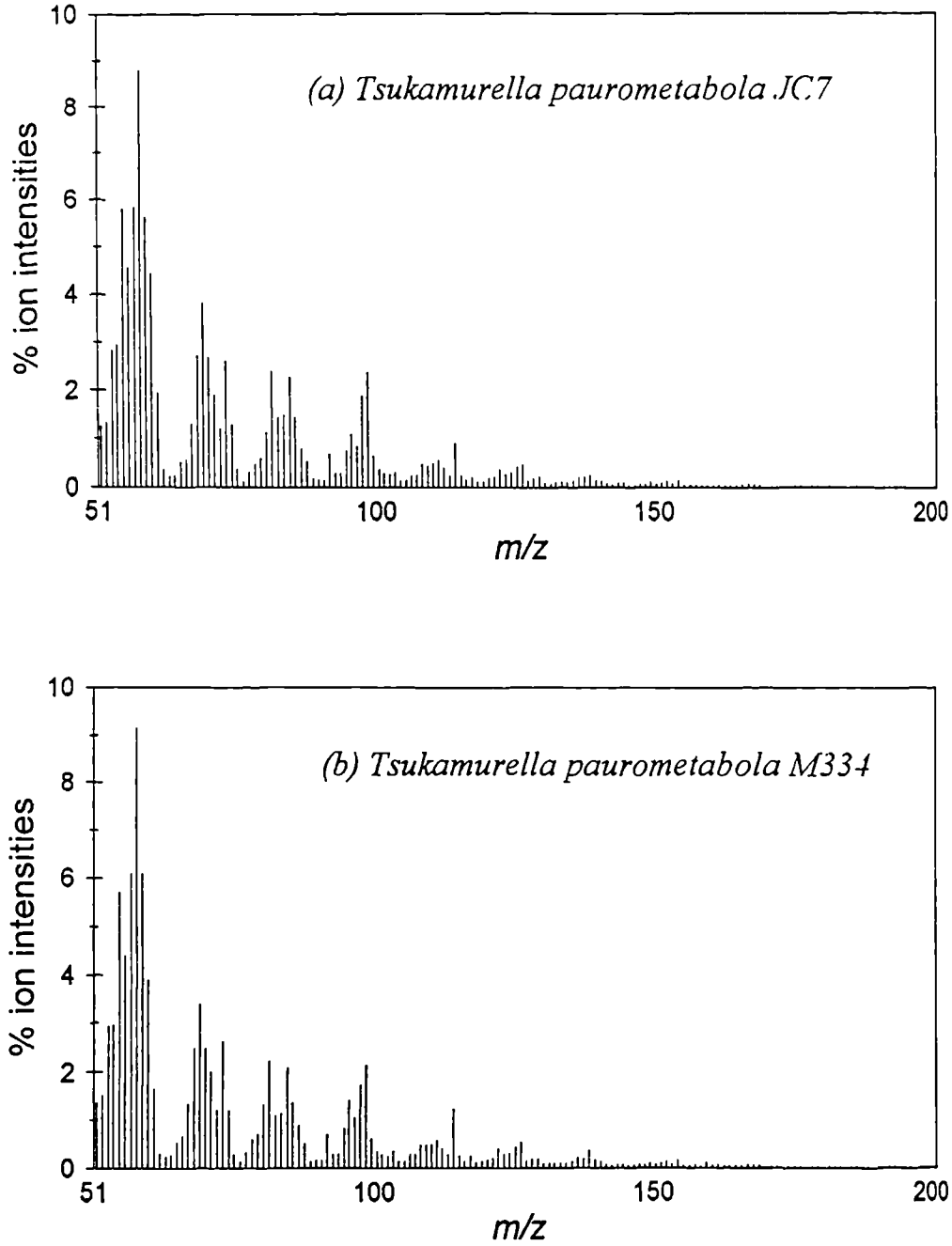
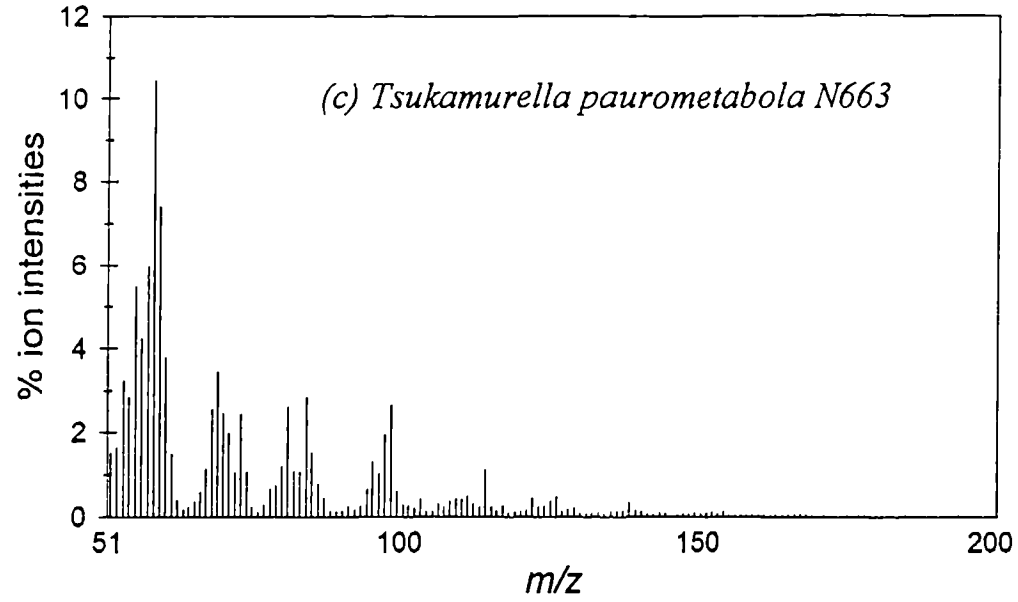


FIGURE 3-10. PYROLYSIS MASS SPECTRA DERIVED FROM THE ANALYSIS OF (a) *TSUKAMURELLA PAUROMETABOLA* STRAIN JC7^T, (b) *TSUKAMURELLA PAUROMETABOLA* STRAIN M334, (c) *TSUKAMURELLA PAUROMETABOLA* STRAIN N663, AND (d) SLUDGE ISOLATE N1171.



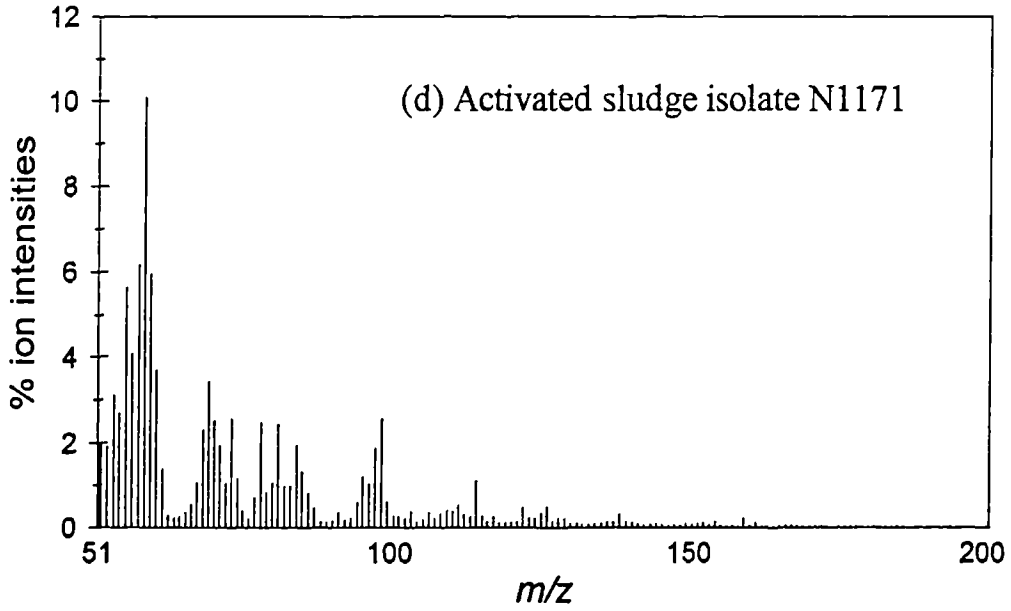


FIGURE 3-10. CONTINUED.

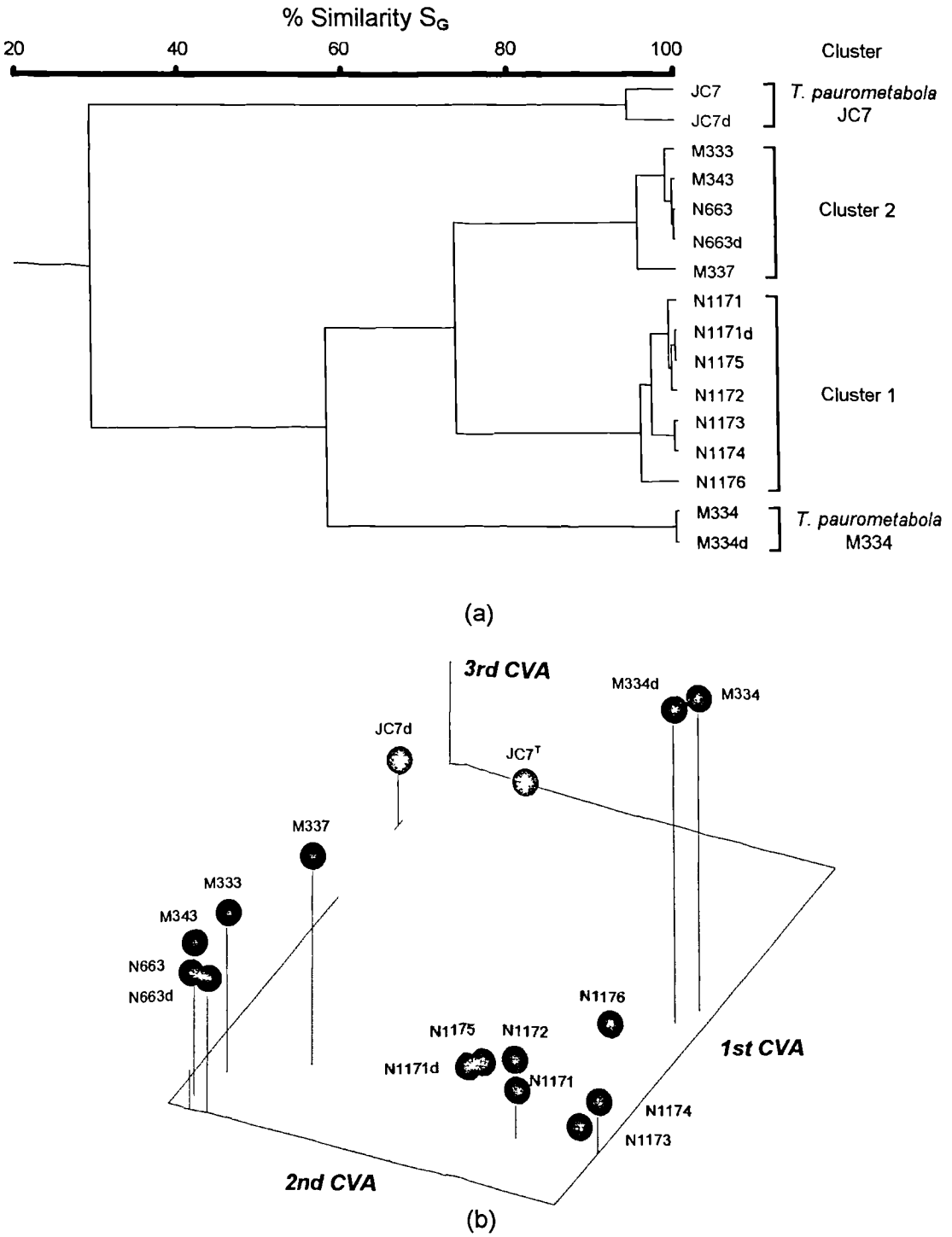


FIGURE 3-11. (a) DENDROGRAM REPRESENTING RELATIONSHIPS FOUND BETWEEN THE SLUDGE ISOLATES AND *TSUKAMURELLA PAUROMETABOLA* STRAINS BASED ON PYROLYSIS MASS SPECTRAL DATA. THE DATA WERE ANALYSED USING THE GENSTAT PACKAGE WITH CLUSTERING ACHIEVED BY APPLYING THE UNWEIGHTED-PAIR GROUP METHOD WITH ARITHMETIC AVERAGES. (b) THREE DIMENSIONAL ORDINATION PLOT DERIVED FROM THE PYROLYSIS MASS SPECTRAL ANALYSIS. THE FIRST THREE CANONICAL VARIATE AXES REPRESENT 45.46%, 39.42% AND 4.58% OUT OF THE TOTAL VARIATION. THE DUPLICATED CULTURES ARE INDICATED BY d.

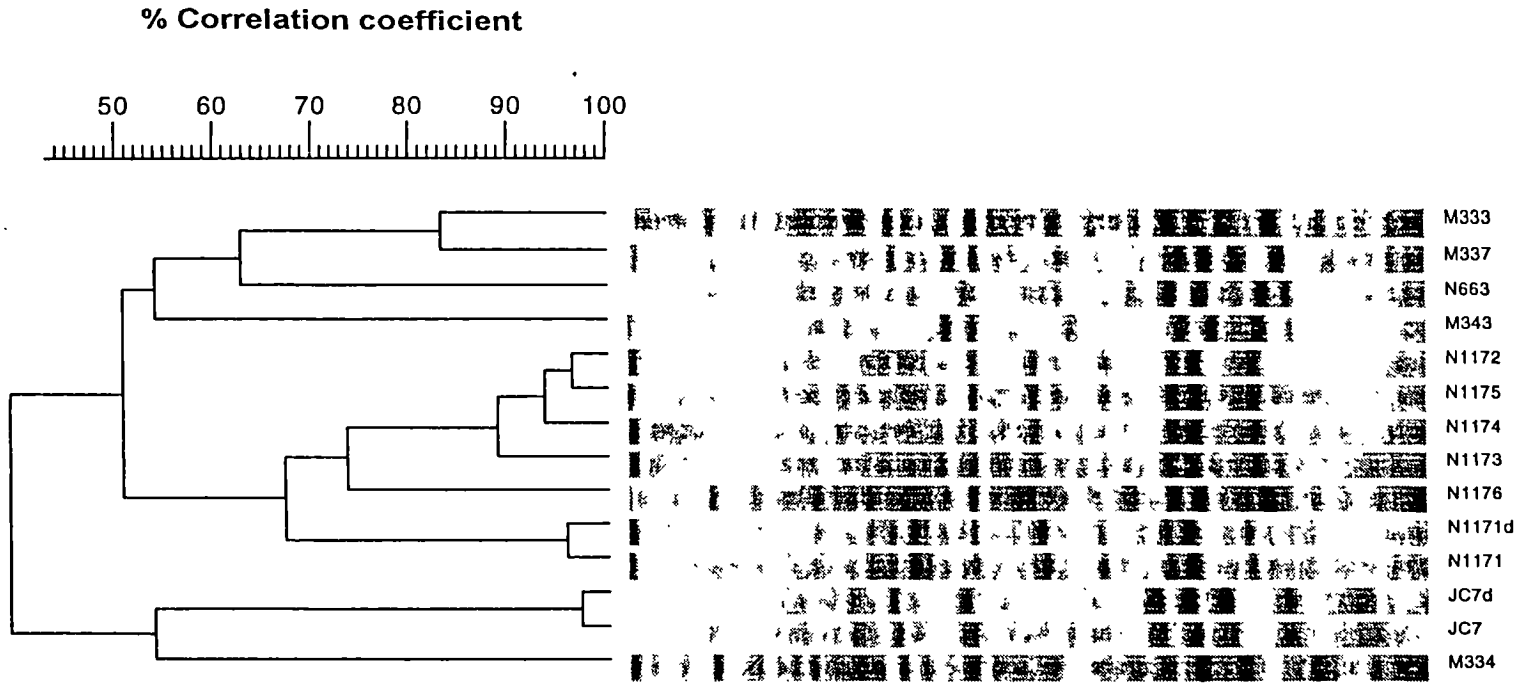


FIGURE 3-12. DENDROGRAM SHOWING RELATIONSHIPS BETWEEN THE SLUDGE ISOLATES AND THE *TSUKAMURELLA PAUROMETABOLA* STRAINS. THE DENDROGRAM IS BASED ON MEAN CORRELATION COEFFICIENT (*r*) VALUES WHICH WERE GROUPED BY USING THE UNWEIGHTED PAIR GROUP METHOD WITH ARITHMETIC AVERAGES ALGORITHM. COMPUTER-PROCESSED GEL IMAGES ARE ALSO PRESENTED. THE *r* VALUES WERE CONVERTED TO PERCENTAGES ACCORDING TO POT *et al.* (1994). d INDICATES DUPLICATED CULTURES.

<BLANK>

D. Discussion

There have been surprisingly few taxonomic studies on microbial agents that cause or are associated with foaming in activated-sludge sewage-treatment plants (Soddell & Seviour, 1990; Jenkins *et al.*, 1993). In most studies, sludge isolates have merely been characterised and identified by using a few phenotypic properties (Lechevalier *et al.*, 1976; Sezgin *et al.*, 1988). It has, therefore, been difficult to determine the precise taxonomic status of microorganisms associated with activated sludge foaming.

The species is usually considered to be the primary, basic and stable unit in bacterial systematics but there is no universally accepted definition of the bacterial species (Goodfellow & O'Donnell, 1993, 1994). In practice, a somewhat fluid species concept is emerging based on the use of both genotypic and phenotypic data. The term **'polyphasic taxonomy'** was introduced by Colwell (1970) to signify successive or simultaneous taxonomic studies on groups of organisms using an array of techniques designed to provide both genotypic and phenotypic data. Polyphasic taxonomic studies can be expected to yield well defined species, a stable nomenclature and improved species definitions. In practice, it is not always possible for individual research groups to gain access to a variety of taxonomic techniques. Consequently, few polyphasic taxonomic studies have been carried out on actinomycetes (e.g., De Boer *et al.*, 1990; Kroppenstedt *et al.*, 1990; Goodfellow *et al.*, 1991). In the present investigation, representative strains isolated from activated sludge foams were examined using the polyphasic taxonomic approach.

The speed and cost of 16S rDNA sequencing methods is now comparable to that of other taxonomic methods mainly due to improvements in molecular biology, notably the invention and application of automated DNA sequencing procedures. It is evident from the 16S rDNA sequence data generated in the present study that the

sludge isolates not only belong to the genus *Tsukamurella* but are closely related to *Tsukamurella paurometabola*. Indeed, all of the test strains, namely *Tsukamurella paurometabola* JC7^T, M334 and N663 and sludge isolate N1171, shared nucleotide sequence similarities of over 99%. It has been pointed out in Chapter II that *Tsukamurella wratislaviensis* Goodfellow *et al.* 1991 should be transferred to the genus *Rhodococcus* leaving *Tsukamurella paurometabola* as the only representative of the genus *Tsukamurella* Collins *et al.* 1988.

Collins *et al.* (1988a) determined the partial 16S rRNA sequences of *Tsukamurella paurometabola* strains JC7^T (1305 nucleotides; X53206) and N663 (1287 nucleotides; X53207) by using the reverse transcriptase sequencing technique (Lane *et al.*, 1985). The comparison between the two strains was limited to a consideration of only 1198 nucleotides as the authors had to cope with experimental difficulties. A number of differences were noted when the 16S rDNA sequences determined in the present study for *Tsukamurella paurometabola* strain JC7^T were compared with those of the corresponding sequence prepared by Collins *et al.* (1988a). Nine nucleotide differences were recorded for the corresponding sequences of *Tsukamurella paurometabola* strain JC7^T, that is, a dissimilarity value of 0.69%. In the corresponding analysis, sixteen nucleotide differences were observed for *Tsukamurella paurometabola* strain N663 giving a dissimilarity value of 1.28%.

Only four nucleotide differences out of a total of 1474, that is, a dissimilarity value of 0.27%, were found between the 16S rDNA sequences obtained for *Tsukamurella paurometabola* strain JC7^T and the corresponding sequence generated by Rainey *et al.* (1995a). The four differences were found in nucleotide positions 720, 1000, 1002 and 1003 using the *Escherichia coli* numbering system (Brosius *et al.*, 1978; Appendix C). The last three nucleotide positions mentioned

above form part of the stem of helix 37; the secondary structure of this region is shown in Figure 3-13. It is clear from the figure that the nucleotide sequence determined in the present investigation makes more sense than the one presented by Rainey *et al.* (1995a) given the estimated secondary structure of the 16S rRNA molecule of *Tsukamurella paurometabola* strain JC7^T.

The results of the chemotaxonomic studies provide further evidence of a close phylogenetic relationship between the activated sludge isolates and the genus *Tsukamurella*. Indeed, all six activated sludge isolates contained *meso*-diaminopimelic acid as the major wall diamino acid and major amounts of arabinose and galactose (wall chemotype IV *sensu* Lechevalier and Lechevalier 1970), were rich in straight chain, unsaturated and tuberculostearic acids (fatty acid type 1b *sensu* Kroppenstedt 1985), contained fully unsaturated menaquinones with nine isoprene units (MK-9) as the predominant isoprenologue and had highly unsaturated long chain mycolic acids with 68 to 72 carbon atoms which were not precipitated by the procedure of Hamid *et al.* (1993). All of these chemical properties are consistent with the assignment of the sludge strains to the genus *Tsukamurella* (Collins *et al.*, 1988a; Goodfellow, 1992; see Table 2-11).

It is clear from the numerical taxonomic analyses that the sludge isolates formed a homogeneous cluster that was readily distinguished from clusters containing the *Tsukamurella paurometabola* strains. The isolates also formed a distinct homogeneous group in the Curie point pyrolysis mass spectrometric and whole-organism protein electrophoretic analyses. These data strongly suggest that the sludge isolates form the nucleus of a novel species in the genus *Tsukamurella*.

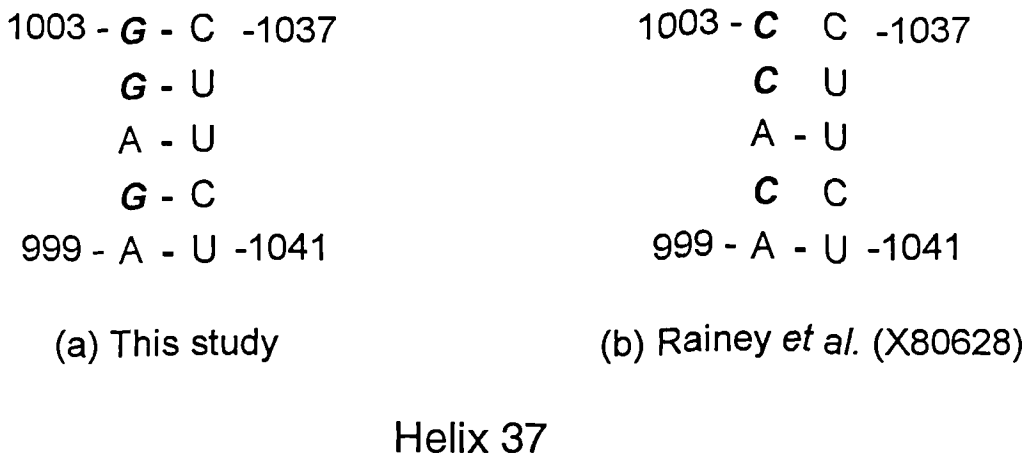


FIGURE 3-13. DIFFERENCES IN SECONDARY STRUCTURE FOUND BETWEEN 16S rDNA SEQUENCES OF *TSUKAMURELLA PAUROMETABOLA* JC7^T WHEN DATA FROM THE PRESENT STUDY WERE COMPARED WITH CORRESPONDING DATA FROM RAINEY *et al.* (1995a). THE BASES WHICH WERE FOUND TO DIFFER ARE HIGHLIGHTED BY THE USE OF BOLD-ITALIC CHARACTERS.

Tsukamurella paurometabola was proposed by Collins *et al.* (1988a) to accommodate members of two morphologically distinct but chemotaxonomically and phylogenetically related species, namely *Corynebacterium paurometabolum* Steinhaus 1941[AL] and *Rhodococcus aurantiacus* (Tsukamura and Mizuno 1971) Tsukamura and Yano 1985. The proposal that these taxa be transferred to *Tsukamurella* gen. nov. as *Tsukamurella paurometabola* was primarily based on 16S rRNA sequence similarity (over 99%) and supporting chemotaxonomic data. However, it has already been outlined in Chapter II that it is not unusual for organisms which belong to different genomic species to have almost identical 16S rRNA sequences (Fox *et al.*, 1992; Stackebrandt & Goebel, 1994; see Figure 2-1). Within this context it was interesting that in the present study *Tsukamurella paurometabola* strains M334 and N663 had an almost identical primary structure (99.39% similarity; nine nucleotide differences out of 1474 positions) but show only a moderate amount of DNA relatedness (44%; Goodfellow *et al.*, 1991).

Auerbach *et al.* (1992) compared tsukamurellas from hospital outbreaks with marker strains of *Tsukamurella paurometabola.* They found that *Tsukamurella paurometabola* strain JC7^T, the original type strain of *Corynebacterium paurometabolum,* and *Tsukamurella paurometabola* strain N663, the original type strain of *Rhodococcus aurantiacus*, showed different phenotypic and ribotype profiles. They concluded that their outbreak isolates were similar to *Tsukamurella paurometabola* strain N663 but not *Tsukamurella paurometabola* strain JC7^T! The conclusions drawn by Auerbach *et al.* (1992) are very interesting for in the present study *Tsukamurella paurometabola* strain JC7^T (*neé Corynebacterium paurometabolum*) and *Tsukamurella paurometabola* strains M333, M337, M343 and N663 (*neé Rhodococcus aurantiacus*) were readily distinguished by a number of phenotypic properties.

It is also evident from the present study that the representatives of *Tsukamurella paurometabola*, namely strains JC7^T, M333, M334, M337, M343 and N663, showed considerable phenotypic heterogeneity. Thus, the *Tsukamurella paurometabola* strains were recovered in one multimembered group (cluster 2) and two single membered clusters, namely *Tsukamurella paurometabola* strains JC7^T and M334. The recovery of *Tsukamurella paurometabola* strains M333, M337, M343 and N663 in a distinct cluster is in line with an earlier numerical taxonomic study (Goodfellow *et al.*, 1991). However, in the present study *Tsukamurella paurometabola* strain M334 was sharply distinguished from the *Tsukamurella paurometabola* strains assigned to cluster 2. The separation of strain M334 from those assigned to cluster 2 is in good agreement with DNA-DNA pairing data, but not with the results of the previous numerical taxonomic study (Goodfellow *et al.* ,1991).

In the present investigation, the separation of *Tsukamurella paurometabola* strains into three phena was strongly supported by the results of the Curie point pyrolysis mass spectrometric and whole-organism electrophoretic analyses. It seems likely, therefore, that *Tsukamurella paurometabola* Collins *et al.* 1988 is heterogeneous and contains organisms that should be assigned to three species. DNA-DNA pairing studies on representatives of the different kinds of *Tsukamurella paurometabola* strains need to be carried out to determine the detected taxonomic relationships that exist between these organisms.

The results of the present study provide further evidence that members of the genus *Tsukamurella* occur in activated sludge (Lemmer & Kroppenstedt, 1984). The consistent isolation of one organism from the foam samples was in good agreement with the results of Lemmer and Kroppenstedt (1984) who also found that a single predominant actinomycete predominated in each of fourteen activated

sludge sewage treatment plants. Tsukamurellas have also been isolated from soil and have been implicated as human pathogens (Shapiro *et al.*, 1991; Lai, 1993; Jones *et al.*, 1994; McNeil & Brown, 1994).

The present study provides further evidence of the value of the polyphasic taxonomic approach in bacterial systematics. The sludge isolates and the representatives of *Tsukamurella paurometabola* were assigned into four taxa using the polyphasic taxonomic data. The four phenons can be tentatively equated with distinct species though DNA-DNA pairing studies are required to confirm this given the low resolution of the 16S rDNA sequence data. However, the 16S rRNA sequence data can be used to generate species-specific probes for *in situ* recognition of tsukamurellas in the environment, notably in activated sludge sewage plants. Several rapid chemical fingerprinting techniques are available for the characterisation of large numbers of organisms from natural and manmade habitats, notably Curie point pyrolysis mass spectrometry (Magee, 1993, 1994; Goodfellow, 1994c), quantitative fatty acid analysis (Embley & Wait, 1994) and whole-organism protein electrophoresis (Pot *et al.*, 1994). The speed and reproducibility of Curie point pyrolysis mass spectrometry and its applicability to a wide range of bacteria make it a particular attractive method for evaluating the integrity of taxa circumscribed in numerical taxonomic studies (Sanglier *et al.*, 1992; Goodfellow, 1994c). In addition, pyrolysis mass spectrometry coupled with artificial neural network analysis should provide an excellent strategy for identifying and monitoring microbial fluxes in the activated sludge plants.

<BLANK>

CHAPTER IV.
IDENTIFICATION OF STREPTOMYCETES USING CURIE POINT PYROLYSIS MASS SPECTROMETRY AND ARTIFICIAL NEURAL NETWORKS

A. Introduction

1. Pyrolysis mass spectrometry

1.1. Introduction

The need to classify, identify and type microorganisms is an ever present theme in microbiology, notably in clinical and industrial microbiology. Identification is critical for distinguishing between potential pathogens or spoilage organisms and between commensals or contaminants. The choice of microorganisms for industrial screening programmes, especially those with low throughputs, is primarily a problem of distinguishing between known organisms and recognising new ones. Further, reliable techniques are needed for inter-strain comparisons for epidemiological tracing and in eliminating sources of microbial contamination. Conventional techniques used for such purposes are often taxon specific, require a varied assortment of methods, media and reagents and are generally slow given the time constraints within which decisions have to be made. This is not the case with analytical chemical technique, notably Curie point pyrolysis mass spectrometry (Magee, 1993, 1994; Goodfellow, 1995c).

Pyrolysis is a chemical process that involves the breakdown of complex organic material, such as whole organisms or cell fractions, in an inert atmosphere or vacuum to produce a series of volatile, lower molecular weight molecules, the **'pyrolysate'** (Irwin, 1982), using heat alone. The breakdown of test material is

reproducible under controlled conditions and the resultant fragments are characteristic of the original material. The fragments are ionized and separated by mass spectrometry on the basis of their mass-to-charge ratio (*m/z*) to give a pyrolysis mass spectrum which can be taken as a **'chemical fingerprint'** of the original material. The resultant data are complex and need to be analysed using suitable statistical routines (Gutteridge, 1987; Magee, 1993, 1994).

One of the major advantages of pyrolysis mass spectrometry (PyMS) over comparable taxonomic methods, such as conventional chemotaxonomic procedures (Suzuki *et al.*, 1993; Embley & Wait, 1994; Pot *et al.*, 1994) and nucleic acid probing (Stahl & Amman, 1991; Schleifer *et al.*, 1993; Amman *et al.*, 1995), is that it is rapid with respect to single and multiple samples. Typical sample times are less than 2 minutes and up to 300 samples can be analysed per batch. Pyrolysis techniques, notably Curie-point pyrolysis mass spectrometry, are currently being introduced to diagnostic and industrial screening laboratories (Sanglier *et al.*, 1992; Sisson *et al.*, 1992a).

The steps involved in Curie-point pyrolysis mass spectrometric analyses have been considered in detail by Magee (1993, 1994). The procedure can be divided into four stages: sample preparation, smearing a small sample (*ca.* 10-100 µg dry weight) onto a metal carrier; pyrolysis, in which the carrier and sample are heated rapidly under vacuum to a fixed temperature in the range of 358 to 1000 °C resulting in the thermal degradation of the sample to a mixture of volatile compounds; mass spectrometry, in which the volatile components are separated by molecular weight to charge ratios and quantified; and computation, in which the spectra are compared using multivariate statistical methods. A schematic representation of an automatic Curie-point pyrolysis mass spectrometer is shown in Figure 4-1.

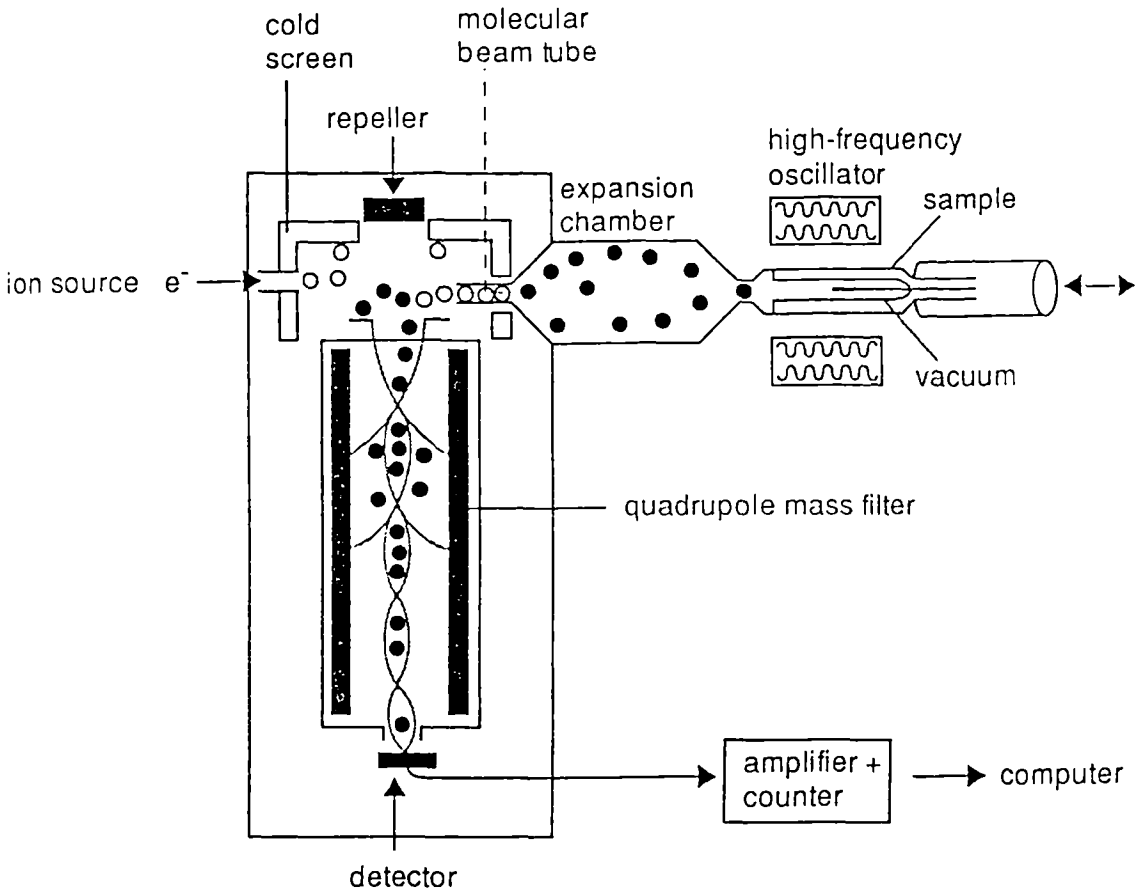


FIGURE 4-1. SCHEMATIC REPRESENTATION OF AN AUTOMATED CURIE POINT PYROLYSIS MASS SPECTROMETER.

1.2. Pyrolysis mass spectrometry

Pyrolysis was first applied to mass spectrometry by Zemany (1952) who showed that complex biological materials degrade in a reproducible way given carefully standardised pyrolysis conditions. Interest soon waned in the technique as workers turned to a less expensive system, namely pyrolysis gas chromatography (PyGC). This technique involves the controlled thermal degradation of organic materials followed by analysis of pyrolysates using gas chromatography. The resultant '**pyrogram**' consists of a series of peaks which may be used for the identification of the original specimen. The chemical nature of each peak may be further determined by mass spectrometry (Simmonds, 1970) but, more usually, the profile is taken as a '**fingerprint**' characteristic of the test material.

The first application of PyGC was by Davison *et al.* (1954) who obtained characteristic patterns when they pyrolysed chemical polymers at 650 °C. The driving force for the application of pyrolysis to microbiological problems came from the United States space exploration programme when a miniaturised automated pyrolysis system was designed to test for the presence of extraterrestrial life (Wilson *et al.*, 1962).

It was not until the work of Reiner and his colleagues (Reiner, 1965; Reiner & Ewing, 1968; Reiner *et al.*, 1973) that the potential value of PyGC in microbiology was appreciated. The technique was subsequently applied to the classification and identification of several microorganisms including aerobic, endospore-forming bacilli (O'Donnell *et al.*, 1980), dermatophytes (Carmichael *et al.*, 1973), pseudomonads (French *et al.*, 1980), staphylococci (Magee *et al.*, 1983) and streptococci (Stack *et al.*, 1981; French *et al.*, 1989). It is also proved possible to distinguish between serotypes of *Escherichia coli* (Reiner & Ewing, 1968), between different subgroups

of *Vibrio cholerae* (Haddadin *et al.*, 1973) and between streptococcal mutants which differed only in polysaccharide capsular antigens (Huis In't Veid *et al.*, 1973).

Despite the attractive features of PyGC outlined above, the technique proved to be of limited value in microbiology given problems with long-term reproducibility, low throughput and inadequate data handling routines (Gutteridge, 1987). Improvements have been made to the technique, especially in the longevity and performance of chromatographic columns (Eudy *et al.*, 1985) and in data handling routines (Magee *et al.*, 1983), but PyMS proved to be superior to PyGC in terms of speed, reproducibility and ease of automation.

The first dedicated PyMS system was designed and built by Meuzelaar and Kistemaker (1973) at the FOM Institute in Amsterdam. Further technical development led to the construction of the first fully automated instrument, the Autopyms, which involved the use of high-speed ion counting and computerised data processing technology (Meuzelaar *et al.*, 1976). This system served as a model for the production of two commercial instruments, the Extranuclear 5000 (Extranuclear Laboratories, Pittsburgh, USA) and the Pyromass 8-80 (VG Gas Analysis Ltd., Middlewich, Cheshire, UK). However, these commercial instruments were expensive, costing upwards of £100,000 (Gutteridge, 1987). The Pyromass 8-80 has been described in detail (Gutteridge *et al.*, 1984; Shute *et al.*, 1985).

At this stage of development PyMS was seen to offer the prospect of processing large numbers of samples rapidly for microbiological purposes. These early systems were used to classify and identify several groups of clinically significant bacteria, notably mycobacteria (Meuzelaar *et al.*, 1976; Wieten *et al.*, 1979, 1981, 1982, 1983). Other early taxonomic studies included those on *Bacillus* (Shute *et al.*, 1984), *Escherichia* (Haverkamp *et al.*, 1980), *Legionella* (Kajiooka & Tang, 1984) and *Neisseria* (Borst *et al.*, 1978). The technique was also successfully

used in fungal systematics (Weijman, 1977), in quality control of vaccines (Haverkamp *et al.*, 1980; Windig *et al.*, 1980) and in the analysis of bacterial wall polymers (Haverkamp *et al.*, 1980; Boon *et al.*, 1981).

These early investigations showed the potential of PyMS though a number of problems prevented the widespread use of the technique as a routine tool in microbiology (Gutteridge, 1987). Manual loading of samples coupled with a prolonged processing time meant that the technique was labour-intensive with a low sample throughput per day. Attempts to compare new spectra with those held in data libraries were frustrated by inherent machine instability which led to poor reproducibility over time. Additional problems included the expensive, cumbersome hardware and lack of suitable computer software to fully analyse the complex pyrolysis data.

The breakthrough came with the introduction of the Horizon PYMS-200X (Horizon Instruments, Heathfield, West Sussex, England, UK), an instrument based on the PyMS quadrupole mass spectrometric system of Prutec Ltd. (Aries *et al.*, 1986). The superior performance of this machine over earlier models can be attributed to an improved electron multiplier, which allowed faster analysis times (2 minutes per sample), enhanced reproducibility and fully integrated software implemented on a personal computer (Gutteridge, 1987; Goodfellow, 1995c). However, the development of a reliable automated inlet system which allowed high-volume throughput was the major achievement of the PYMS-200X (Gutteridge, 1987).

The PYMS-200X instrument has been extensively used to characterise and classify a variety of microorganisms. These include *Actinomadura* (Trujillo, 1995), *Aeromonas* (Magee *et al.*, 1993), *Bacteroides* (Duerden *et al.*, 1989), *Clostridium*

(Wilkinson *et al.*, 1995), *Corynebacterium* (Hindmarch *et al.*, 1990), *Fusobacterium* (Magee *et al.*, 1989a), *Mycobacterium* (Hamid, 1993), *Propionibacterium* (Goodacre *et al.*, 1994b), *Rhizobium* (Goodacre *et al.*, 1991), *Streptococcus* (Winstanley *et al.*, 1992) and *Streptomyces* (Sanglier *et al.*, 1992). Most of these applications were designed to evaluate the integrity of taxa circumscribed using conventional taxonomic criteria. In general, good agreement has been found between the pyroclassifications and corresponding conventional classifications. Strategies to confirm the homogeneity of taxa highlighted in such comparative studies have been discussed by Magee (1993). It has also been shown that PyMS and DNA-DNA pairing procedures give similar profiles of relatedness (Goodacre *et al.*, 1991; Sanglier *et al.*, 1992).

To date, the most important application of Curie point pyrolysis mass spectrometry has been in microbial epidemiology (Magee, 1993; Goodfellow, 1995c). Pyrolysis mass spectrometry is not a typing method *per se* as a permanent type designation is not assigned to test strains but it has proved to be a quick and effective method for inter-strain comparisons of bacteria that commonly cause outbreaks of disease. This conclusion is based on studies of clinically significant bacteria, recent examples include *Bacteroides ureolyticus* (Duerden *et al.*, 1989), *Campylobacter jejuni* (Orr *et al.*, 1995), *Corynebacterium jeikeium* (Hindmarch *et al.*, 1990), *Legionella pneumoniae* (Sisson *et al.*, 1991b), *Listeria monocytogenes* (Freeman *et al.*, 1991a; Low *et al.*, 1992), *Pseudomonas aeruginosa* (Sisson *et al.*, 1991c), *Pseudomonas cepacia* (Corkhill *et al.*, 1994), *Salmonella enteritidis* (Freeman *et al.*, 1990), *Streptococcus pneumoniae* (Freeman *et al.*, 1991b), *Staphylococcus aureus* (Gould *et al.*, 1991), *Staphylococcus pyogenes* (Magee *et al.*, 1989b, 1991; Freeman *et al.*, 1990) and *Xanthomonas maltophilia* (Orr *et al.*, 1991).

It is evident that PyMS can be used to discriminate between strains as accurately as routine typing systems (Freeman *et al.*, 1991a; Sisson *et al.*, 1991c). Indeed, in some cases it has been used to separate isolates beyond the resolution of such systems (Freeman *et al.*, 1991b; Gould *et al.*, 1991). The results of PyMS analyses have also been shown to correspond to those from molecular based techniques, including restriction length fragment polymorphism (Sisson *et al.*, 1991c; Low *et al.*, 1992) and random amplification of polymorphic DNA analyses (Kay *et al.*, 1994). Similarly, PyMS studies on isolates from outbreaks of infection due to untypable isolates have been shown to agree with epidemiological data that have subsequently become available (Orr *et al.*, 1991; Cartmill *et al.*, 1992).

The discriminating capacity of PyMS has also been used to distinguish between very closely related strains of bacteria, notably to detect small genotypic changes in *Escherichia coli* (Goodacre & Berkeley, 1990). It has also proved possible to distinguish between representatives of staphylococcal species by PyMS of extracted DNA (Mathers *et al.*, 1995). These results challenge previous assumptions that PyMS is restricted to detecting phenotypic differences though the basis for the differentiation of the DNA extracts has still to be determined.

Pyrolysis mass spectrometry has also been used to classify and identify industrially significant actinomycetes (Saddler *et al.*, 1988; Sanglier *et al.*, 1992). In this latter study, members of representative actinomycete genera were pyrolysed in order to determine the effects of medium design, incubation time and sample preparation on experimental data; it was concluded that reproducible results could be obtained given rigorous standardisation of growth and pyrolysis conditions. Sanglier and his colleagues also showed that PyMS data could be used to objectively select strains for pharmacological screens, as unknown or putatively novel actinomycetes appeared as outliers on ordination diagrams. They were also

able to distinguish between actinomycetes at and below the species level. In particular, representatives of three closely related *Streptomyces* species, namely *Streptomyces albidoflavus, Streptomyces anulatus* and *Streptomyces halstedii*, were distinguished. The separation of these numerically circumscribed streptomycete species indicated that PyMS can provide a rapid way of establishing the taxonomic integrity of established or putatively novel *actinomycete species.*

1.3. Reproducibility of pyrolysis mass spectrometric data

Short-term reproducibility. The *'destructive'* nature of pyrolysis and the undefinable multiple sources of masses led many workers to the view that the technique could not be expected to give highly reproducible results. However, it has been shown that pyrolysis has little effect on reproducibility provided temperature rise and total heating times are strictly controlled and pyrolysates are rapidly removed from the pyrolysis zone to prevent secondary reactions (Schulten & Lattimer, 1984).

Windig *et al.* (1979) showed that reproducible spectra were obtained using pyrolysis temperatures between 510°C and 610°C; the total heating time of samples was not found to be a crucial parameter provided that the pyrolysate was driven off rapidly nor was the temperature rise time critical as long as it was fast enough to avoid secondary reactions. In contrast, sample size was shown to influence the short-term reproducibility of pyrolysis mass spectra. When samples were too large secondary reactions occurred and when they were too small spectra were susceptible to background noise that did not originated from the sample (Windig *et al.*, 1979).

A number of investigators studied the role of the interface between the pyrolyser and mass spectrometer in the short-term reproducibility of PyMS (Windig *et al.*, 1979; Schulten & Lattimer, 1984). It was considered that the expansion

chamber should allow efficient pyrolysate transfer between the pyrolyser and the mass spectrometer so as to prevent loss of information and irreproducible catalytic reactions. These requirements were met by gold-plating the inner surfaces of expansion chambers and heating to reduce condensation. Windig *et al.* (1979) considered that the optimal temperature for expansion chambers was 150°C for this temperature allowed a reduction in condensation and facilitated reproducible electron impact fragmentation.

The short-term reproducibility of PyMS has been examined by several investigators working with microorganisms. Eshuis *et al.* (1977) showed that the intensities of some masses in the replicate spectra of a Listeria strain differed by over 10% when the automated system developed by Meuzelaar *et al.* (1976) was used. Shute *et al.* (1984) found a mean dissimilarity of 4% between replicated *Bacillus* samples over an eight week period using the Pyromass 8-80 system though they also noted that some of the mass spectra showed far less reproducibility than this mean value. However, such levels of variability were not considered to be an obstacle in microbiological work, notably for the identification and typing of microorganisms (Gutteridge, 1987).

Long-term reproducibility. Relatively few studies have been carried out to determine the long-term reproducibility of PyMS. Wieten *et al.* (1981) generated a database for the identification of members of the '*Mycobacterium tuberculosis* complex' and were able to assign fresh isolates to the complex over a one year period. However, not a lot of credence can be given to this study as only seven masses were compared.

Shute *et al.* (1988) examined the long-term reproducibility of the Pyromass 8-80 instrument over a fourteen month period using a set of aerobic, endospore-forming bacilli and found that within each dataset the test strains were assigned to

the correct groups though the characteristic masses involved differed in the corresponding studies. However, when two datasets derived from runs at the begining and end of a fourteen month interval were combined and analysed by using multivariate statistics not only were the mass spectra not directly comparable but the two groups obtained corresponded to the time of the analysis. Shute and her colleagues concluded that the overall spectral patterns had changed over time even though the relationships found between samples pyrolysed at the same time remained. The lack of reproducibility over time was attributed not to variations in pyrolysis conditions but to **'mass spectrometer drift'**, that is, to changes in ion transmissivity due to contamination and ageing of the ion source of the mass spectrometer (Windig *et al.*, 1979; Meuzelaar *et al.*, 1982).

The view that long-term comparisons of pyrolysis mass spectra derived from microorganisms are impossible has been widely accepted (Gutteridge, 1987; Berkeley *et al.*, 1990). Consequently, the use of PyMS in microbial identification and inter-strain comparisons is largely restricted to the use of a procedure known as **'operational fingerprinting'** (Meuzelaar *et al.*, 1982) whereby unknown strains are pyrolysed with reference strains in the same batch experiment. However, operational fingerprinting should be seen as a temporary strategy not as an end in itself.

Phenotypic and genotypic variation. Pyrolysis mass spectrometry has also been used to detect phenotypic variation. Boon *et al.* (1981) showed that PyMS could be used to detect changes in cell wall components attributed to differences in the composition of growth media. Shute *et al.* (1984) found that sporulation made it difficult to discriminate between members of some *Bacillus* species, namely *Bacillus amyloliquefasciens*, *Bacillus subtilis* and *Bacillus pumilus*. Discrimination between representatives of selected *Bacillus* species was not affected by batch to batch

variations in growth media though different media formulations caused marked changes in pyrolysis mass spectra (Shute *et al.*, 1988). Voorhees *et al.* (1988) examined *Bacillus subtilis, Escherichia coli* and *Proteus vulgaris* strains to determine the effect of culture age on pyrolysis mass spectra and found that samples were correctly discriminated despite differences in culture age.

Colonial variation shown by a single strain can also effect the discriminatory ability of PyMS. Freeman *et al.* (1994b) showed that phenotypic variations found between three sets of pyrolysis mass spectra derived from a single strain of *Staphylococcus aureus* were so different that they hampered inter-strain discrimination or strain definition on multivariate statistical analysis. However, variations based on the different subcultures did not affect species or subspecific differentiation.

In practice, phenotypic variation between closely related strains can be minimised by standardising media and growth conditions. It is not always possible to meet these requirements, notably in ecological studies involving diverse organisms which grow on the different media and require different growth conditions. Suitable data handling methods need to be developed to overcome these limitations.

The studies outlined above have been designed to study the effects of genotypic changes on PyMS data. Goodacre and Berkeley (1990) demonstrated that *Escherichia coli* strains which differed only in the presence or absence of a single plasmid could be distinguished though it was not clear whether this differentiation was due to genotypic or phenotypic variation. More work needs to be carried out to determine the effect of genotypic variation on pyrolysis mass spectra.

2. Artificial neural networks

2.1. Introduction

Artificial neural networks (ANNs), which are also referred to as neural networks, neurocomputers, adaptive systems and parallel distributed processors, are widely used to determine complex, non-linear relationships in multivariate data (Simpson, 1990; Hertz *et al.*, 1991; Freeman & Skapura, 1991). The primary aim in developing ANNs is to explore and reproduce human information processing tasks, such as knowledge processing, motor-control, speech and vision (Simpson, 1990). In addition, ANNs are used for data compression, near-optimal solution to combinatorial optimisation problems, pattern recognition, system modelling and function approximation (Pao, 1989; Simpson, 1990; Freeman & Skapura, 1991).

Artificial neural networks have a multiple-processor architecture which has been described as '**parallel distributed processing**', and which is highly interconnected. Artificial neural networks, like the human brain, can learn and memorise. These brain-like features of ANNs allow this computing technique to outperform conventional computation methods based on single-processors. A general, albeit rigorous, definition of an ANN was given by Hecht-Nielsen (1988):

> *"A neural network is a parallel, distributed information processing structure consisting of processing elements (which can process a local memory and carry out localized information processing operations) interconnected together with unidirectional signal channels called connections. Each processing element has a single output connection which branches ("fan out") into as many collateral connections as desired (each carrying the same signal - the processing elements output signal). The processing element output signal can be of any mathematical type desired. All of the processing that goes on within each processing element must be completely local; that is, it must depend only upon the current values of the input signal arriving at the processing element* via *impinging connections and upon values stored in the processing element's local memory."*

Artificial neural networks can be divided into two types on the basis of the learning mechanism, that is, into supervised and unsupervised models. These models learn from examples but exhibit different concepts in learning. In supervised learning the connections (weights) between neurons are adjusted according to the outputs required by the user whereas in unsupervised learning ANNs are exposed only to input stimuli and are allowed to organised by themselves. The supervised learning model is considered here since all of the ANNs developed in the present study were trained by using a supervised learning algorithm, namely the backpropagation algorithm (Rumelhart *et al.*, 1986).

2.2. History of artificial neural networks

The first neural network model, which was formulated by McCulloch and Pitts (1943), featured digital neurons but did not have any ability to learn. Donald Hebb, a psychologist, introduced the concept of '**Hebbian learning**' whereby changes in synaptic strengths (connections between neurons) were proportional to the activation of the neurons (Hebb, 1949). This concept formed the basis for the creation of neural networks since it described a rule for updating synaptic strengths in two-layered networks thereby enabling the networks to learn.

The learning concept was incorporated into a two-layered network, the **'perceptron',** by Rosenblatt (1957). He not only formulated a learning rule based on weights adjusted in proportion to error between output neurons and target outputs, but also proposed a three-layered perceptron into which he attempted to incorporate a learning mechanism. However, Rosenblatt was unable to derive a sound way of adjusting the weights between input and hidden layer neurons in the three layered network but, even so, many problems were solved by using two-layered perceptrons (Simpson, 1990).

The lack of a mathematically rigorous procedure to allow learning in multi-layered networks was a major stumbling block in the development and application of neural networks. Minsky and Papert (1969) pointed out the limitations of perceptrons but acknowledged the possibility that multi-layered networks could overcome these limitations. However, they were unable to show that three-layered networks could learn and concluded that multi-layered networks were not a viable proposition.

Interest in multi-layered ANNs was rekindled by the invention of the **'backpropagation'** algorithm which allows training of multi-layered networks. Several investigators working in wide-ranging disciplines developed this algorithm. The first gradient approach to training multi-layered ANNs was introduced by Amari (1967) who developed a system to perform nonlinear classification. However, Amari's approach, while on the right lines, did not provide a complete description of how multi-layered mapping could be developed. Werbos (1974) independently discovered the backpropagation algorithm, which he called **'dynamic feedback'**, when working on his doctoral thesis in statistics (Werbos, 1974). This algorithm was rediscovered by Paker (1982), a student at Stanford University, who called his algorithm **'learning logic'**. Nevertheless, it was not until Rumelhart *et al.* (1986) introduced the backpropagation algorithm in their work on stimulating cognitive processes that the use of ANNs was introduced to various industrial and scientific fields (Simpson, 1990; Crick, 1989; Casey Klimasauskas, 1992).

2.3. Structure of artificial neural networks

A comparison can be made between the structures of biological neurons and processing elements (Figure 4-2). Neurons are the fundamental cellular elements of the nervous system and human brain. The average human brain has about 10^{11} interconnected neurons (Boddy *et al.*,1990). A single neuron can receive inputs

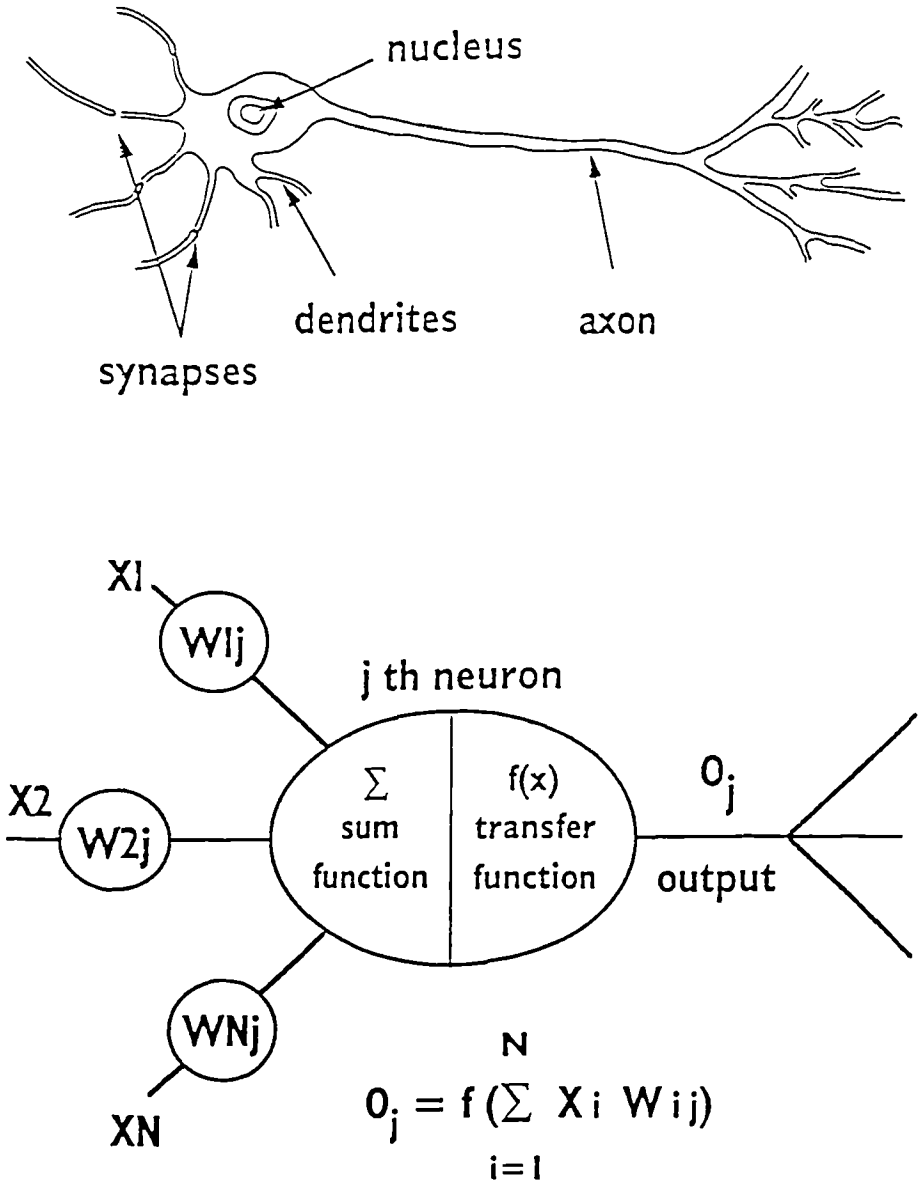
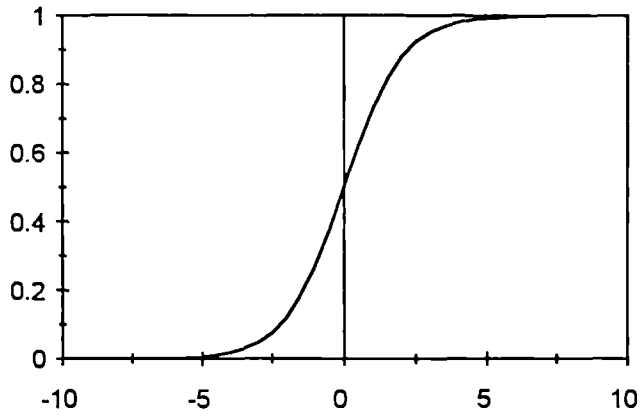


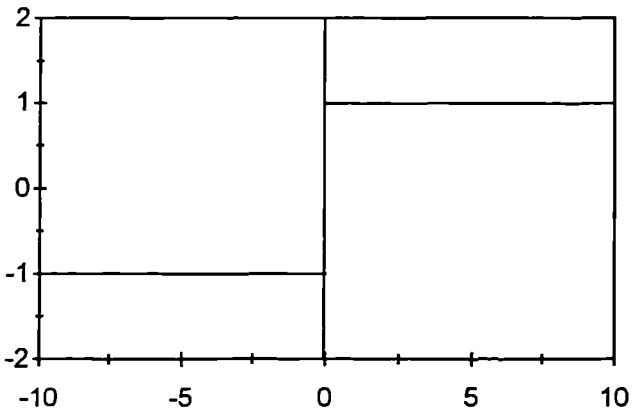
FIGURE 4-2. STRUCTURAL COMPARISON OF THE BIOLOGICAL NEURON AND THE PROCESSING ELEMENT (PE). SUMMATION OF THE WEIGHTED INPUTS OCCURS WITHIN THE PROCESSING ELEMENT, $s_j = \sum W_{(j,i)} x_i$. A TRANSFER FUNCTION IS APPLIED TO THE SUMMATION, $y_i = f(s_i)$. MODIFIED AFTER BODDY *et al.* (1990).

from 5000 to 10000 adjacent neurons through dendrites and can send output signals to adjacent neurons *via* axons. The signals are in the form of short pulses of electrical activity that occur at about 1000 *Hz*. The junctions between axons and dendrites are called synapses; signals are transferred across synapses as chemicals, that is, by neurotransmitters. The strength of the signal transferred depends on the amount of these chemicals. This synaptic function provides the basic memory mechanism of the brain but also provides the basic concept for ANNs.

The building blocks of ANNs, **'processing elements (PE)'**, are analogous to neurons. A PE receives many input signals from other PEs. Each element has two mathematical functions, that is, the sum and transfer functions. The signals accepted from other PEs are summed and passed on to other PEs *via* transfer functions. The role of the transfer function is to transform the summed signal to values between zero and one. A variety of transfer functions are available, but the most popular ones are the sigmoid and step functions (Figure 4-3). Each connection between two adjacent PEs is given a weight that is equivalent to the synapse of the neuron. The magnitude of the weight is analogous to the synaptic strength or to the amount of neurotransmitters. The typical topology of an ANN is shown in Figure 4-4. Network learning can be improved by adding bias neurons (PEs) which receive one as an input value.



Sigmoid function $f(x) = \dfrac{1}{(1+e^{-x})}$



Step function $f(x) = +1(x \geq 0)$
$f(x) = -1(x < 0)$

FIGURE 4-3. THE SIGMOID AND STEP TRANSFER FUNCTIONS.

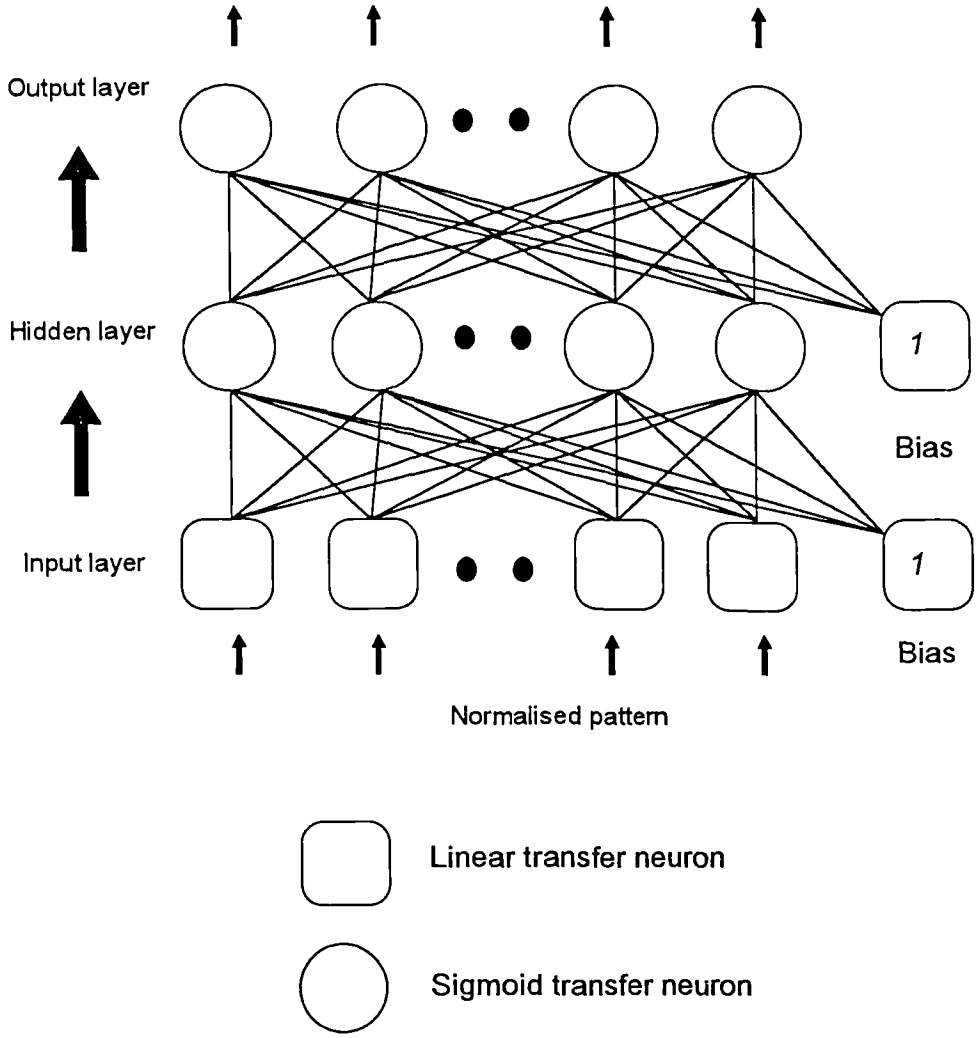


FIGURE 4-4. TYPICAL STRUCTURE OF A MULTI-LAYERED ARTIFICIAL NEURAL NETWORK WHICH CONSISTS OF THREE LAYERS, NAMELY INPUT, HIDDEN AND OUTPUT LAYERS.

2.4. Application of artificial neural networks to biology and chemistry

Artificial neural networks have been widely used in disciplines that require pattern recognition (Table 4-1). Spectral patterns generated from the application of techniques such as infrared spectroscopy, mass spectrometry, laser-induced fluorescent spectrometry, nuclear magnetic resonance spectroscopy, pyrolysis mass spectrometry, Raman spectroscopy and ultraviolet spectroscopy have been identified by using ANNs. Artificial neural networks are especially useful for distinguishing between minute differences within complex patterns. Sellers *et al.* (1990) were able to differentiate between the isomers of partially methylated alditol acetates by using mass spectrometry; this process is difficult to achieve using conventional pattern recognition methods.

Goodacre and his colleagues used ANNs to derive qualitative and quantitative information from pyrolysis mass spectra of chemical and biological mixtures using an Horizon Instruments PyMS-200X mass spectrometer (Goodacre *et al.*, 1992, 1993a, 1993b, 1994c; Goodacre & Kell, 1993). The first application of ANNs to PyMS data was designed to differentiate between adulterated and non-adulterated virgin olive oil samples (Goodacre *et al.*, 1992; 1993b). The system basically gave crude quantitative results.

Artificial neural networks were subsequently used to quantify the content of casamino acid in a mixture with glycogen (Goodacre, 1993a) and to determine the amount of indole produced by an *Escherichia coli* strain (Goodacre & Kell, 1993). The same strategy was used to quantify the amount of ampicillin in fermentation broths of *Escherichia coli* and *Staphylococcus aureus* strains (Goodacre *et al.*, 1994c) and to determine the amount of a single component in artificial mixtures which include glycogen and lysozyme, glycogen and nucleic acid, and a mixture

TABLE 4-1. SOME APPLICATIONS OF ARTIFICIAL NEURAL NETWORKS IN BIOLOGY AND CHEMISTRY.

Applications	References
Chemical process engineering and fermentation	Cleran *et al.* (1991) Willis *et al.* (1991) Collins (1993) Glassey *et al.* (1994a,b) Morris *et al.* (1994)
Clinical diagnosis	Jervis *et al.* (1994) Siebler *et al.* (1994) Wilding *et al.* (1994)
Ecological modelling	Colasanti (1991) Fu & Poch (1995)
Infrared spectroscopy	Rob & Munk (1990) Fessenden & Gyorgyi (1991) Munk *et al.* (1991) Tanabe *et al.* (1992) Klawun & Wilkins (1995)
Laser induced fluoroscent spectrometry	Andrews & Lieberman (1994)
Mass spectrometry	Sellers *et al.* (1990) Werther *et al.* (1994)
Nuclear magnetic resonance spectroscopy	Meyer *et al.* (1991) Kjaer & Poulsen (1991) Anker & Jurs (1992) Radomski *et al.* (1994)
Nucleic acid analysis	Delmeler & Zhou (1991) Snyder & Stormo (1993) Taylor *et al.* (1994)
Protein sequence analysis	Qian & Sejnowski (1988) Vieth & Kolinski (1991) Cohen *et al.* (1993) Wu (1993) Schneider & Wrede (1994)
Pyrolysis mass spectrometry	Goodacre *et al.* (1992) Goodacre *et al.* (1993a,b) Goodacre *et al.* (1994a,c)
Raman spectroscopy	Lewis *et al.* (1994) Schulze *et al.* (1994)
Ultraviolet spectroscopy	Mittermayr *et al.* (1994)

containing biomass from *Bacillus subtilis*, *Escherichia coli* and *Staphylococcus aureus* strains (Goodacre *et al.*, 1994a).

Artificial neural networks have also been used to recognise sequence patterns within macromolecules such as DNA, RNA and proteins (Qian & Sejnowski, 1988; Delmeler & Zhou, 1991; Vieth & Kolinski, 1991; Snyder & Stormo, 1993). Complex biological interactions that occur in fermentors and bioreactors can be monitored and the output (normally biomass) predicted by using ANNs (Willis *et al.*, 1991;Glassey *et al.*, 1994a,b). The advantages and prospects of the use of ANNs in biotechnological fields has been reviewed by Montague and Morris (1994).

2.5. Artificial neural networks and microbial systematics

Microbial systematics is markedly data dependent. It is, therefore, not surprising that many of the more recent taxonomic advances have resulted from the ways in which data are collected and analysed. In bacterial systematics there is a continuing trend towards automation with an ever increasing reliance placed on analytical instrumentation, notably on the use of gas chromatography, high-performance liquid chromatography and mass spectrometry. A common feature of such instruments is that they generate data rapidly, often involving multiple measurements on individual samples. Artificial neural networks are recognised as an appropriate way of handling such data (Simpson, 1990).

Artificial neural networks have been successfully used for the identification of algae, bacteria and fungi (Table 4-2). The distinctive patterns produced by the application of various chemical, molecular and physical fingerprinting methods to microorganisms can be detected by using ANNs following suitable *a priori* digitisation and normalisation steps. Artificial neural networks are especially useful when dealing with fairly **'noisy'** or **'incomplete'** data.

TABLE 4-2. APPLICATION OF ARTIFICIAL NEURAL NETWORKS IN MICROBIAL SYSTEMATICS.

Source of fingerprints	Target organisms	References
Microscopic image analysis	Phytoplankton	Culverhouse *et al.* (1994)
Flow cytometry	Algae	Balfoort *et al.* (1992) Smits *et al.* (1992)
	Basidiomycetes	Morris *et al.* (1992)
	Phytoplankton	Wilkins *et al.* (1994)
Phenotypic or phenetic data	Gram-negative bacteria	Rataj & Schindler (1991)
	Haemophilus	Kennedy & Thakur (1993)
	Enterobacteriaceae and *Vibrionaceae*	Schindler *et al.* (1994)
Pyrolysis mass spectrometry	*Mycobacterium tuberculosis*	Freeman *et al.* (1994a)
	Propionibacterium acnes	Goodacre *et al.* (1994b)
Whole-organism protein electrophoresis	*Aeromonas* and *Mycobacterium*	Millership (1993)

Schindler *et al.* (1994) used ANNs for the identification of Gram-negative bacteria belonging to the families *Enterobacteriaceae* and *Vibrionaceae*. In this study, which was an extended version of previous work (Rataj & Schindler, 1991), over 95% of the test strains were correctly identified to species and nearly 98% to genera. Polyacrylamide gel electrophoresis of cellular proteins, a well established technique for classification and identification of bacteria (Pot *et al.*, 1994), has also been coupled with ANNs. Millership (1993) showed that members of species belonging to the genera *Aeromonas* and *Mycobacterium* can be distinguished by ANN analysis of Fourier-transformed electrophoretic data. The application of ANNs to flow cytometric data is interesting as this approach opens up the prospect of achieving a large throughput system of organisms for identification. This approach should be suitable for ecological studies that involve many diverse microorganisms (Balfoort *et al.*, 1992; Morris *et al.*, 1992; Smits *et al.*, 1992; Wilkins *et al.*, 1994).

Artificial neural networks have also been used to identify pyrolysis mass spectra derived from clinically significant bacteria. Freeman *et al.* (1994a) were able to differentiate between *Mycobacterium bovis* and *Mycobacterium tuberculosis* isolates by training an ANN with mean spectra derived from triplicated samples. A combination of PyMS and ANNs was used to distinguish between *Propionibacterium acnes* strains isolated from dogs (Goodacre *et al.*, 1994b). These workers successfully identified isolates to species both by using an ANN constructed with the backpropagation algorithm, that is, by a supervised learning process, and by using a conventional statistical procedure.

The primary aims of the present study were to generate ANNs for the analysis of pyrolysis mass spectral data to see whether this approach could be used to overcome problems that currently impede multivariate statistical analyses of PyMS data and to evaluate whether the system could be applied to long-term

identification of streptomycetes. The test strains were representatives of three putatively novel *Streptomyces* species, provisionally labelled *Streptomyces* species A, B and C (Atalan, 1993; Manfio, 1995). Members of these taxa have repeatedly been isolated, using a standard dilution plate procedure (Goodfellow *et al.*, 1990), from soil taken from the Palace Leas site at Cockle Park Experimental Farm, Northumberland, England, UK (Grid reference NZ200913). *Streptomyces* group A and B strains were isolated from raffinose-histidine agar plates supplemented with cyclohexamide (50 µg/ml) and nystatin (50 µg/ml) and incubated at 25 °C for 14 days (Vickers *et al.*, 1984), and the group C strains from similarly treated starch-casein plates supplemented with the two antifungal antibiotics and novobiocin (25 µg/ml). Representatives of these taxa form homogeneous taxa on the basis of numerical phenetic, whole-organism protein electrophoretic, fatty acid, DNA pairing and 16S rDNA sequencing data (Atalan, 1993; Manfio, 1995).

<BLANK>

B. Materials and Methods

1. Test strains

The test strains consisted of sixteen well studied representatives of *Streptomyces* groups A, B and C and thirteen fresh isolates putatively identified to these taxa solely on the basis of morphological criteria (Table 4-3). The test strains were maintained as glycerol suspensions (20 %, v/v) at -20 °C (Wellington & Williams, 1978). In addition, information from earlier PyMS analyses on representative strains belonging to the genera *Actinomadura, Mycobacterium, Nocardia, Nocardiopsis, Saccharomonospora* and *Streptosporangium* was used (Appendix J).

2. Datasets

2.1. *Streptomyces* (target) strains

The twenty-nine streptomycete strains were analysed on six different occasions over a period of twenty months in order to investigate the long-term reproducibility of the PyMS system. The results from each of the batch experiments were assigned to different datasets (Table 4-3). The first experiment was carried out in duplicate at different times of the day and the datasets designed 1A and 1B.

2.2. Non-*Streptomyces* (non-target) strains

The pyrolysis mass spectra of members of non-target taxa were obtained from previous studies carried out in the Newcastle laboratory (Kim, 1993; Hamid, 1994; Trujillo, 1994; Kim, 1995) and the database maintained by using the FoxPro database language (version 2.0; Fox Software Inc., 134 W. South Bounary, Perrysburg, Ohio 43551, USA).

TABLE 4-3. *STREPTOMYCES* STRAINS AND DATASETS.

Strain	Dataset					
	1A,1B	2	3	4	5	6
No. of days*	0	241	510	518	619	622
• Reference strains						
Streptomyces group A						
A2	√	√	√	√		
A34	√	√	√	√	√	
A35	√	√	√	√		
A40	√	√	√	√	√	√
A41	√	√	√	√	√	
A42	√	√	√	√		
Streptomyces group B						
B41	√					
B42	√	√	√	√	√	
B43	√	√	√	√	√	√
B76	√	√	√	√		
B77	√	√	√	√	√	
Streptomyces group C						
C51	√		√	√		
C54	√	√	√		√	
C55	√	√	√	√		√
C56	√	√	√		√	
C58	√	√	√	√	√	
• Putatively identified isolates						
Streptomyces group A						
A2102					√	
A2108					√	
A46R47					√	
A46R51					√	√
A46R62					√	√
A5843					√	
A62BR2					√	
A81319					√	
Streptomyces group B						
B7316					√	
B7324					√	√
B7342					√	√
Streptomyces group C						
C765						√
C768						√

* Represents the number of days from the analysis of the first dataset.

3. Curie-point pyrolysis mass spectrometry

3.1. Growth conditions and sample preparation

Glycerol stock cultures of the streptomycete and streptosporangia strains were used to inoculate sterile polyvinyl membrane filters (0.45mm, HV type; Millipore Ltd., Watford, England, UK) placed over a non-sporulating medium designed to inhibit sporulation of streptomycetes (Sanglier *et al.*, 1992; Appendix A). The inoculated plates were incubated for 3 days at 25 °C. The same procedure was used for the *Actinomadura* and *Nocardiopsis* strains grown on modified Bennett's agar (Jones, 1949; Appendix A) for 3 days at 30 °C, for the *Nocardia* and *Mycobacterium* strains grown on glucose-yeast extract agar (GYEA; Gordon & Mihm, 1962; Appendix A) for 3 days at 30 °C, and for the *Saccharomonospora* strains on GYEA for 3 days at 45 °C. After incubation, biomass was taken from the whole surface area of the filter and applied, using sterile disposable loops, onto clean ferro-nickel alloy foils that had been inserted into pyrolysis tubes (Horizon Instruments Ltd.). Triplicate samples were prepared for each strain. The assembled tubes plus foils were dried in a hot air oven at 80 °C for 5 minutes to ensure that the biomass adhered to the foils.

3.2. Pyrolysis mass spectrometry

Curie point pyrolysis mass spectrometry was carried out at 530 °C for 2.4 seconds using a Horizon Instrument PyMS-200X mass spectrometer (Horizon Instruments). The spectrometer was left to stabilise for at least an hour before the analyses. The inlet heater was set at 110 °C and the heated tube loader at 120 °C. Curie point pyrolysis was carried out under vacuum with a temperature rise time of 0.6 of a second (Windig *et al.*, 1979). The pyrolysate was ionised by collision with a cross-beam of low energy (20 eV) electrons and the ions separated in a quadrupole

mass spectrometer that scanned pyrolysates at 0.35 second intervals. Integrated ion counts were recorded for each sample at unit mass intervals from 51 to 200 using an IMB-PC compatible computer and software provided by the manufacturer. Masses between 11 and 50 were not included in the dataset given their poor reproducibility (Berkeley *et al.*, 1990); these mass ions tend to be derived from low molecular weight compounds such as methane (*m/z*: 15, 16), ammonia (*m/z*: 16, 17), water (*m/z*: 17, 18), methanol (*m/z*: 31, 32) and hydrogen sulphide (*m/z*: 34).

4. Statistical analyses

The **PYMENU** program (Horizon Instruments) and **GENSTAT** statistical package (Nelder, 1979) were used to carry out the statistical analyses as described by Gutteridge *et al.* (1985) and Goodfellow (1995c). The major steps involved are outlined in Figure 4-5.

The first stage in the analysis of data was normalisation which is also known as pattern scaling. Each mass intensity is expressed as a percentage of the total ion intensity to compensate for variations caused by factors unrelated to the analytical problem such as sample size and instrument sensitivity. The normalised data were subject to the **'characteristicity'** test (Eshuis *et al.*, 1977), a procedure which allows the selection of masses which show good reproducibility within a group of sample replicates (in this case the triplicate samples), and good specificity compared with other groups (that is, other strains). The characteristicity values were calculated using the following equations:

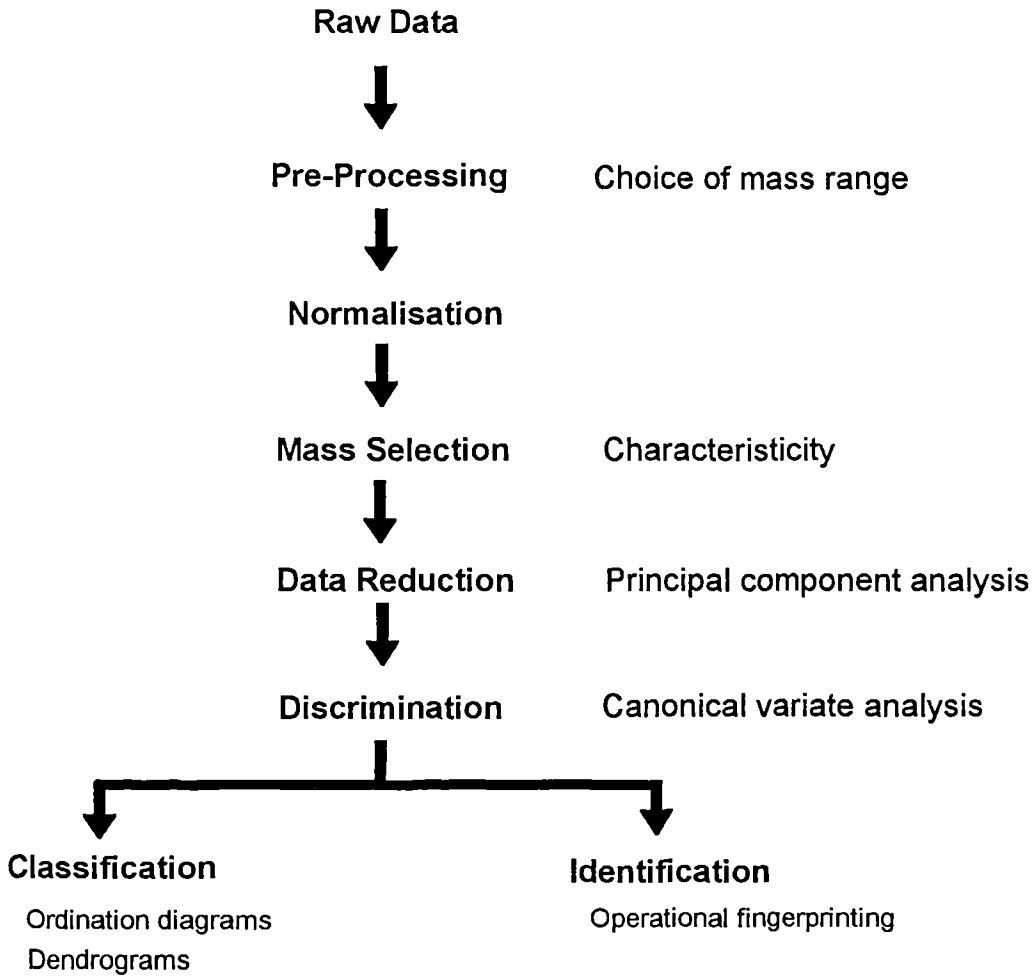


FIGURE 4-5. MAJOR STEPS INVOLVED IN MULTIVARIATE STATISTICAL ANALYSIS OF PYROLYSIS MASS SPECTRA.

- Inner variance or reproducibility (Kruscal, 1964a)

$$R_i = [1/n\sum_{j=1}^{n} V_{(i,j)}]^{-1}$$

where n is the number of strains (groups); R_i is the reproducibility of peak i; and $V_{(i,j)}$ is the variance of peak i in strain j.

- Outer variance or specificity (Kruscal, 1964b)

$$S_i = 1/n\sum_{j=1}^{n}(M_{(i,j)} - M_i)^2$$

where n is the number of strains (groups); S is the specificity of peak i; $M_{(i,j)}$ is the mean of peak i in strain j; and M_i is the mean of $M_{(i,j)}$ for all strains.

- Characteristicity (C_i)

$$C_i = R_i \times S_i$$

One hundred masses with the highest characteristicity values were selected and processed for all of the analyses.

The normalised data were analysed by using principal component analysis (PCA), a well known means of data reduction, to reduce the number of variables before discriminant analysis. Principal components accounting for over 0.1 % of the total variance were used as input data for the canonical variate analyses (CVA; MacFie *et al.*, 1978) in which successive axes are derived that are independent and seek to maximise the ratio of variation between *a priori* groups to that within these groups assuming that the structure of variation within the groups is the same. The results from the combined multivariate statistical procedure (PC-CVA) were presented as two or three dimensional ordination diagrams projecting the first two or three canonical variates (CVs). The data from the PC-CVA analyses were also

presented as Mahalanobis distances (Mahalanobis, 1936) using ***GENSTAT*** software. The resultant Mahalanobis distance matrices were converted to similarity matrices using Gower's coefficient S_G (Gower, 1971) and used to generate dendrograms by applying the UPGMA algorithm (Sneath & Sokal, 1973).

5. Artificial neural network analysis

5.1. Computer programs

The raw mass spectral data were transformed into a text-format file using the ***PYMENU*** program (Horizon Instruments). The text files containing the integrated ion counts were then transferred into a database program, ***PYMSFOX***, written in the FoxPro language (version 2.0; Fox Software Inc.). Creation of training and test data files, and training of the ANNs, were achieved by using the ***ANN*** program written in the C++ language and complied by using the Borland C++ package (version 3.0; Borland International Ltd., 100 Borland Way, Scotts Valley, CA 95067-3249, USA). IBM-PC compatible computers were used to train the ANNs.

5.2. Pre-processing

The raw data were normalised before training in order to ensure that all input variables were below one. The normalisation methods shown below were compared for training efficiency:

- Normalisation $i_j = v_j / v_{SUM}$
- Scaling $i_j = v_j / v_{MAX}$

where i_j is the input value of peak j; v_j is the raw ion count of peak j; v_{SUM} is the sum of ion counts of the mass spectra; and v_{MAX} is the biggest peak of the spectrum.

5.3. Training of artificial neural networks

All of the ANNs were trained using the epoch-based backpropagation algorithm (Rumelhart *et al.*, 1986) with a constant learning rate of 0.1 and a momentum of 0.9 as outlined in Figure 4-6. The efficacy of training was expressed as the root mean squared (RMS) error between the output (*o*) and target (*t*) vectors:

$$RMS\ error = \sqrt{[\sum_{j=1}^{n}(o_j - t_j)^2]} / n$$

where *n* is the number of training data; *o* the output vector of the mass spectra (data) *j*; and *t* the target vector for spectra *j*.

The initial values of the weight matrix in the ANNs were set randomly between -0.5 and +0.5. The initial RMS errors were always around 0.5 given this random assignment. When an RMS error of 0.005 was reached the ANNs were considered to be trained. The time taken to train ANNs is called an **'epoch'**, that is, the period for all of the input data to propagate the ANN. Root mean squared errors during training were recorded at intervals of 10 epochs and then used to construct a **'training curve'**. The weight matrix was also saved with a certain interval (e.g., 100 epochs) to calculate the RMS errors from the test data. The number of hidden layers and hidden neurons were set as 1 and 5, respectively, throughout the study.

5.4. Identification of mass spectra

Simple artificial neural network. The ANN topology consisted of 150 input, 5 hidden and 3 output neurons (150-5-3). The target vectors for the *Streptomyces* groups were set as (1,0,0) for the group A strains, (0,1,0) for the group B strains and (0,0,1) for the group C strains. The threshold value of 0.8 was used for strain identification. Strains were considered to be identified to one of the three streptomycete groups when only one of the three output neurons showed activity

Step 1 Set values of weights with small random values.

Step 2 Set δ (learning rate) and β (momentum term).

Step 3 Present a single set of inputs, and propagate data forward to obtain the predicted output (o).

$h=f(W1i+bias1)$ $o=f(W2h+bias2)$

Step 4 Calculate output layer error vector.

$d=o(1-o)(o-t)$

Step 5 Calculate hidden layer error vector.

$e=h(1-h)W2d$

Step 6 Adjust weights and bias between hidden and output layers.

$\Delta W2_t=\delta hd+\beta\Delta W2_{t-1}$, $bias2=\delta d$

$W2=W2+\Delta W2$

Step 7 Adjust weights and bias between input and hidden layers.

$\Delta W1_t=\delta ie+\beta\Delta W1_{t-1}$, $bias1=\delta e$

$W1=W1+\Delta W1$

Step 8 Repeat steps 3 to 7 until the error between output and target vectors is within the desired tolerance.

• The weight adjustments can be made after the entire set of input vectors has been propagated (**Epoch-based backpropagation**; Leonard & Kramer, 1990).

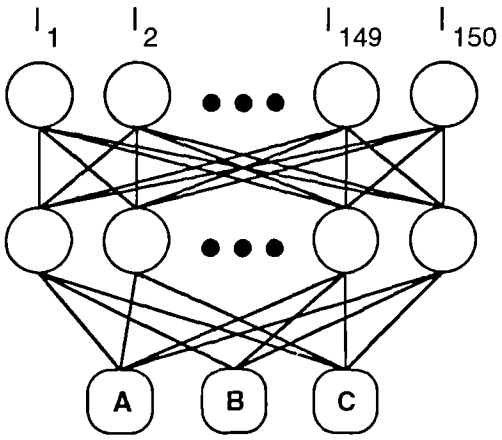
FIGURE 4-6. THE BACKPROPAGATION ALGORITHM (RUMELHART *et al.*, 1986). SYMBOLS: *i*, INPUT LAYER; *h*, HIDDEN LAYER; *o*, OUTPUT LAYER; *t*, TARGET VECTOR; *d*, OUTPUT; *e*, HIDDEN LAYER ERROR VECTORS; *bias1* AND *bias2*, BIAS VECTORS; AND *W1* AND *W2*, WEIGHT MATRICES.

over the threshold value. The topology of the simple ANN is illustrated in Figure 4-7.

Artificial neural network with an additional neuron (complex ANN). A network topology was developed to try and overcome the problem of strain misidentification, that is, assignment of non-streptomycete marker cultures to *Streptomyces* groups A, B and C. The complex ANN included a fourth neuron designed to accommodate members of the non-target taxa. The target vectors were (1, 0, 0, 1) for streptomycete group A strains, (0,1,0,1) for streptomycete group B strains, (0,0,1,1) for streptomycete group C strains and (0,0,0,0) for members of the non-target taxa. The network topology is given in Figure 4-7 and the identification scheme outlined in Figure 4-8. The threshold of the system was set at 0.8 for all of the output neurons. The identification scheme was implemented in the ***ANN*** program which allowed automated identification.

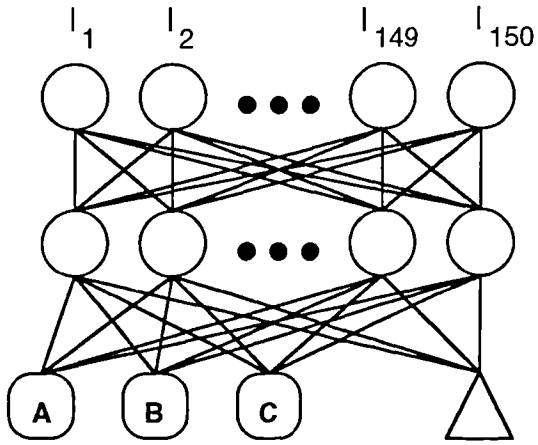
5.5. Generation of artificial neural networks

Six ANNs were constructed and evaluated (Table 4-4).



Target vectors	Target neurons		
Group A	1	0	0
Group B	0	1	0
Group C	0	0	1
Unknown	0	0	0

(A) SIMPLE ANN



Target vectors	Target neurons			Additional neuron
Group A	1	0	0	1
Group B	0	1	0	1
Group C	0	0	1	1
Unknown	0	0	0	0

(B) ANN WITH AN ADDITIONAL NEURON (COMPLEX ANN)

FIGURE 4-7. TEST TOPOLOGIES OF THE NEURAL NETWORKS.

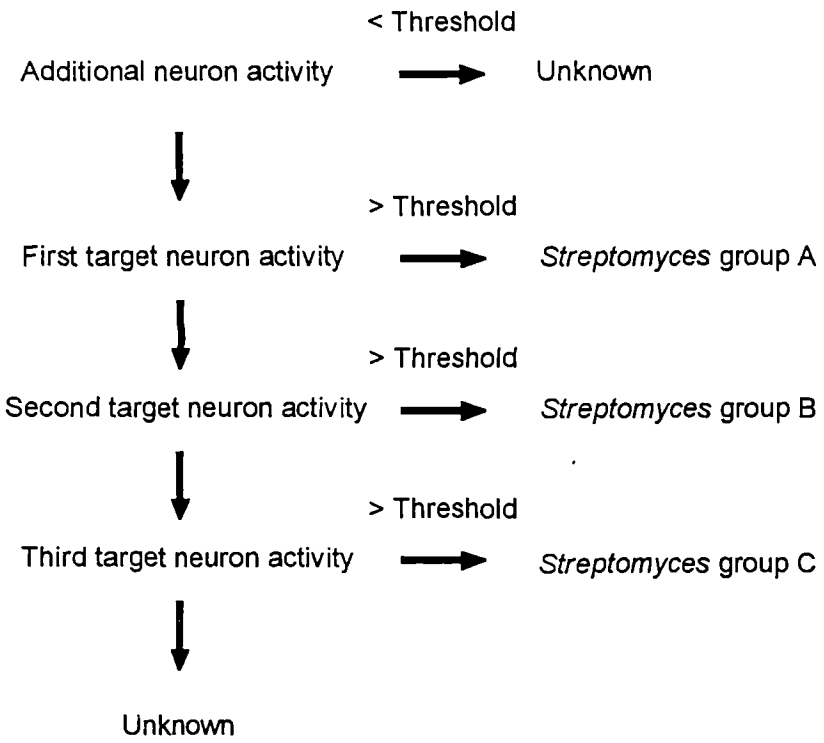


FIGURE 4-8. THE PROCEDURE USED FOR THE IDENTIFICATION OF *STREPTOMYCES* GROUP A, B AND C STRAINS USING AN ARTIFICIAL NEURAL NETWORK WITH AN ADDITIONAL NEURON. THE ORGANISMS WERE CONSIDERED TO BE UNIDENTIFIED WHEN MORE THAN TWO TARGET NEURONS SIMULTANEOUSLY SHOWED ACTIVITIES OVER THE THRESHOLD VALUE OF 0.8.

TABLE 4-4. ARTIFICIAL NEURAL NETWORKS.

Network[a]	Training dataset (No. of spectra)	Type of artificial neural network	Topology[b]
NET-1AS	1A (27)	Simple	150-5-3
NET-1AA	1A (27) & NT (33)	Complex	150-5-4
NET-1	1A (48) & NT (33)	Complex	150-5-4
NET-2	2 (42) & NT (33)	Complex	150-5-4
NET-3	3 (45) & NT (33)	Complex	150-5-4
NET-C	1A (48), 2 (42), 3 (45) & NT (33)	Complex	150-5-4

Abbreviation: NT, non-target database.

[a],All of the networks were trained using the backpropagation algorithm (Rumelhart *et al.*, 1986).

[b],Topology 150-5-3 represents 150 input neurons, 5 hidden neurons and 3 output neurons.

<BLANK>

C. Results and Discussion

1. Pyrolysis mass spectra

Typical pyrolysis mass spectra of representatives of the three *Streptomyces* groups are shown in Figure 4-9.

2. Effect of scaling on learning

Three ANNs were trained with normalised and scaled spectra from dataset 1A. The six resultant training curves are shown in Figure 4-10. The ANNs trained with the normalised data converged at 7760, 7540 and 8470 epochs given a RMS error of 0.005, and the ANNs with the scaled data at 1600, 790 and 1090 epochs, respectively. It is clear from Figure 4-10 that the scaling procedure facilitates training hence all spectral data for the subsequent ANN analyses were scaled prior to training. The advantage of scaling individual mass spectra was also observed by Goodacre *et al.* (1994a).

3. Comparison of datasets 1A and 1B in multivariate statistical and artificial neural network analyses

Datasets 1A and 1B (Table 4-3) were the subject of separate PC-CVA analyses (Figure 4-11). The *Streptomyces* strains from dataset 1A were assigned to three groups which corresponded to *Streptomyces* groups A, B and C. A similar result was obtained in the analysis of dataset 1B though in this case *Streptomyces* strain A35 was recovered as a single membered cluster. It is evident that a substantial amount of noise is present in the mass spectra of *Streptomyces* strain A35 though the cause of this is not clear. It can be concluded that *Streptomyces* strain A35 was not identified when the **'operational fingerprinting'** procedure was applied to dataset 1B.

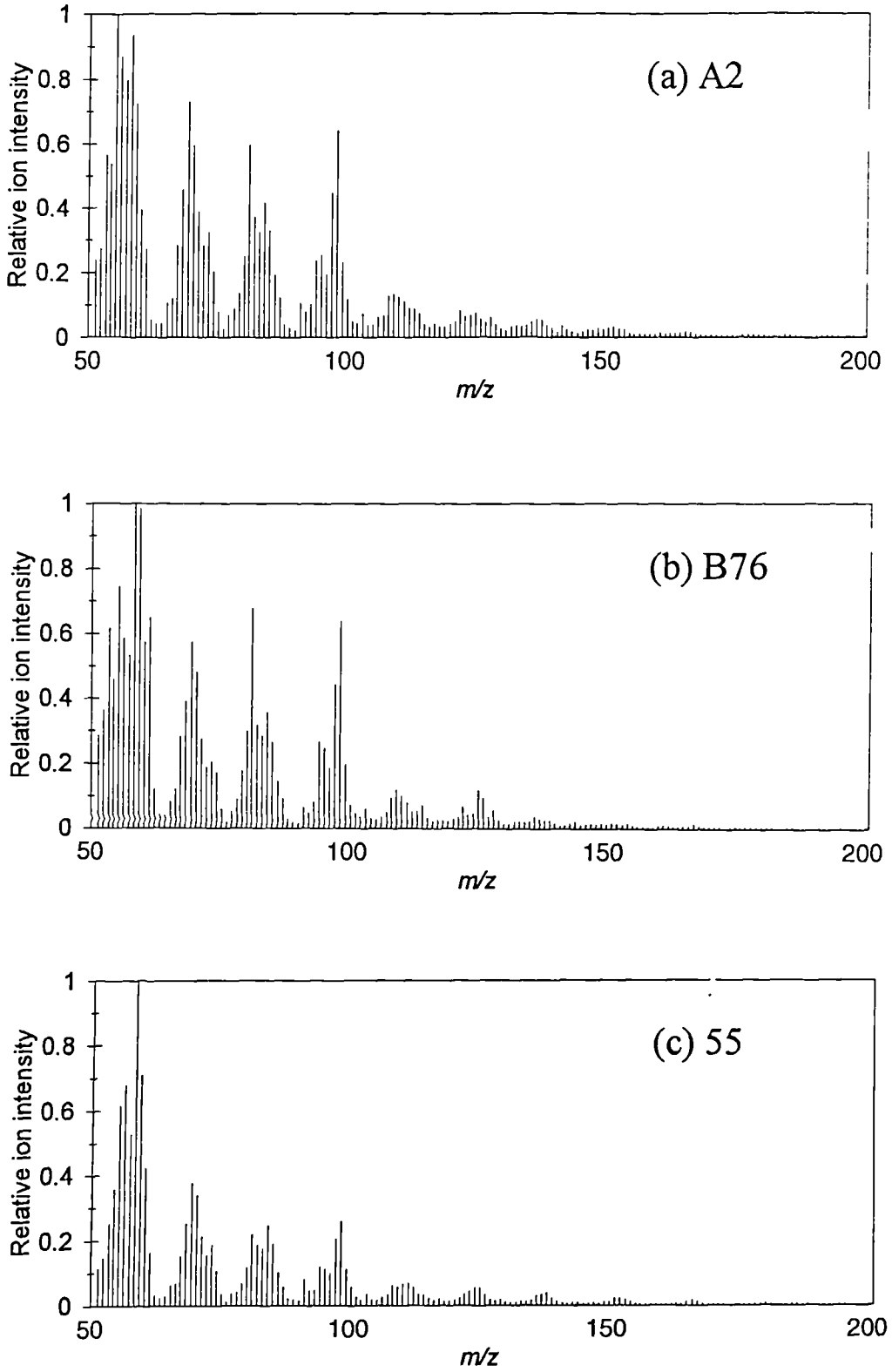


FIGURE 4-9. PYROLYSIS MASS SPECTRA OF REPRESENTATIVES OF *STREPTOMYCES* GROUPS A, B AND C.

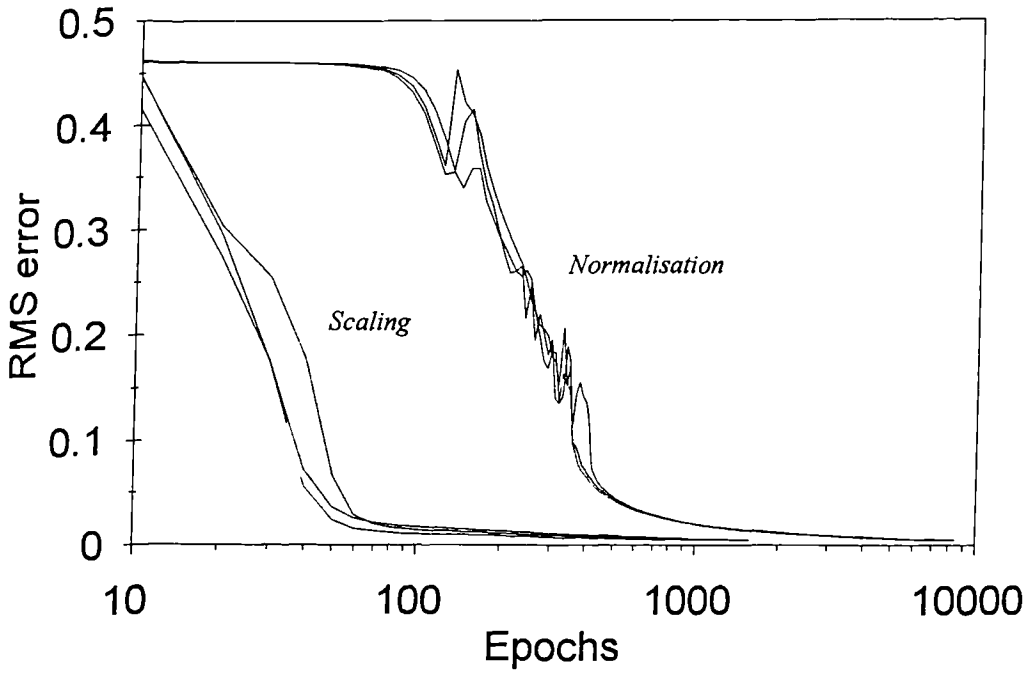


FIGURE 4-10. TRAINING CURVES OF SIX NEURAL NETWORKS. THREE WERE TRAINED WITH THE NORMALISED DATA AND THE OTHER THREE WITH THE SCALED TRAINING DATA IN EACH CASE USING THE FORTY-EIGHT PYROLYSIS MASS SPECTRA FROM DATASET 1A. RMS, ROOT MEAN SQUARED.

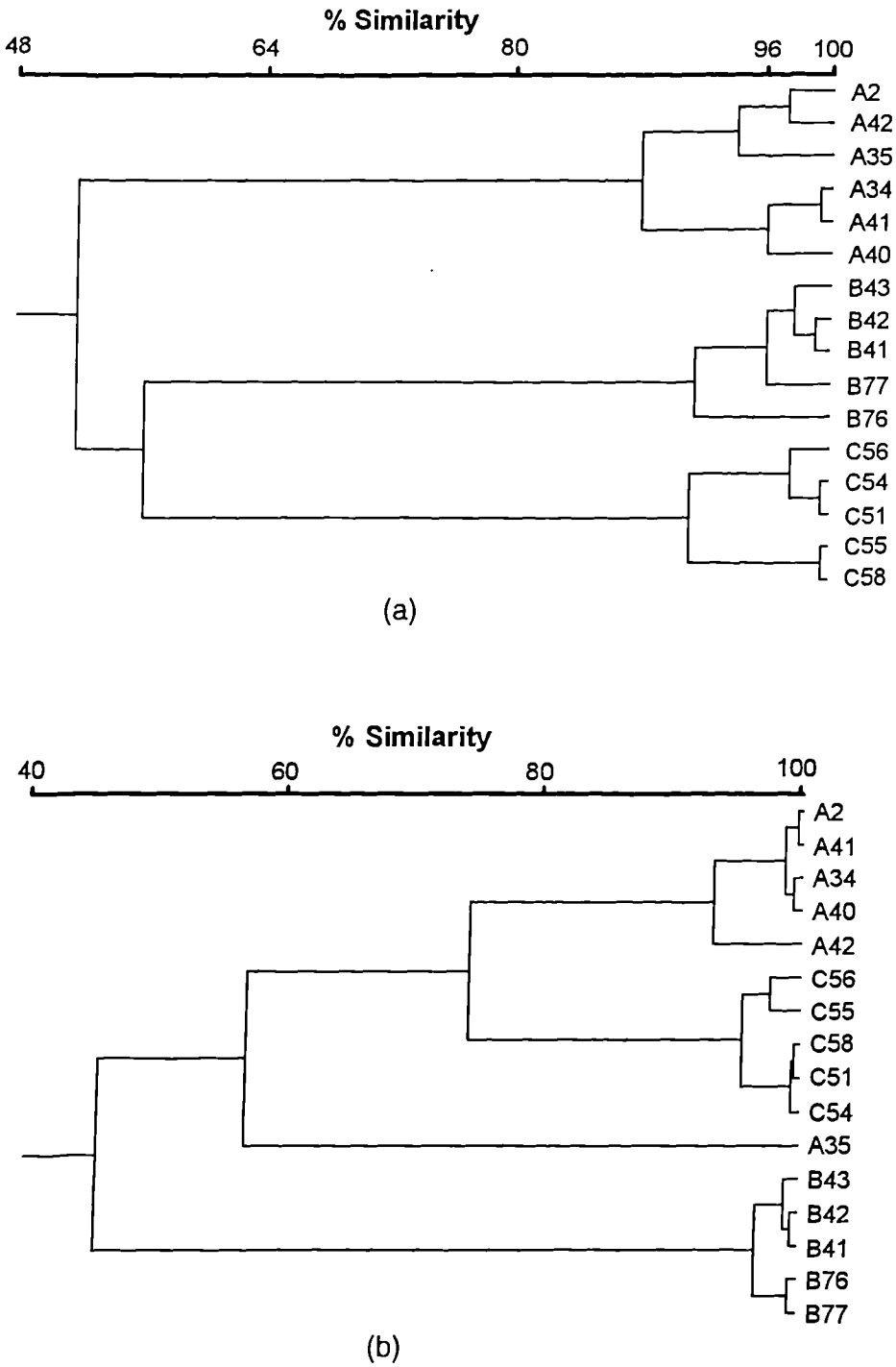


FIGURE 4-11. DENDROGRAMS SHOWING RELATIONSHIPS BETWEEN REPRESENTATIVE *STREPTOMYCES* STRAINS BASED ON: (a) DATASET 1A AND (b) DATASET 1B. DATA WERE ANALYSED USING GENSTAT SOFTWARE WITH CLUSTERING ACHIEVED BY APPLYING THE UNWEIGHTED-PAIR GROUP METHOD WITH ARITHMETIC AVERAGES ALGORITHM.

An ANN (NET-1AS; Table 4-4) was constructed using twenty-seven of the pyrolysis mass spectra from dataset 1A, namely information on *Streptomyces* strains A2, A34, A42, B42, B43, B76, C51, C55 and C56 (Table 4-3). The network was used to identify the remaining mass spectra held in dataset 1A and all of the mass spectra from dataset 1B; the results of the strain identifications are summarised in Table 4-5. The twenty-one mass spectra from dataset 1A and the forty-eight mass spectra from dataset 1B were identified correctly, including those derived from *Streptomyces* strains A35, A40, A41, B41, B77, C54 and C58 which were not been used to train the network.

The correctness and overall performance of the strain identifications were assessed by using the mean RMS error of all of the tested mass spectra. The mean RMS error of NET-1AS with respect to the streptomycete strains was 0.0081±0.0094. The highest error was with the spectra of *Streptomyces* strain B42 (0.0404) and the lowest error with those of *Streptomyces* strain B43 (0.0013). It is very encouraging that *Streptomyces* strain A35, which was not identified using the **'operational fingerprinting'** procedure, was accurately identified in the ANN analysis. It can also be said that the ANN ignored the noise in the mass spectra of *Streptomyces* strain A35 in dataset 1B and hence was able to detect the specific pattern exhibited by members of streptomycete group A. It can be concluded from these results that the analysis of the pyrolysis mass spectral data using the ANN provides a relatively robust way of identifying the streptomycete isolates.

TABLE 4-5. IDENTIFICATION RESULTS OBTAINED FOR TEST SET 1 USING NET-1AS BASED ON TRIPLICATE PYROLYSIS MASS SPECTRA.

Strain	Dataset	Activity of output neuron*			Identification: *Streptomyces* group
		1	2	3	
A2	1B	0.998±0.001	0.003±0.000	0.003±0.001	A
A34	1B	0.998±0.000	0.003±0.000	0.003±0.000	A
A35	1A	0.998±0.000	0.002±0.000	0.006±0.002	A
A35	1B	0.999±0.000	0.002±0.000	0.003±0.001	A
A40	1A	0.999±0.000	0.004±0.000	0.001±0.000	A
A40	1B	0.997±0.001	0.003±0.000	0.003±0.001	A
A41	1A	0.998±0.001	0.003±0.000	0.003±0.001	A
A41	1B	0.994±0.005	0.003±0.001	0.006±0.003	A
A42	1B	0.999±0.000	0.004±0.001	0.002±0.000	A
B41	1A	0.001±0.000	0.990±0.006	0.014±0.010	B
B41	1B	0.001±0.000	0.991±0.002	0.010±0.003	B
B42	1B	0.001±0.001	0.977±0.015	0.039±0.027	B
B43	1B	0.002±0.000	0.998±0.001	0.001±0.001	B
B76	1B	0.008±0.005	0.997±0.001	0.001±0.000	B
B77	1A	0.001±0.000	0.970±0.005	0.035±0.005	B
B77	1B	0.044±0.025	0.992±0.004	0.000±0.000	B
C51	1B	0.002±0.001	0.008±0.001	0.995±0.001	C
C54	1A	0.007±0.003	0.005±0.000	0.990±0.003	C
C54	1B	0.003±0.000	0.009±0.001	0.991±0.001	C
C55	1B	0.016±0.012	0.005±0.001	0.983±0.009	C
C56	1B	0.002±0.000	0.012±0.003	0.993±0.001	C
C58	1A	0.003±0.001	0.007±0.002	0.994±0.001	C
C58	1B	0.001±0.000	0.010±0.001	0.996±0.001	C

*, Activity values are shown as the mean and standard deviations of triplicate spectra.

4. Comparison of simple and complex artificial neural networks

Network NET-1AS (Table 4-4) consisted of only three output neurons and was only trained using data from streptomycete (target) strains. The network was also used to examine 414 pyrolysis mass spectra derived from the non-target actinomycetes in order to see whether the system could recognise new patterns. The three training curves of network NET-1AS based on the training data, the additional streptomycete test data and on the non-streptomycete test data, as well as on the strain identification results, at the corresponding epochs are shown in Figure 4-12. It is clear from this figure that the level of identification of the target strains, presented as the RMS error, improves along the training curve. However, the mean RMS error for the non-target strains was constant around 0.5 and the identification rates for non-target strains, that is, the per cent of non-target test strains that were recognised as unknown decreased from 100% to 7% based on the 414 pyrolysis mass spectra from the database containing information on the non-streptomycete strains. The results of the strain identification are summarised in Table 4-6.

It is evident from the results that network NET-1AS has the ability to identify members of the target streptomycete groups but is unable to recognise non-target strains. The network tried to assign all input pyrolysis mass spectral data, including those derive from non-target strains, to one of the three target taxa. These results are not unexpected as it is well known that the supervised learning algorithm, notably the backpropagation algorithm, generates neural networks which behave poorly when challenged by unknown patterns outside the range of the training data (Freeman & Skapura, 1991).

An ANN topology was developed to facilitate the rejection of non-target strains. The improved ANN contained a fourth neuron designed to help distinguish

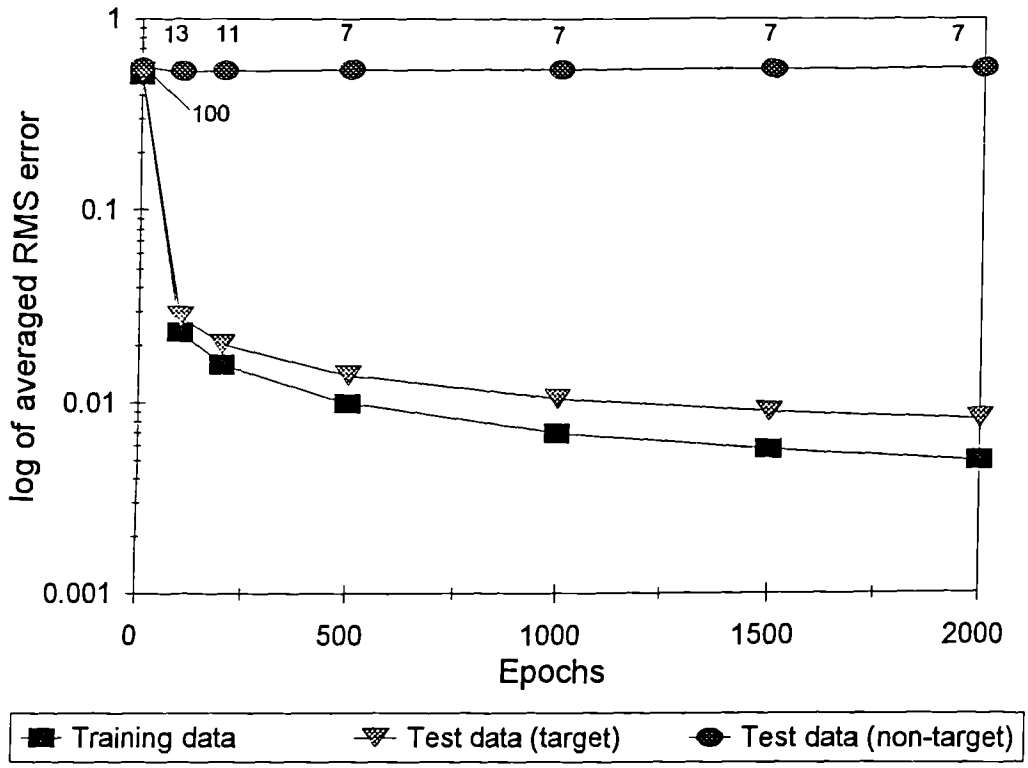


FIGURE 4-12. TRAINING CURVES OF NETWORK NET-1AS. THE WEIGHT MATRIX WAS SAVED AT 0, 100, 200, 500, 1000 AND 1500 EPOCHS AND THE CORRESPONDING RMS ERRORS CALCULATED FROM TEST DATA SETS CONSISTING OF TARGET (*STREPTOMYCES* GROUPS A, B AND C) AND NON-TARGET DATA, RESPECTIVELY. THE PER CENT OF UNIDENTIFIED PYROLYSIS MASS SPECTRA (AS CORRECT ANSWER) IN THE NON-TARGET DATA (414 SPECTRA) ARE INDICATED AT THE POINTS OF CORRESPONDING RMS ERROR.

TABLE 4-6. COMPARISON OF IDENTIFICATION RESULTS FOR NETWORKS NET-1AS AND NET-1AA.

Taxon	Number of strains	NET-1AS Strains identified as: *Streptomyces*				NET-1AA Strains identified as: *Streptomyces*			
		Group A	Group B	Group C	Unknown	Group A	Group B	Group C	Unknown
Streptomyces group A	6 (36)	6(36)	0(0)	0(0)	0(0)	6(36)	0(0)	0(0)	0(0)
Streptomyces group B	5 (30)	0(0)	5(30)	0(0)	0(0)	0(0)	5(30)	0(0)	0(0)
Streptomyces group C	5 (30)	0(0)	0(0)	5(30)	0(0)	0(0)	0(0)	5(30)	0(0)
Actinomadura spp.	8 (24)	0(0)	23(19)	0(1)	1(4)	0(0)	0(0)	0(0)	8(24)
Mycobacterium spp.	34 (102)	32(97)	0(0)	2(4)	0(1)	0(0)	0(0)	0(0)	34(102)
Nocardia spp.	40 (120)	38(106)	0(0)	0(0)	2(14)	0(0)	0(0)	0(0)	40(120)
Nocardiopsis spp.	9 (27)	0(0)	1(3)	6(20)	2(4)	0(0)	0(0)	0(0)	9(27)
Saccharomonospora spp.	28 (84)	28(74)	0(5)	0(0)	3(5)	0(0)	0(0)	0(0)	28(84)
Streptosporangium spp.	19 (57)	0(8)	19(48)	0(0)	0(1)	0(0)	0(0)	0(0)	19(57)

target streptomycete groups from members of non-target taxa. This complex ANN, named NET-1AA, was constructed using training data which consisted of the twenty-seven mass spectra used to train NET-1AS and additional pyrolysis mass spectra from *Actinomadura* strains A9, A16 and A32, *Nocardia* strains N36, N317, N318, N898 and N1116, and *Nocardiopsis* strains A14 and A120. The results obtained using the NET-1AA network are summarised in Table 4-6; the training curves obtained for this network are presented in Figure 4-13. It is evident from the training curves that the ability of NET-1AA to recognise non-target patterns was gained in the early stages of training (before 500 epochs), that is, at around a RMS error of 0.11 for the non-target pyrolysis mass spectra; this level of RMS error was constant until the training of the network ceased at 3070 epochs.

It is very encouraging that the complex ANN, namely NET-1AA, was successfully used to identify the target strains following the identification scheme outlined in Figure 4-8. The network also recognised differences between the target mass spectral patterns and those held in the non-target database, the latter included data new to the network on representatives of the genera *Mycobacterium*, *Saccharomonospora* and *Streptosporangium*. Complex ANNs, such as the one generated in this investigation, should prove to be very useful for detecting novel organisms and in circumstances where mis-identification needs to be avoided, particularly in identification of clinically significant microorganisms (Freeman *et al.*, 1994a).

5. Statistical comparison of pyrolysis mass spectral data from datasets 1A, 2 and 3

The pyrolysis mass spectral results from the first three datasets, namely datasets 1A, 2 and 3, were analysed separately using the multivariate statistical procedure. The three *Streptomyces* groups were recovered consistently without exception (data not shown). The results from the three datasets were then

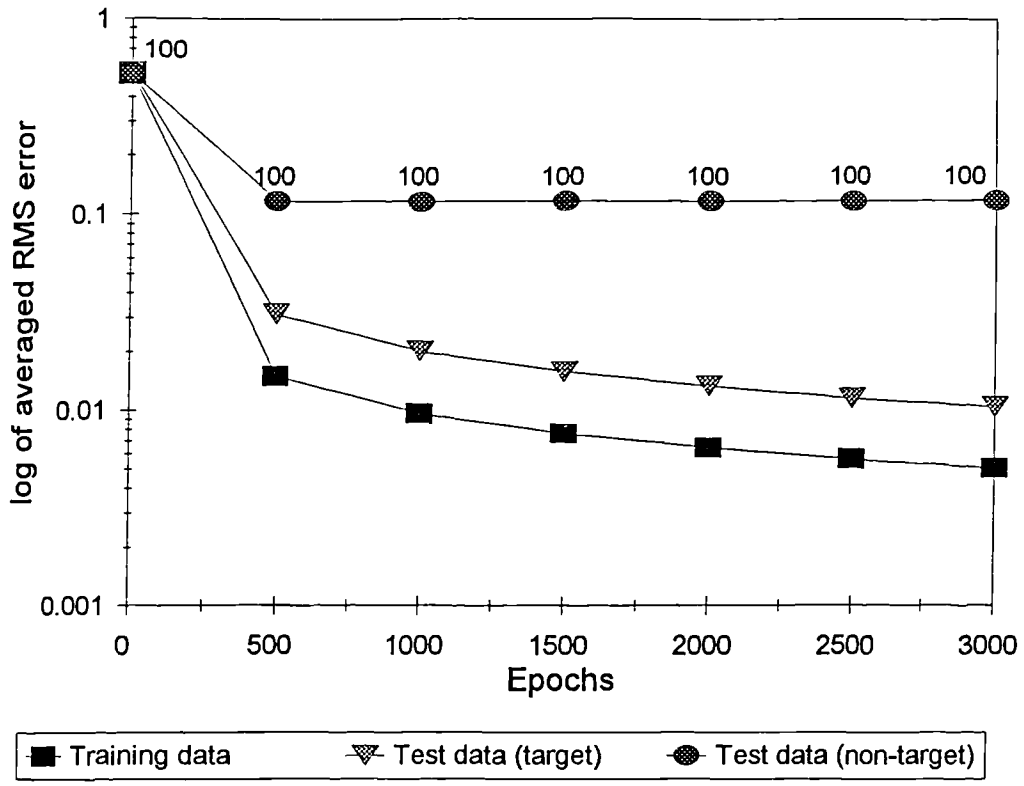


FIGURE 4-13. TRAINING CURVES OF NETWORK NET-1AA. THE WEIGHT MATRIX WAS SAVED AT 0, 500, 1000, 1500, 2000, 2500 AND 3000 EPOCHS. FOR THE DETAILS SEE LEGEND TO FIGURE 4-12.

combined and analysed to statistically evaluate the long-term reproducibility of the system (Figure 4-14). It is evident that the resultant groups correspond not to the three taxonomic groups but to the times when the organisms were pyrolysed. It is also clear from the dendrogram that the variation found within the different datasets was more significant than that found between members of the three test taxa. It can be concluded from this study, and from the earlier investigation of Shute *et al.* (1988), that pyrolysis mass spectra derived from different batch analyses cannot be compared directly using conventional statistical analyses. It has already been pointed out that Shute *et al.* (1988) attributed this problem to instrumental instability.

It can also be seen from Figure 4-14 that the pyrolysis mass spectra held in dataset 3 were no closer to the corresponding results from dataset 2 than to those from dataset 1. These results imply that there may not be any consistent time-dependent drift in the performance of instrument. However, it is important to realise that despite the lack of long-term reproducibility of PyMS, the system can still be of value in the classification and identification of micro-organisms provided that test strains are analysed in the same batch analysis (Gutteridge, 1987; Goodfellow 1995c). Thus, in the present preliminary investigation, the three *Streptomyces* groups were consistently recovered in each of the batch analyses. Nevertheless, it is also clear that the *Streptomyces* strains cannot be identified by comparison with reference mass spectra derived from different batches. This means that the generation of a conventional pyrolysis mass spectral database is not feasible for long-term studies.

6. Artificial neural network comparison of pyrolysis mass spectral data of datasets 1A, 2 and 3

Three ANNs showing the complex topology described in the previous section were constructed using the training data given in Table 4-4. The mass spectra held

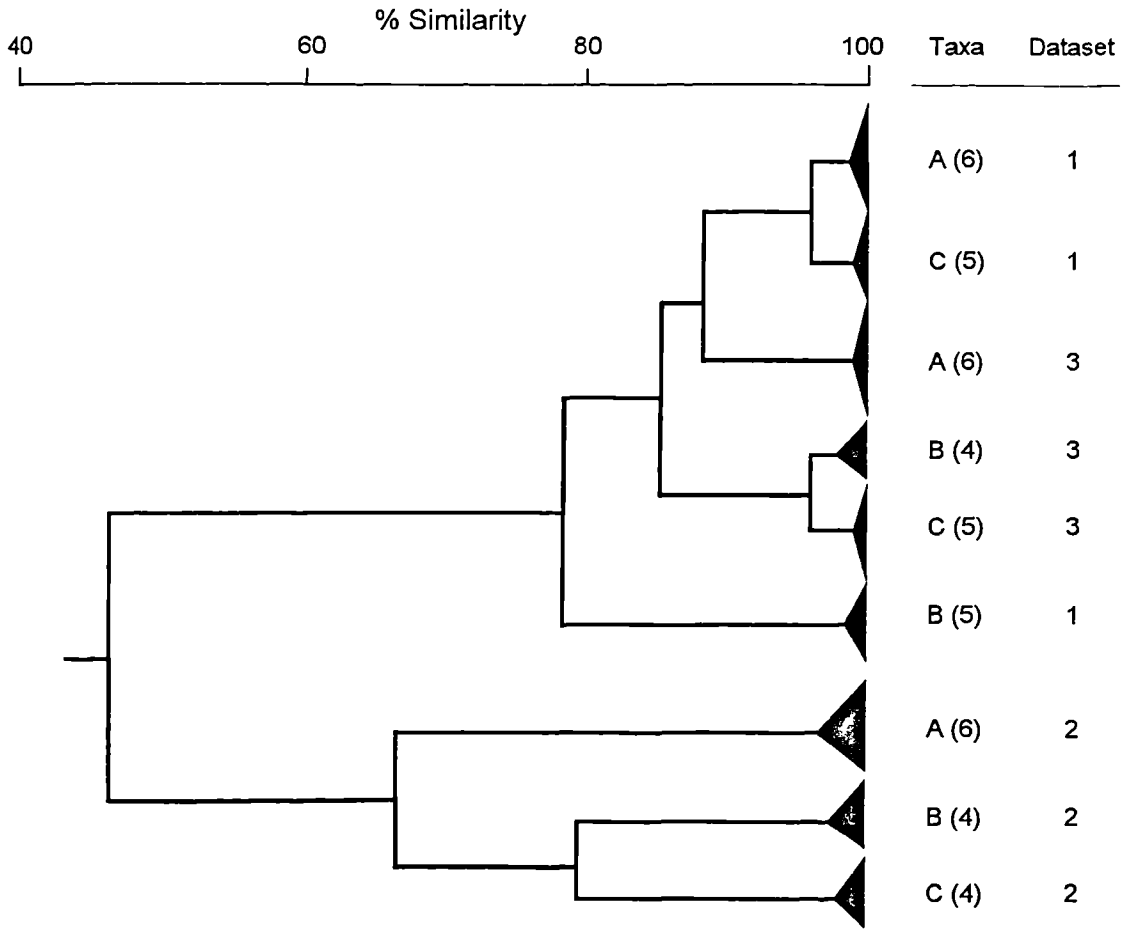


FIGURE 4-14. RESULTS OF MULTIVARIATE STATISTICAL ANALYSIS OF THE COMBINED DATASET WHICH WAS COMPOSED OF DATASETS 1A, 2 AND 3. NUMBERS IN PARENTHESES INDICATE THE NUMBER OF PYROLYSIS MASS SPECTRA.

in each dataset, namely datasets 1,2 and 3, were identified using two ANNs which were trained using the remaining two datasets; this process can be called 'cross identification' since the three datasets were composed of the same strains (see Table 4-3). The cross identifications are summarised in Table 4-7.

When the pyrolysis mass spectra of dataset 2 were examined using network NET-1, which was trained using the mass spectra held in dataset 1A and those from representatives of the non-target taxa, all of the *Streptomyces* group C strains were correctly identified. In contrast, none of representatives of the *Streptomyces* groups A and B were identified. Similarly, only the *Streptomyces* group A strains of dataset 3 were correctly identified using network NET-1. In contrast, none of the organisms in datasets 1 and 3 were identified by using network NET-2. These poor results may be due to the fact that the mass spectra held in dataset 2 are completely different from those held in the other two datasets (Figure 4-14). Only *Streptomyces* group A spectra of dataset 1 were recognised using network NET-3. This result was anticipated as network NET-1 only recognised the *Streptomyces* group A strains of dataset 3 (Table 4-7).

Some of the test strains were identified using the ANNs but none of them were identified statistically. It can, therefore, be concluded from Table 4-7 that ANNs trained with pyrolysis mass spectra from only one dataset (batch) are not sufficiently robust to identify the same streptomycete strains pyrolysed at different times. However, all of the non-target strains were correctly recognised as unidentified with all three networks when members of the non-target taxa were pyrolysed in the different batches.

TABLE 4-7. CROSS IDENTIFICATION RESULTS FROM NETWORKS NET-1, NET-2 AND NET-3. ALL OF THE 414 NON-TARGET MASS SPECTRA WERE RECOGNISED AS UNKNOWN [+].

Dataset		No. of spectra	NET-1 Identified to :				NET-2 Identified to :				NET-3 Identified to :			
			A	B	C	U	A	B	C	U	A	B	C	U
1A	A	18	-	-	-	-	0	0	0	18	***8***	0	0	10
1A	B	15	-	-	-	-	0	0	0	15	0	0	4	11
1A	C	15	-	-	-	-	0	0	0	15	0	0	0	15
2	A	18	***1***	0	8	9	-	-	-	-	0	0	0	18
2	B	12	0	***5***	3	4	-	-	-	-	0	0	1	11
2	C	12	0	0	***12***	0	-	-	-	-	0	0	0	12
3	A	18	***18***	0	0	0	0	0	0	18	-	-	-	-
3	B	12	1	***1***	0	10	0	0	0	12	-	-	-	-
3	C	15	0	9	***0***	6	0	0	0	15	-	-	-	-

[+], correct identifications are shown in bold-italic type.

7. Spectral identification of data in datasets 4, 5 and 6 using 'operational fingerprinting' and artificial neural network analysis

Network NET-C was trained on the combined pyrolysis mass spectra from datasets 1A, 2 and 3 in order to construct a more generalised neural network (Table 4-4). The information held in dataset 4 was derived from reference *Streptomyces* strains whereas the mass spectral data held in datasets 5 and 6 were from reference *Streptomyces* strains and soil isolates putatively assigned to *Streptomyces* groups A, B and C (Table 4-3). The information held in the three test datasets were used to identify the *Streptomyces* strains by using conventional **'operational fingerprinting'** and ANN procedures.

All of the reference strains of *Streptomyces*, apart from *Streptomyces* strains A35 and B76, formed three groups when the information held in dataset 4 was the subject of multivariate statistical analysis (Figure 4-15). *Streptomyces* strains A35 and B76 were not assigned to any of the target groups and hence were not correctly identified using the **'operational fingerprinting'** procedure. Most of the *Streptomyces* group A and B strains, apart from *Streptomyces* strains A34 and B76, were correctly identified using networks NET-1 and NET-3. In contrast, none of the *Streptomyces* strains were correctly identified using network NET-2. All thirteen reference strains were successfully identified when network NET-C was used. The identified organisms included *Streptomyces* strains C51 and C58 which were not identified using networks NET-1, 2 and 3.

Two mechanisms may account for the improvement in the predictive ability of neural network NET-C. One is **'summation'**, which can be seen in the cases of *Streptomyces* strains A2, B76 and C55, whereby network NET-C gained the combined pattern recognition knowledge of networks NET-1, NET-2 and NET-3. In the case of *Streptomyces* strains C51 and C58 the mechanism can be explained in

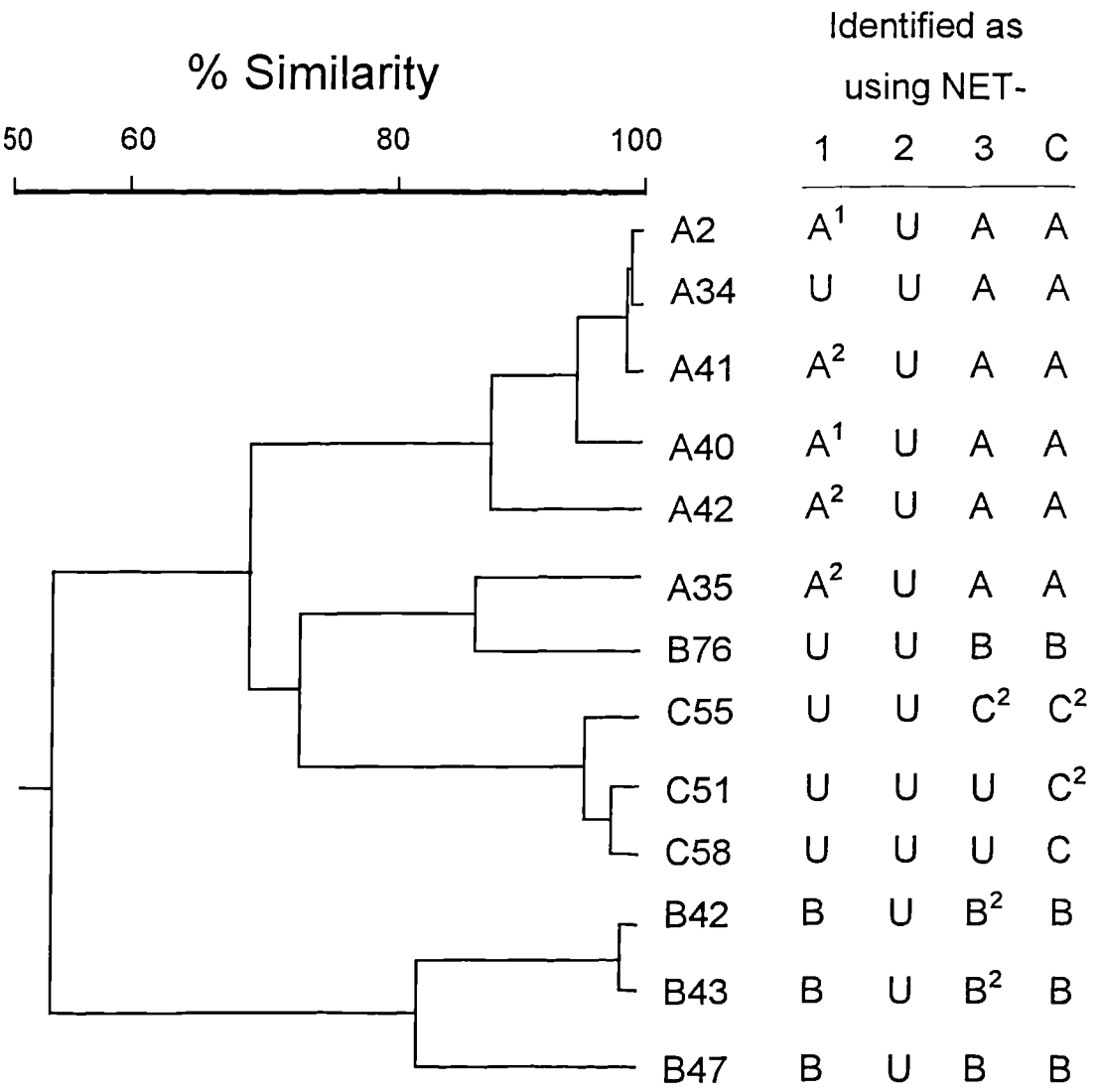


FIGURE 4-15. DENDROGRAM OBTAINED FROM THE PC-CVA ANALYSIS OF INFORMATION HELD IN DATASET 4. THE IDENTIFICATION RESULTS OBTAINED WITH NETWORKS NET-1, NET-2, NET-3 AND NET-C ARE INDICATED ON THE RIGHT-HAND SIDE OF THE DENDROGRAM. ALL OF 414 NON-TARGET MASS SPECTRA WERE IDENTIFIED AS UNKNOWN. A^N INDICATES THAT ONLY N OUT OF TRIPLICATE SPECTRA WAS/WERE ASSIGNED TO *STREPTOMYCES* GROUP A WHERE N IS 1 OR 2.

terms of **'generalisation'**, that is, when ANNs generalise they can predict patterns beyond the ones previously seen by the network provided the training data spans the data space (Freeman & Skapura, 1991).

Datasets 5 and 6 were designed to facilitate practical **'operational fingerprinting'** since fresh soil isolates were pyrolysed together with reference strains of *Streptomyces* groups A, B and C (Figures 4-16 and 4-17). All of the fresh isolates were correctly identified using the **'operational fingerprinting'** technique. Only some of the *Streptomyces* strains in dataset 5 were correctly identified using networks NET-1, NET-2 and NET-3 (Figures 4-16). Network NET-1 performed relatively well but NET-2 and NET-3 were poorly predictive. In contrast, all of the streptomycete pyrolysis mass spectral data held in dataset 5 were correctly identified using network NET-C. It is particularly encouraging that *Streptomyces* strains B42, B7316, C54, C56 and C58, which were mis-identified with at least one of the three other networks, namely NET1, NET-2 and NET-3, were correctly identified by using network NET-C.

The *Streptomyces* strains used to generate dataset 6 were examined by the **'operational fingerprinting'** and ANN procedures (Figure 4-17). As expected, all nine of the representative strains of *Streptomyces* groups A, B and C were assigned to groups corresponding with these taxa indicating that all of the test strains were correctly identified using the **'operational fingerprinting'** procedure. When the mass spectra held in dataset 6 were examined by the ANNs all of the spectra were correctly identified using networks NET-1 and NET-C whereas only one out of the nine strains was correctly recognised by networks NET-2 and NET-C. The close relatedness found between the mass spectra held in datasets 1 and 6 was also evident from the multivariate statistical analysis of the combined dataset, that is, the dataset which consisted of information derived from datasets 1 and 6 (Figure 4-18).

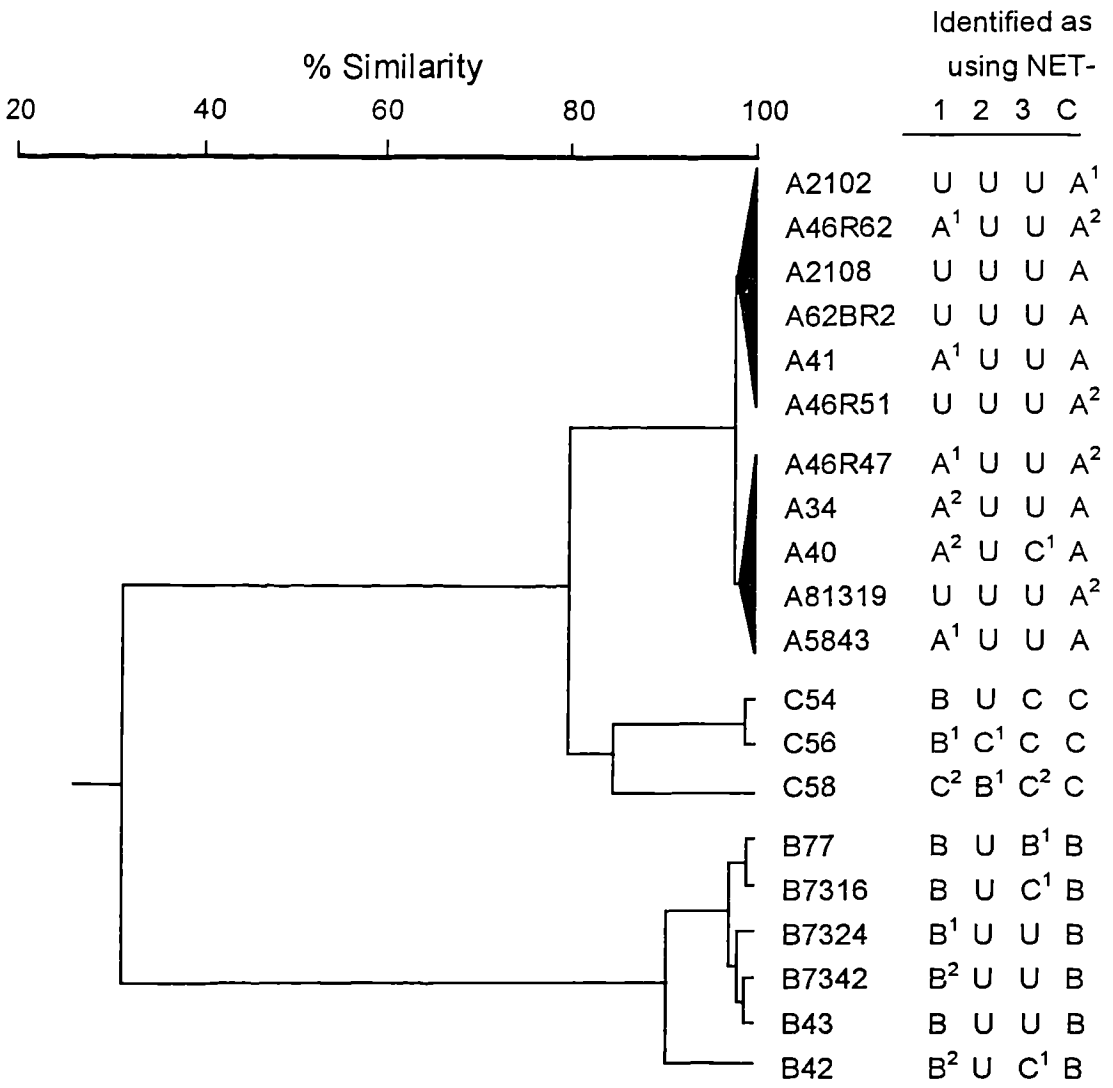


FIGURE 4-16. DENDROGRAM SHOWING RELATIONSHIPS BETWEEN REPRESENTATIVE *STREPTOMYCES* STRAINS BASED ON MULTIVARIATE STATISTICAL ANALYSIS OF THE PYROLYSIS MASS SPECTRA RESULTS HELD IN DATASET 5. THE IDENTIFICATION RESULTS OF THE ANN ANALYSES ARE INDICATED AT THE RIGHT-HAND SIDE OF THE DENDROGRAM. ALL OF 414 NON-TARGET MASS SPECTRA WERE IDENTIFIED AS UNKNOWNS. FOR DETAILS SEE LEGEND TO FIGURE 4-15.

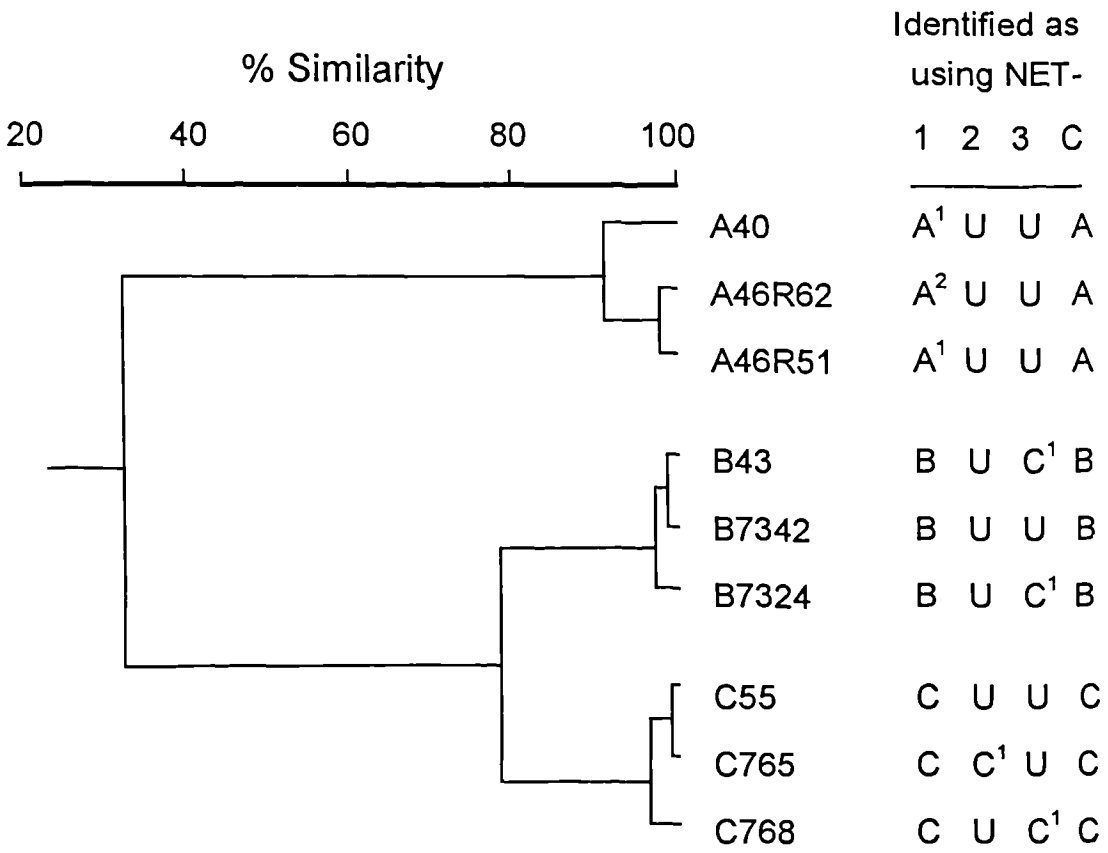


FIGURE 4-17. DENDROGRAM SHOWING RELATIONSHIPS BASED ON MULTIVARIATE STATISTICAL ANALYSIS OF THE PYROLYSIS MASS SPECTRA OF DATASET 6. THE IDENTIFICATION RESULTS OF THE ARTIFICIAL NEURAL NETWORK ANALYSES ARE SHOWN AT THE RIGHT-HAND SIDE OF THE DENDROGRAM. ALL 414 NON-TARGET MASS SPECTRA WERE IDENTIFIED AS UNKNOWNS. FOR DETAILS SEE LEGEND TO FIGURE 4-15.

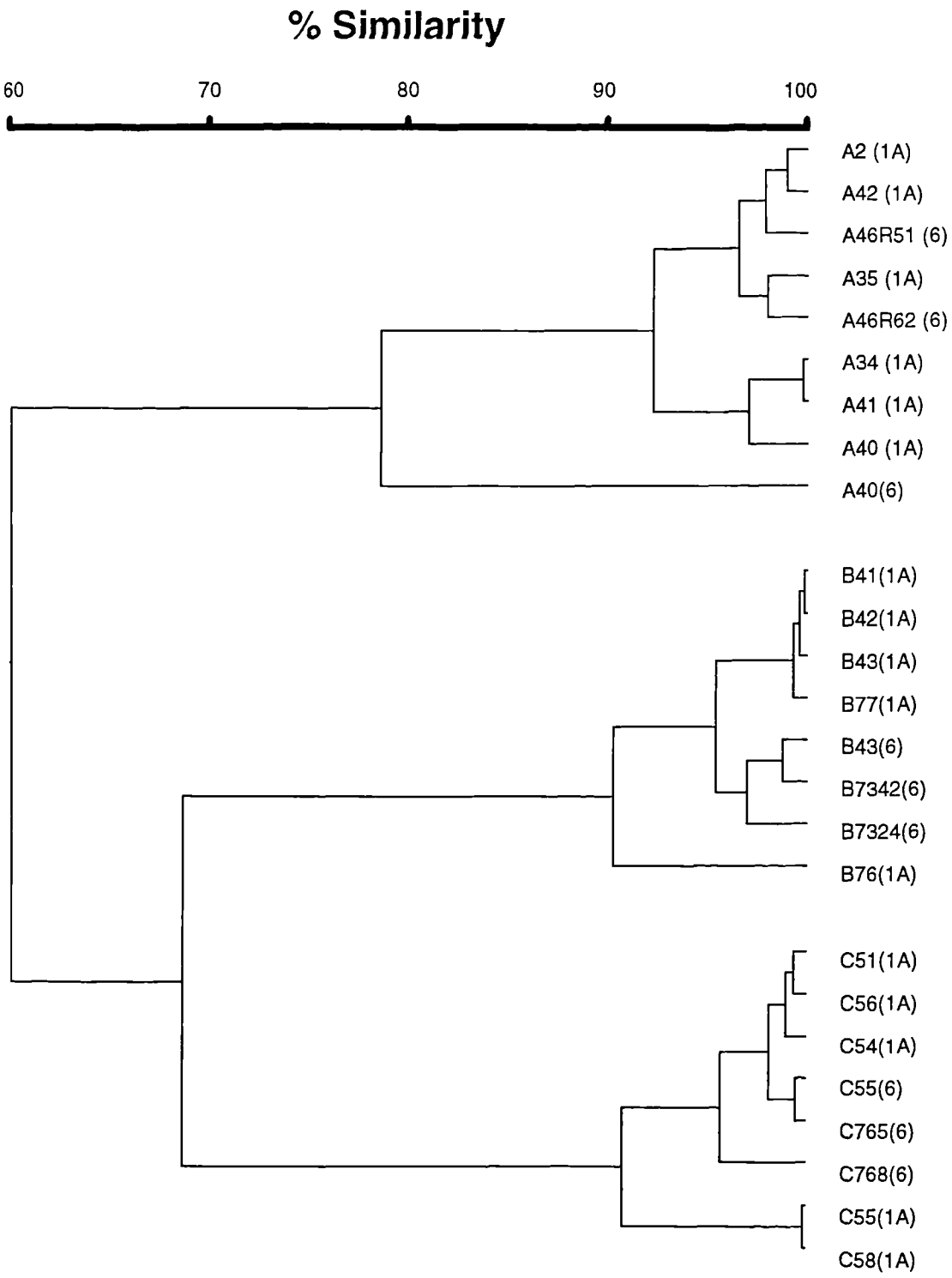


FIGURE 4-18. PRINCIPAL COMPONENT-CANONICAL VARIATE ANALYSIS OF THE COMBINED DATASET COMPRISING THE MASS SPECTRA OF DATASETS 1A AND 6.

This result implies that the nature of the variation caused by 'instrumental drift' is random rather than time-dependent.

It has already been mentioned that the development of rapid, reliable and cost-effective methods are at a premium for the classification, identification and typing of microorganisms. Curie point pyrolysis mass spectrometry is increasingly being used for such purposes as it allows rapid automated acquisition of data derived from whole organism components and requires minimal sample preparation (Magee, 1993, 1994). The system can, therefore, be used to evaluate taxonomic structures derived from the application of other taxonomic procedures as well as for the selection of representative strains prior to the use of more time-consuming and laborious methods, notably numerical taxonomy and nucleic acid sequencing. The use of the system for identification purposes has been hampered by the relatively high cost of the instrument and by inappropriate data handling techniques which have not been able, until now, to overcome the lack of long-term reproducibility of the system.

It can be concluded from the present investigation that the combined use of PyMS and ANN provided an effective procedure for the identification of members of *Streptomyces* groups A, B and C. In addition, the identification of streptomycete isolates was achieved a twenty month period. This new procedure eliminates the necessity for **'operational fingerprinting'**, that is, the need to analyse reference and test strains at the same time, as well as the requirement to duplicate samples which is current practice in conventional PC-CVA analyses. However, it must be emphasised that the confidence that can be placed in the identification of unknown strains using ANNs depends on the topology of the network, the training algorithm, and eventually on the quality of the training dataset (Freeman & Skapura, 1991). In

addition, the performance of trained ANNs should be carefully checked before being applied to real situations.

It will be interesting to see whether ANNs applied for purposes other than for the identification of bacteria, such as in chemical quantification studies (Goodacre *et al.*, 1992, 1993a, 1993b, 1994a; Goodacre & Kell, 1993), highlight the need to address the long-term reproducibility problems found in the present study whereby networks trained using data from a single batch were unable to recognise pyrolysis mass spectra derived from different batches. Studies on ANN analysis of PyMS data for chemical quantification have yet to address long-term reproducibility problems.

Goodacre *et al.* (1994b) attempted to use the self-organising network (Kohonen, 1989) to classify pyrolysis mass spectra derived from *Propionibacterium acne* strains. The result was promising though there is a need for more comparative studies before this approach can be considered as a standard clustering technique, such as the UPGMA algorithm (Sneath & Sokal, 1973).

<BLANK>

CHAPTER V.
OVERVIEW AND PERSPECTIVES FOR FUTURE WORK

Developments in modern bacterial systematics are dependent on three fundamental processes, namely generation of high quality information, development of appropriate data handling techniques, and compilation and management of taxonomic databases. The focus in recent years has been on the generation of good quality taxonomic information, notably numerical phenetic and nucleotide sequence data. There is now a need to pay greater attention to the other two processes, especially since the generation of genotypic and phenotypic data is becoming increasingly automated and hence relatively straightforward (Felsenstein, 1988; O'Donnell *et al.*, 1993; Kämpfer, 1995). It is also clear that developments in bacterial systematics will increasingly depend upon gaining rapid access to taxonomic information through international electronic networks (Canhos *et al.*, 1993).

Computer software designed for phylogenetic analyses of molecular sequence data are regarded as 'blackboxes' by some taxonomists. It is clear from the review of relevant publications in the recent issue of the International Journal of Systematic Bacteriology (see pages 25-28) that misuse of phylogenetic terms and lack of understanding of the theoretical background on the algorithms used to generate microbial phylogenies are not unusual amongst bacterial systematists. In contrast, the theory and practice of numerical phenetic taxonomy is now relevantly well understood (Gyllenberg, 1965; Sneath & Sokal, 1973; Lapage *et al.*, 1973; Priest & Williams, 1993; Sackin & Jones, 1993; Goodfellow, 1995a). It is also fair to say that the success of numerical phenetic taxonomy was partly based on the availability of computers and on the development of a series of computer programs

(Sneath, 1979b,c,d,e, 1980a,b,c; Sneath & Langham, 1989). It is, therefore, important that suitable and cost-effective software tools are developed to handle information held in taxonomic databases.

The popularity and success of the ribosomal RNA database (Larsen *et al.*, 1993) is a paradigm for future work in microbial systematics. In this database, all of the available small and large subunit rRNA sequences are compiled and constantly released in a ready aligned form. This database project provides a number of additional services which include routines for the determination of rRNA targeted oligonucleotide probes, for detecting chimeric PCR products and for phylogenetic analyses. Other types of taxonomic knowledge bases integrated in global networks are also available, these include ones on strains, nomenclature, amino acid and nucleotide sequences, and chemosystematic data (Canhos *et al.*, 1993). In the near future, global taxonomic databases with high information contents together with appropriate data handling techniques will prove to be invaluable for all microbiologists, not just for taxonomists.

Topics which need further investigation in light of the present study are outlined below.

- ***Software tools***

1. The ***AL16S*** program.
   - Routines for displaying secondary structures.
   - Routines for detecting chimeric molecules and products of possible lateral gene transfer.
2. The ***X*** program
   - Incorporation of the routine ***OUTLIER*** (Sneath & Langham, 1989).
   - Improved strategy for handling multistate data.
   - Routines for artificial neural network analyses, notably the backpropagation, radial-basis function and Kohonen self-organising network algorithms.

- Designing phenotypic databases for the internet.

• *Classification*

- Evaluation of the taxonomic status of *Nocardia pinensis*, *Corynebacterium amycolatum* and *Turicella otitidis* strains by sequencing genes independent of the 16S rRNA macromolecules, notably genes encoding for ATPases, elongation factors and RNA polymerases.

- Polyphasic study to unravel the taxonomic structure of the '*Nocardia asteroides* complex'.

- Chemotaxonomic analyses on biopolymers present in the walls of chemotype IV actinomycetes, notably arabinogalactans and lipopolysaccharides.

• *Identification*

- Monitoring population fluxes of tsukamurellae in activated sludge by *in situ* probing and by the use of artificial neural network analysis of pyrolysis mass spectrometric data.

- Comparative studies need to be carried out to determine the reproducibility of pyrolysis mass spectrometric data derived from different instruments.

<BLANK>

REFERENCES

Adams, A. G. (1969). Areas under the normal curve. *Computer Journal* **12**, 197-198.

Adriaans, B. & Shah, H. (1988). *Fusobacterium ulcerans* sp. nov. from tropical ulcers. *International Journal of Systematic Bacteriology* **38**, 447-448.

Alderson, G. (1985). Nonhierarchic methods. In *Computer-Assisted Bacterial Systematics,* pp. 227-263. M. Goodfellow, D. Jones & F. G. Priest, eds. Academic Press: London.

Alderson, G., Athalye, M. & White, R. P. (1984). Numerical methods in the taxonomy of sporoactinomycetes. In *Biological, Biochemical and Biomedical Aspectes of Actinomycetes,* pp. 597-615. L. Ortiz-Oritiz, L. F. Bajalil & V. Yakoleff, eds. Academic Press: Orlando.

Alshamaony, L., Goodfellow, M. & Minnikin, D. E. (1976). Free mycolic acids as criteria in the classification of *Nocardia* and the '*rhodochrous*' complex. *Journal of General Microbiology* **92**, 188-199.

Amann, R., Ludwig, W. & Schleifer, K. H. (1988). ß-Subunit of ATP-synthetase: a useful marker for studying the phylogenetic relationship of eubacteria. *Journal of General Microbiology* **134**, 2815-2821.

Amann, R., Ludwig, W. & Schleifer, K. H. (1994). Identification of uncultured bacteria: a challenging task for molecular taxonomists. *ASM News* **60**, 360-365.

Amann, R., Ludwig, W. & Schleifer, K. H. (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiological Review* **59**, 143-169.

Amari, S. I. (1967). A theory of adaptive pattern classifier. *IEEE Transactions on Electronic Computers* **EC-16**, 299-307.

Andrews, J. M. & Lieberman, S. H. (1994). Neural-network approach to qualitative identification of fuels and oils from laser-induced fluorescence-spectra. *Analytica Chimica Acta* **285**, 237-246.

Anker, L. S. & Jurs, P. C. (1992). Prediction of carbon-13 nuclear magnetic resonance chemical shifts by artificial neural networks. *Analytical Chemistry* **62**, 1157-1164.

Apajalahti, J. H. A., Kärpänoja, P. & Salkinoja-Salonen, M. S. (1986). *Rhodococcus chlorophenolicus* sp. nov., a chlorophenol-mineralising actinomycete. *International Journal of Systematic Bacteriology* **36**, 246-251.

Aries, R. E., Gutteridge, C. S. & Ottley, T. W. (1986). Evaluation of a low cost, automated pyrolysis mass-spectrometer. *Journal of Analytical and Applied Pyrolysis* **9**, 81-98.

Arnheim, N., Krystal, M., Schmickel, R., Wilson, G., Ryder, O. & Zimmer, E. (1980). Molecular evidence for genetic exchanges among ribosomal genes on nonhomologous chromosomes in man and ape. *Proceedings of National Academic Society USA* **77**, 7323-7327.

Arrigo, P., Giuliano, F., Scalia, F., Rapallo, A. & Damiani, G. (1991). Identification of a new motif on nucleic acid sequence data using Kohonen's self-organising map. *Computer Applications in Biological Sciences* **7**, 353-357.

Asano, K., Sano, H., Masunaga, I. & Kawamoto, I. (1989). 3-*O*-methylrhamnose: identification and distribution in *Catellatospora* species and related actinomycetes. *International Journal of Systematic Bacteriology* **39**, 56-60.

Asturias, J. A., Moore, E., Yakimov, M. M., Klatte, S. & Timmis, K. N. (1994). Reclassification of the polychlorinated biphenyl-degraders *Acinetobacter* sp. strain P6 and *Corynbacterium* sp. strain MB1 as *Rhodococcus globerulus*. *Systematic and Applied Microbiology* **17**, 226-231.

Atalan, E. (1993). *Selective Isolation, Characterisation and Identification of Some Streptomyces Species*. Ph. D. Thesis, University of Newcastle upon Tyne.

Auerbach, S. B., McNeil, M. M., Brown, J. M., Lasker, B. A. & Jarvis, W. R. (1992). Outbreak of pseudoinfection with *Tsukamurella paurometabolum* traced to laboratory contamination: Efficacy of joint epidermiologica and laboratory investigation. *Clinical Infectious Diseases* **14**, 1015-1022.

Austin, B. & Colwell, R. R. (1977). Evaluation of some coefficients for use in numerical taxomony of microorganisms. *International Journal of Systematic Bacteriology* **27**, 204-210.

Bacon, Lord F. (1605). The advancement of learning. Book I Chapter 5. Reprinted 1974, Clarendon Press: Oxford.

Baharaeen, S., Melcher, U. & Vishniac, H. S. (1983). Complementary DNA-23S ribosomal RNA hybridization: an improved method for phylogenetic studies. *Canadian Journal of Microbiology* **29**, 546-551.

Bains, W. (1986). MULTAN: a program for align mutiple DNA sequences. *Nucleic Acid Research* **14**, 159-177.

Balfoort, H. W., Snoek, J., Smits, J. R. M., Breedveld, L. W., Hofstraat, J. W. & Ringelberg, J. (1992). Automatic identification of algae: Neural network analysis of flow cytometric data. *Journal of Plankton Research* **14**, 575-589.

Barrett, S. J. & Sneath, P. H. A. (1994). A numerical phenotypic taxonomic study of the genus *Neisseria*. *Microbiology* **140**, 2867-2891.

Barton, G. J. & Sternberg, M. J. E. (1987). A strategy for the rapid multiple alignment of protein sequences. *Journal of Molecular Biology* **198**, 327-337.

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.

Baylis, H. A. & Bibb, M. J. (1988). Organisation of the ribosomal RNA genes in *Streptomyces coelicolor* A3(2). **211**, 191-196.

Becking, J. H. (1970). *Frankiaceae* fam. nov. (*Actinomycetales*) with one new combination and six new species of the genus *Frankia* Brunchorst 1886, 174. *International Journal of Systematic Bacteriology* **20**, 201-220.

Beers, R. J. & Lockhart, W. R. (1962). Experimental methods in computer taxonomy. *Journal of General Microbiology* **28**, 633-640.

Bello (1989). Computer applications in microbiology: making sense out of data. *ASM News* **55**, 71-73.

Bercovier, H., Kafri, O. & Sela, S. (1986). *Mycobacteria* posses a surprisingly small number of ribosomal RNA genes in relation to the size of their genome. *Biochemical and Biophysical Research Communications* **136**, 1136-1141.

Bergey, D. H., Harrison, F. C., Breed, R. S., Hammer, B. W. & Huntoon, F. M. (1923). *Bergey's Manual of Determinative Bacteriology, Volume 1.* The Williams and Wilkins Company, Baltimore.

Besra, G. S. & Chatterjee, D. (1994). Lipids and carbohydrates of *Mycobacterium tuberculosis*. In *Tuberculosis: Pathogenicity, Protection and Control.* pp. 285-306. B. R. Bloom, ed. American Society for Microbiology, Washington D.C.

BIOLOG (1992). *BIOLOG Microstation: Automated Bacteria Identification System.* BIOLOG Inc., Hayward, USA.

Blackall, L. L. (1991). Use of the Skerman micromanipulator for isolating actinomycetes in the waste water field. *Actinomycetes* **2**, 8-12.

Blackall, L. L. (1994). Molecular identification of activated sludge foaming bacteria. *Water Science and Technology* **29**, 35-42.

Blackall, L. L., Hayward, A. C., Pettigrew, A. E. & Greenfield, P. F. (1985). Biological foam and scum formation in activated sludge treatment systems. *Australian Water and Wastewater Association, 11th Federal Convention*, 338-345.

Blackall, L. L., Harbers, A. E., Greenfield, P. F. & Hayward, A. C. (1989a). Actinomycete scum problems in Australian activated sludge plants. *Water Science and Technology* **20**, 493-495.

Blackall, L. L., Parlett, J. H., Hayward, A. C., Minnikin, D. E., Greenfield, P. F. & Harbers, A. E. (1989b). *Nocardia pinensis* sp. nov., an actinomycete found in activated sludge foams in Australia. *Journal of General Microbiology* **135**, 1547-1558.

Blackall, L. L., Seviour, E. M., Cunningham, M. A., Seviour, R. J. & Hugenholtz, P. (1994). "*Microthrix parvicella*" is a novel, deep branching member of the actinomycetes subphylum. *Systematic and Applied Microbiology* **17**, 513-518.

Blanchard, R. (1896). Parasites végetaux à l'exclusion des bactéries. In *Traité de Pathologie Générale*, vol. II. C. Bouchard, ed. G. Masson: Paris.

Blum, A. (1992). *Neural Networks in C++.* John Wiley & Sons: New York.

Böddinghaus, B., Rogall, T., Flohr, T., Blöcker, H. & Böttger, E. C. (1990). Detection and identification of mycobacteria by amplification of rRNA. *Journal of Clinical Microbiology* **28**, 1751-1759.

Boddy, L. & Morris, C. W. (1993). Neural network analysis of flow cytometry data. In *Flow Cytometry in Microbiology*, pp. 159-169. D. Lloyd ed. Springer-Verlag: London.

Boddy, L., Morris, C. W. & Wimpenny, J. W. T. (1990). Introduction to neural networks. *Binary* 2, 179-185.

Boiron, P., Provost, F. & Duport, B. (1993). *Laboratory Methods for the Diagnosis of Nocardiosis.* Institut Pasteur: Paris.

Boon, J. J., De Boer, W. R., Kruyssen, F. J. & Wouters, J. T. M. (1981). Pyrolysis mass spectrometry of whole cells, cell walls and isolated cell wall polymers of *Bacillus subtilis* var. *niger* WM. *Journal of General Microbiology* **122**, 119-127.

Borst, J., Van der Snee-Enkelaar, A. C. & Meuzelaar, H. L. C. (1978). Typing of *Neisseria gonorrhoeae* by pyrolysis mass spectrometry. *Antonie van Leeuwenhoek* **44**, 253.

Böttger, E. C. (1989). Rapid determination of bacterial ribosomal RNA sequences by direct sequencing of enzymatically amplified DNA. *FEMS Microbiology Letters* **65**, 171-176.

Bousfiled, I, J. & Goodfellow, M. (1976). The '*rhodochrous*' complex and its relationships with allied taxa. In *The Biology of Nocardiae*, pp. 39-65. M. Goodfellow, G. H. Brownell & J. A. Serrano, eds. Academic Press: London.

Branlant, C., Krol, A., Machatt, M. A., Pouyet, J., Ebel, J. P., Edwards, K. & Kössel, (1981). Primary and secondary structures of *Escherichia coli* MRE 600 23S ribosomal RNA. Comparison with models of secondary structure for maize chloroplast 23S rRNA and for large portions of mouse and human 16S mitochondrial rRNAs. *Nucleic Acids Research* **9**, 4303-4321.

Breed, R. S. (1953). The families developed from *Brevibacteriaceae* Cohn with a description on the family *Brevibacteriaceae* Breed, 1953. VI Congresso Internazionale Microbiologia Roma. **1**, 10-15.

Bremer, H. & Dennis, P. P. (1987). Modulation of chemical composition and other parameters of the cell growth rate. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, pp. 1527-1542. F. C. Neidhardt *et al.* eds. American Society for Microbiology, Washington, D.C.

Bridge, P. D. & Sneath, P. H. A. (1983). Numerical taxonomy of *Streptococcus*. *Journal of General Microbiology* **129**, 565-597.

Brierley, J. A. (1978). Thermophilic iron-oxidizing bacteria found in copper leaching dumps. *Applied Environmental Microbiology* **36**, 523-525.

Briglia, M., Eggen, R. I. L., van Elsas, D. J. & De Vos, W. M. (1994). Phylogenetic evidence for transfer of pentachlorophenol-mineralizing *Rhodococcus chlorophenolicus* PCP-I to the genus *Mycobacterium*. *International Journal of Systematic Bacteriology* **44**, 494-498.

Brosius, J., Palmer, M. L., Kennedy, P. J. & Noller, H. F. (1978). Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. *Proceedings of National Academic Society USA* **75**, 4801-4805.

Bryant, T. N. (1991). *Bacterial Identifier: a Utility for Probabilistic Identification of Bacteria*. Blackwell Scientific Publification, Oxford.

Buchanan, R. E. (1917). Studies in the nomenclature and classification of the bacteria. II. The primary subdivisions of the *Schizomycetes*. *Journal of Bacteriology* **2**, 155-164.

Buchanan, R. E. (1918). Studies in the classification and nomenclature of the bacteria. VIII. Subgroups and genera of the *Actinomycetales*. *Journal of Bacteriology* **3**, 403-406.

Cai, J. & Collins, M. D. (1994). Phylogenetic analysis of species of the *meso*-diaminopimelic acid-containing genera *Brevibacterium* and *Dermabacter*. *International Journal of Systematic Bacteriology* **44**, 583-585.

Cain, A. J. & Harrison (1960). Phyletic weighting. *Proceedings inb the Zoological Society of London* **135**, 1-31.

Canhos, V. P., Manfio, G. P. & Blaine, L. D. (1993). Software tools and databases for bacterial systematics and their disseminaztion *via* global networks. *Antonie van Leeuwenhoek* **64**, 205-229.

Carmichael, J. W., Sekhon, A. S. & Sigler, L. (1973). Classification of some dermatophytes by pyrolysis gas-liquid chromatography. *Canadian Journal of Microbiology* **19**, 403-407.

Cartmill, T. D. I., Orr, K., Freeman, R. & Sisson, P. R. (1992). Nosocomial infection with *Clostridium difficile* investigated by pyrolysis mass spectrometry. *Journal of Medical Microbiology* **37**, 352-356.

Casey Klimasauskas, C. C. (1992). Neural networks: an engineering perspective. *IEEE Communications Magazine* **September**, 50-53.

Castellani, A. & Chalmers, A. J. (1913). *Manual of Tropical Medicine*. 2nd ed., Bailliere, Tindall and Cox: London.

Castellani, A. & Chalmers, A. J. (1919). *Manual of Tropical Medicine*. 3rd ed., William, Wood and Co.: New York.

Cavalli-Sforza, L. L. & Edwards, A. W. F. (1967). Phylogenetic analysis: models and estimation procedures. *Evolution* **32**, 550-570.

Chatterjee, D., Hunter, S. W., McNeil, M. & Brennan, P. J. (1992). Lipoarabionomannan, multiglycosylated form of the mycobacterial mannosylphosphatidylinositols. *Journal of Biological Chemistry* **267**, 6228-6233.

Chester, F. D. (1897). Report of the mycologist: bacteriological work. *Delaware Agricultural Experimental Station, Bulletin* **9**:38-145.

Claridge, M. F. & Boddy, L. (1994). Species recognition systems in insects and fungi. In *The identification and characterization of pest organisms*, pp. 261-274. D. L. Hawksworth, ed. CAB International: Wallingford.

Cleran, Y.,Thibault, J., Cheruy, A. & Corrieu, G. (1991). Comparison of prediction performances between models obtained by the group method of data handling and neural networks for the alcoholic fermentation rate in enology. *Journal of Fermentation and Biotechnology* **71**, 356-362.

Clifford, H. T. & Stephenson, W. (1975). *An Introduction to Numerical Classification.* Academic Press: New York.

Codon, C., Philips, J., Fu, Z. Y., Squires, C. & Squires, C. L. (1992). Comparison of the expression of the seven ribosomal RNA operons in *Escherichia coli. EMBO Journal* **11**, 4175-4185.

Cohen, B. I., Presnell, S. R. & Cohen, F. E. (1993). Origins of structural diversity within sequentially identical hexapeptides. *Protein Science* **2**, 2134-2145.

Colasanti, R. L. (1991). Discussions of the possible use of neural network algorithms in ecological modelling. *Binary* **3**, 13-15.

Colless, D. H. (1970). The phenogram as an estimate of phylogeny. *Systematic Zoology* **19**, 352-362.

Collins, M. A. (1993). Practical applications of neural nets in fermenter control. *Binary* **5**, 11-16.

Collins, M .D. (1994). Isoprenoid quinones. In *Chemical methods in Prokaryotic Systematics*, pp. 265-309. M. Goodfellow & A. G. O'Donnell, eds. John Wiley and Sons: Chichester.

Collins, M. D. & Jones, D. (1982). Lipid composition of *Corynebacterium paurometabolum* Steinhaus. *FEMS Microbiology Letters* **13**, 13-16.

Collins, M. D. & Stackebrandt, E. (1989). Molecular taxonomic studies on some LL-diaminopimelic acid-containing coryneforms from herbage: description of *Nocardioides fastidiosa* sp. nov. *FEMS Microbiology Letters* **57**, 289-293.

Collins, M. D. & Wallbanks, S. (1992). Comparative sequence analyses of the 16S rRNA genes of *Lactobacillus minutus, Lactobacillus rimae* and *Streptococcus parvulus*: proposal for the creation of a new genus *Atopobium. FEMS Microbiology Letters* **95**, 235-240.

Collins, M. D., Pirouz, T., Goodfellow, M. & Minnikin, D. E. (1977). Distribution of menaquinones in actinomycetes and corynebacteria. *Journal of General Microbiology* **100**, 221-230.

Collins, M. D., Goodfellow, M. & Minnikin, D. E. (1982). A survey of the structures of mycolic acid in *Corynebacterium* and related taxa. *Journal of General Microbiology* **128**, 129-149.

Collins, M. D., Goodfellow, M., Minnikin, D. E. & Alderson, G. (1985). Menaquinone composition of mycolic acid-containing actinomycetes and some sporoactinomycetes. *Journal of Applied Bacteriology* **58**, 77-86.

Collins, M. D., Howarth, o. W., Grund, E. & Kroppenstedt, R. M. (1987). Isolation and structural determination of new members of the vitamin K2 series in *Nocardia brasiliensis. FEMS Microbiology Letters* **41**, 35-39.

Collins, M. D., Smida, J., Dorsch, M. & Stackebrandt, E. (1988a). *Tsukamurella* gen. nov. harboring *Corynebacterium paurometabolum* and *Rhodococcus aurantiacus. International Journal of Systematic Bacteriology* **38**, 385-391.

Collins, M. D., Burton, R. A. & Jones, D. (1988b). *Corynebacterium amycolatum* sp. nov. a new mycolic acid-less *Corynebacterium* species from human skin. *FEMS Microbiology Letters* **49**, 349-352.

Collins, M. D., Kroppenstedt, R. M., Tamaoka, J., Komagata, K. & Kinoshita, T. (1988c). Structures of the tetrahydrogenated menaquinones from *Actinomadura angiospora*, *Faenia rectivirgula* and *Saccharothrix australiensis*. *Current Microbiology* **17**, 275-279.

Collins, M. D., Dorsch, M. & Stackebrandt, E. (1989a). Transfer of *Pimelobacter tumescens* to *Terrabacter* gen. nov. as *Terrabacter tumescens* comb. nov. and of *Pimelobacter jensenii* to *Nocardioides* as *Nocardioides jensenii* comb. nov. *International Journal of Systematic Bacteriology* **39**, 1-6.

Collins, M. D., Smida, J. & Stackebrandt, E. (1989b). Phylogenetic evidence for the transfer of *Caseobacter polymorphus* (Crombach) to the genus *Corynebacterium*. *International Journal of Systematic Bacteriology* **39**, 7-9.

Collins, M. D., Cockcroft, S. & Wallbanks, S. (1994a). Phylogenetic analysis of a new LL-diaminopimelic acid-containing coryneform bacterium from herbage, *Nocardioides plantarum* sp. nov. *International Journal of Systematic Bacteriology* **44**, 523-526.

Collins, M. D., Love, D. N., Karjalainen, J., Kanervo, A., Forsblom, B., Willens, A., Stubbs, S., Sarkiala, E., Bailey, G. D., Wigney, D. I. & Jousimies-Somer, H. (1994b). Phylogenetic analysis of members of the genus *Porphyromonas* and description of *Porphyromonas cangingivalis* sp. nov. and *Porphyromonas cansulci* sp. nov. *International Journal of Systematic Bacteriology* **44**, 674-679.

Collins, M. D., Lawson, P. A., Willems, A., Cordoba, J. J., Fernamdez-Garayzabal, J., Gercia, P., Cai, J., Hippe, H. & Farrow, J. A. E. (1994c). The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. *International Journal of Systematic Bacteriology* **44**, 812-826.

Colloch, N., Etchebest, C., Thoreau, E., Henrissat, B. & Mornon, J. P. (1993). Comparison of 3 algorithms for the assignment of secondary structure in proteins - the advantages of a consensus assignment. *Protein Engineering* **6**, 377-382.

Colwell, R. R. (1970). Polyphasic taxonomy of bacteria. In *Culture collections of microorganisms, pp. 421-436. H. Iizuka & T. Hasegawa, eds.* University Park Press: Baltimore.

Consden, R. & Gordon, A. H. (1948). Effect of salt on partition chromatograms. *Nature (London)* **162**, 180-181.

Corkill, J. E., Sisson, P. R., Smyth, A., Deveney, J., Freeman, R., Shears, P., Heaf, D. & Hart, C. A. (1994). Application of pyrolysis mass spectrometry and SDS-PAGE in the study of the epimediology of *Pseudomonas cepacia* in cystic fibrosis. *Journal of Medical Microbiology* **41**, 106-111.

Couch, J. N. (1955). A new genus and family of the *Actinomycetales*, with a revision of the genus *Actinoplanes*. *Journal of Elisha Mitchell Scientific Society* **71**, 148-155.

Cowan, S. T. (1978). *A Dictionary of Microbial Taxonomy.* Cambridge University Press: Cambridge.

Crick, F. (1989). The recent excitement about neural networks. *Nature* **337**, 129-132.

Cross, T. & Goodfellow, M. (1973). Taxonomy and classification of the actinomycetes. In *Actinomycetales: Characteristics and Practical Importance,* pp. 11-112. Academic Press: London.

Culverhouse, P. F., Ellis, R., Simpson, R. G., Williams, R., Pierce, R. W. & Turner, J. T. (1994). Automatic categorization of 5 species of *Cymatocylis* (Protozoa, Tintinnida) by artificial neural network. *Marine Ecology-Progress Series* **107**, 273-280.

Cummins, C. S. & Harris, H. (1956). The chemical composition of the cell wall in some Gram-positive bacteria and its possible value as a taxonomic character. *Journal of General Microbiology* **14**, 583-600.

Daffé, M., Brennan, P. J. & McNeil, M. (1990). Predominantal structural features of the cell wall arabinogalactan of *Mycobacterium tuberculosis* as revealed through characterization of oligoglycosyl alditol fragments by gas chromatography/mass spectrometry and by ^{1}H and ^{13}C NMR analyses. *Journal of Biological Chemistry* **265**, 6734-6743.

Daffé, M., McNeil, M. & Brennan, P. J. (1993). Major structural features of the cell wall arabinogalactans of *Mycobacterium*, *Rhodococcus*, and *Nocardia*. **249**, 383-398.

Dams, E., Yamada, T., De Baere, R., Huysamans, E., Vandenberghe, A. & De Wachter, R. (1987). Structure of 5S rRNA in actinomycetes and relatives and evolution of eubacteria. *Journal of Molecular Evolution* **25**, 255-260.

Davison, W. H. T., Slaney, S. & Wragg, A. L. (1954). A novel method of identification of polymers. *Chem. Ind.* 1356.

De Boer, L., Dijkhuizen, L., Grobben, G., Goodfellow, M., Stackebrandt, E, Parlett, J. H., Whitehead, D. & Witt, D. (1990). *Amycolatopsis methanolica* sp. nov., a facultatively methylotrophic actinomycete. *International Journal of Systematic Bacteriology* **40**, 194-204.

De Ley, J. & De Smedt, J. (1975). Improvement of the membrane filter method for DNA:rRNA hybridization. *Antonie van Leeuwenhoek* **41**, 287-307.

De Smedt, J. & De Ley, J. (1977). Intra- and intergeneric similarities of *Agrobacterium* ribosomal ribonucleic acid cistrons. *International Journal of Systematic Bacteriology* **27**, 222-240.

De Soete, G. (1983). A least squares algorithm for fitting additive trees to proximity data. *Psychometrica* **48**, 621-626.

De Soete, G. (1986). Optimal weighting for ultrametric and additive tree clustering. *Quality and Quantity* **20**, 169-180.

De Soete, G. (1988). OVWRTE: A program for optimal variable weighting for ultrametric and additive tree fitting. *Journal of Classification* **5**, 101-104.

De Vos, P., Van Landshoot, A., Segers, P., Tijtgat, R., Gillis, M., Bauwens, M., Rossau, R., Goor, M., Pot, B., Kersters, K., Lizzaraga, P. & De Ley, J. (1989). Genotypic relationships and taxonomic localization of unclassified *Pseudomonas* and *Pseudomonas*-like strains by deoxyribonucleic acid:ribosomal ribonucleic acid hybridizations. *International Journal of Systematic Bacteriology* **39**, 35-49.

DeBorde, D. C., Naeve, C. W., Herlocher, M. L. & Maassab, H. F. (1986). Resolution of a common RNA sequencing ambiguity by terminal deoxynucleotidyl transferase. *Analytical Biochemistry* **157**, 275-282.

DeBry, R. W. & Slade (1985). Cladistic analysis of restriction endonuclease cleavage maps within a maximum-likelihood framework. *Systematic Zoology* **34**, 21-34.

Deiko, S., Yamasaki, R., Jidoi, J., Hori, H. & Osawa, S. (1984). Secondary structure and phylogeny of *Staphylococcus* and *Micrococcus* 5S rRNAs. *Journal of Bacteriology* **159**, 233-237.

Demaree, J. B. & Smith, N. R. (1952). *Nocardia vaccinii* n. sp. causing galls on blueberry plants. *Phytopathology* **42**, 249-252.

Demeler, B. & Zhou, G. (1991). Neural network optimization for *E. coli* promoter prediction. *Nucleic Acids Research* 19, 1593-1599.

Demharter, W., Hensel, R., Smida, J. & Stackebrandt, E. (1989). *Sphaerobacter thermophilus* gen. nov., sp. nov. A deeply rooting member of the actinomycetes subdivision isolated from thermophilically treated sewage sludge. *Systematic and Applied Microbiology* **11**, 261-266.

Dias, F. F. & Bhat, J. V. (1964). Microbial ecology of actiavted sludge. I. Dominant bacteria. *Applied Microbiology* **16**, 1191

Dittmer, J. C. & Lester, R. L. (1964). A simple specific spray for the detection of phospholipids on thin-layer chromatography. *Journal of Lipid Research* **5**, 126-127.

Doi, R. H. & Igarashi, R. T. (1965). Conservation of ribosomal and messenger ribonucleic acid cistrons in *Bacillus* species. *Journal of Bacteriology* **90**, 384-390.

Dolzani, L., Tonin, E., Lagatolla, C. & Monti-Bragadin, C. (1994). Typing of *Staphylococcus aureus* by amplification of the 16S-23S rRNA intergenic spacer sequences. *FEMS Microbiology Letters* **119**, 167-174.

Donnison, A. M., Gutteridge, C. S., Norris, J. R., Morgan, H. W. & Daniel, R. (1986). A preliminary grouping of New Zealand *Thermus* strains by pyrolysis mass spectrometry. *Journal of Analytical and Applied Pyrolysis* **9**, 281-285.

Dopazo, J. (1994). Estimating errors and confidence intervals for branch lengths in phylogenetic trees by a bootstrap approach. *Journal of Molecular Evolution* **38**, 300-304.

Dott, W. & Kämpfer, P. (1988). Biochemical methods for automated bacterial identification and testing metabolic activities in water and wastewater. *Water Science and Technology* **20**, 221-227.

Dover, G. (1982a). A molecular drive through evolutiuon. *BioScience* **32**, 526-533.

Dover, G. (1982b). Molecular drive: a cohesive mode of species evolution. *Nature* **299**, 111-117.

Dover, G. A. (1987). DNA turnover and the molecular clock. *Journal of Molecular Evolution* **26**, 47-58.

Drancourt, M., McNeil, M. M., Brown, J. M., Lasker, B. A., Maurin, M., Choux, M. & Raoult, D. (1994). Brain abscess due to *Gordona terrae* in an immunocompromised child: case report and review of infectious caused by *G. terrae*. *Clinical Infectious Diseases* **19**, 258-262.

Dryden, S. & Kaplan, S. (1990). Localization and structural analysis of the ribosomal RNA operons of *Rhodobacter sphaeroides*. *Nucleic Acids Research* **18**, 7267-7277.

Dubnau, D., Smith, I., Morell, P. & Marmur, J. (1965). Gene conservation in *Bacillus* species. I. Conserved genetic and nucleic acid sequence homologies. *Proceedings of National Academic Society USA* **54**, 491-498.

Duerden, B. I., Eley, A., Goodwin, L., Magee, J. T., Hindmarch, J. M. & Bennett, K. W. (1989). A comparison of *Bacteroides ureolyticus* isolates from different clinical sources. *Journal of Medical Microbiology* **29**, 63-73.

Dunn, G. & Everitt, B. S. (1982). *An Introduction to Mathematical Taxonomy.* Cambridge University Press: Cambridge.

Dunning, A. M., Talmud, P. & Humphries, S. E. (1988). Errors in the polymerase chain reation. *Nucleic Acids Research* **16**, 10393.

Dupuy, N., Willems, A., Pot, B., Dewettinck, D., Vandenbruaene, I., Maestrojuan, G., Dreyfus, B., Kersters, K., Collins, M. D. & Gillis, M. (1994). Phenotypic and genotypic characterization of bradyrhizobia nodulating the leguminous tree *Acacia albida*. *International Journal of Systematic Bacteriology* **44**, 461-473.

East, A. K., Allaway, D. & Collins, M. D. (1992). Analysis of DNA encoding 23S rRNA and 16S-23S rRNA intergenic spacer regions from *Plesiomonas shigelloides*. *FEMS Microbiology Letters* **95**, 57-62.

Eck, R. V. & Dayhoff, M. O. (1966). *Atalas of Protein Sequence and Structure.* National Biomedical Research Foundation: Silver spring, Maryland.

Edwards, A. W. F. & Cavalli-Sforza, L. L. (1964). Reconstruction of evolutionary trees. In *Phenetic and Phylogenetic Classification*, pp. 67-76. V. H. Heywood & J. McNeill, eds. Systematics Association Volume No. 6. Systematics Association: London.

Efron, B. & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *Am. Stat.* **37**, 36-48.

Eguchi, T., Takada, N., Nakamura, S., Tanaka, T., Makino, T. & Oshima, Y. (1993). *Streptomyces bungoensis* sp. nov. *International Journal of Systematic Bacteriology* **43**, 794-798.

Eikelboom, D. H. (1994). The *Microthrix parvicella* puzzle. *Water Science and Technology* **29**, 271-279.

Embley, T. M. (1991). The linear PCR reaction: a simple and robust method for sequencing amplified rDNA genes. *Letters in Applied Microbiology* **13**, 171-174.

Embley, T. M. & Stackebrandt, E. (1994). The molecular phylogeny and systematics of the actinomycetes. *Annual Reviews of Microbiology* **48**, 257-289.

Embley, T. M. & Wait, R. (1994). Structural lipids of eubacteria. In *Chemical methods in Prokaryotic Systematics*, pp. 121-161. M. Goodfellow & A. G. O'Donnell eds. Chichester: John Wiley and Sons.

Embley, T. M., Goodfellow, M., Minnikin, D. E. & Austin, B. (1983). Fatty acids, isoprenoid quinone and polar lipid composition of *Renibacterium salmoninarum*. *Journal of Applied Bacteriology* **55**, 37.

Embley, T. M., Smida, J. & Stackebrandt, E. (1988a). Reverse transcriptase sequencing of 16S ribosomal RNA from *Faenia rectivirgula*, *Pseudonocardia thermophila* and *Saccharopolyspora hirsuta*, three wall type IV actinomycetes which lack mycolic acids. *Journal of General Microbiology* **134**, 961-966.

Embley, T. M., Smida, J. & Stackebrandt, E. (1988b). The phylogeny of mycolate-less wall chemotype IV actinomycetes and description of *Pseudonocardiaceae* fam. nov. *Systematic and Applied Microbiology* **11**, 44-52.

Enea, V. & Corredor, V. (1991). The evolution of plasmodial stage specific rRNA genes is dominated by gene conversion. *Journal of Molecular Evolution* **32**, 183-186.

Enright, M. C., Carter, P. E., MacLean, I. A. & McKenzie, H. (1994). Phylogenetic relationships between some members of the genera *Neisseria, Acinetobacter, Moraxella,* and *Kingella* based on partial 16S ribosomal DNA sequence analysis. *International Journal of Systematic Bacteriology* **44**, 387-391.

Ensign, J. C. (1992). Introduction to actinomycetes. In *The Prokaryotes*, 2nd edition, pp. 811-815. A. Balows, H. G. Trüper, M. Dworkin, W. Harder and K. H. Schleifer, eds. Springer-Verlag, New York.

Eppinger, H. (1891). Über eine neue pathogene *Cladothrix* und eine durch sie hervorgerufene Pseudotuberculosis (Cladothrichia). *Beitrage zur pathologischen Anatomie* **9**, 287-328.

Eshuis, W., Kistemaker, P. G. & Meuzelaar, H. L. C. (1977). Some numerical aspects of reproducibility and specificity. In *Analytical Pyrolysis*, pp. 151-166. C. E. R. Jones & C. A. Cramers, eds. Elsevier, Amsterdam.

Eudy, L. W., Walla, M. D., Hudson, J. R., Morgan, S. L. & Fox, A. (1985). Gas chromatography-mass spectrometry studies on the occurrence of acetamide, propionamide, and furfuryl alcohol in pyrolysates of bacteria, bacterial fractions and model compouds. *Journal of Analytical and Applied Pyrolysis* **7**, 231-247.

Everitt, B. S. (1980). *Cluster Analysis*, 2nd edition. Heinemann, London.

Ezaki, T., Li, N., Hashimoto, Y., Miura, H. & Yamamoto, H. (1994). 16S ribosomal DNA sequences of anaerobic cocci and proposal of *Ruminococcus hansenii* comb. nov. and *Ruminococcus productus* comb. nov. *International Journal of Systematic Bacteriology* **44**, 130-136.

Farris, F. S. (1970). Methods for computing Wagner trees. *Systematic Zoology* **19**, 83-92.

Farris, F. S. (1977). Phylogenetic analysis under Dollo's law. *Systematic Zoology* **26**, 77-88.

Felsenstein, J. (1978a). The number of evolutionary trees. *Journal of Molecular Evolution* **17**, 368-376.

Felsenstein, J. (1978b). Cases in which parsiomony and compatibility methods will be positively misleading. *Systematic Zoology* **27**, 410-410.

Felsenstein, J. (1981a). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* **17**, 368-376.

Felsenstein, J. (1981b). Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. *Evolution* **35**, 1229-1242.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783-791.

Felsenstein, J. (1988). Phylogenies from molecular sequences: inference and reliability. *Annual Reviews of Genetics* **22**, 521-565.

Felsenstein, J. (1993). *PHYLIP (Phylogeny Inference Package),* version 3.5c. University of Washington.

Felsenstein, J. & Kishino, H. (1993). Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. *Systematic Biology* **42**, 193-200.

Feng, D. F. & Doolittle, R. F. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *Journal of Molecular Evolution* **25**, 351-360.

Fessenden, R. J. & Gyorgyi, L. (1991). Identifying functional groups in IR spectra using an artificial neural network. *Journal of Chemical Society Perkin Transection 2* 1755-1762.

Fitch, W. M. (1971). Toward defining the course of evolution: minimal change for a specific tree topology. *Systematic Zoology* **20**, 406-416.

Fitch, W. M. & Margoliash, E. (1967). Construction of phylogenetic trees. *Science* **155**, 279-284.

Fowler, V. J., Ludwig, W. & Stackebrandt, E. (1985). Ribosomal ribonucleic acid cataloguing in bacterial systematics: the phylogeny of *Actinomadura.* In *Chemical Methods in Bacterial Systematics,* pp. 17-40. M. Goodfellow & D. E. Minnikin, eds. Academic Press: London.

Fox, G. E. & Stackebrandt, E. (1987). The application of 16S rRNA sequencing to bacterial systematics. *Methods in Microbiology* **19**, 406-458.

Fox, G. E., Pechman, K. R. & Woese, C. R. (1977). Comparative cataloging of 16S ribosomal ribonucleic acid: molecular approach to prokaryotic systematics. *International Journal of Systematic Bacteriology* **27**, 44-57.

Fox, G. E., Wisotzkey, J. D. & Jurtshuk, P., Jr. (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *International Journal of Systematic Bacteriology* **42**, 166-170.

Freeman, J. A. & Skapura, D. M. (1991). *Neural Networks: Algorithms, Applications, and Programming Techniques.* Addison-Wesley Publishing Company: Reading, MA.

Freeman, R., Goodfellow, M., Gould, F. K., Hudson, S. J. & Lightfoot, N. F. (1990). Pyrolysis mass spectrometry for the rapid epidemiological typing of clinically significant bacterial pathogens. *Journal of Medical Microbiology* **32**, 283-286.

Freeman, R., Sisson, P. R., Lightfoot, N. F. & McLauchlin, J. (1991a). Analysis of epidemic and sporadic strains of *Listeria monocytogenes* by pyrolysis mass spectrometry. *Letters in Applied Microbiology* **12**, 133-136.

Freeman, R., Gould, F. K., Sisson, P. R. & Lightfoot, N. F. (1991b). Strain differentiation of capsule type 23 penicillin-resistant *Streptococcus pneumoniae* from nosocomial infections by pyrolysis mass spectrometry. *Letters in Applied Microbiology* **13**, 28-31.

Freeman, R., Goodacre, R., Sisson, P. R., Magee, J. G., Ward, A. C. & Lightfoot, N. F. (1994a). Rapid identification of species within the *Mycobacterium tuberculosis* complex by artificial neural network analysis of pyrolysis mass spectra. *Journal of Medical Microbiology* **40**, 170-173.

Freeman, R., Sisson, P. R., Ward, A. C. & Hetherington, C. (1994b). Phenotypic variation as a significant within-strain factor in comparisons of bacteria based on pyrolysis-mass spectrometry. *Journal of Analytical and Applied Pyrolysis* **28**, 29-37.

French, G. I., Gutteridge, C. S. & Phillips, I. (1980). Pyrolysis gas chromatography of *Pseudomonas* and *Acenitobacter* species. *Journal of Applied Bacteriology* **49**, 505-516.

French, G. I., Talsania, H. & Philips, I. (1989). Identification of viridnas streptococci by pyrolysis-gas chromatography. *Journal of Medical Microbiology* **29**, 19-27.

Friedman, C. S. & Hedrick, R. P. (1991). Pacific oyster nocardiosis: isolation of the bacterium and induction of laboratory infections. *Journal of Invertebrate Pathology* **57**, 109-120.

Frothingham, R. & Wilson, K. H. (1993). Sequence-based differentiation of strains in the *Mycobacterium avium* complex. *Journal of Bacteriology* **175**, 2818-2825.

Fu, C. S. & Poch, M. (1995). System-identification and real-time pattern-recognition by neural networks for an activated-sludge process. *Environment International* **21**, 57-69.

Fuhrman, J. A., McCallum, K. & Davis, A. A. (1993). Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific oceans. *Applied Environmental Microbiology* **59**, 1294-1302.

Fukami, K. & Tateno, Y. (1989). On the maximum likelihood method for estimating molecular trees: uniqueness of the likelihood point. *Journal of Molecular Evolution* **28**, 460-464.

Funke, G., Stubbs, S., Altwegg, M., Carlotti, A. & Collins, M. D. (1994). *Turicella otitidis* gen. nov., sp. nov., a coryneform bacterium isolated from patients with otitis media. *International Journal of Systematic Bacteriology* **44**, 270-273.

Garland, J. L. & Mills, A. L. (1991). Classification and characterization of heterotrophic microbial communities on the basis of patterns of community-level sole-carbon-source utilization. *Applied Environmental Microbiology* **57**, 2351-2359.

Gillespie, D. & Spiegelman, S. (1965). A quantitative assay for DNA-RNA hybrids with DNA immobilized on a membrane. *Journal of Molecular Biology* **12**, 829-842.

Gillis, M. & De Ley, J. (1980). Intra- and intergeneric similarities of the ribosomal ribonucleic acid cistrons of *Acetobacter* and *Gluconobacter*. *International Journal of Systematic Bacteriology* **30**, 7-27.

Gilmour, J. S. L. (1937). A taxonomic problem. *Nature* **139**, 1040-1042.

Glassey, J., Montague, G. A., Ward, A. C. & Kara, B. V. (1994a). Enhanced supervision of recombinant *E. coli* fermentations *via* artificial neural network. *Process Biochemistry* **29**, 387-398.

Glassey, J., Montague, G. A., Ward, A. C. & Kara, B. V. (1994b). Artificial neural network based experimental design procedures for enhancing fermentation development. *Biotechnology and Bioengineering* **44**, 397-405.

Gogarten, J. P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E. J., Bowman, B. J., Manolson, M. F., Poole, R. J., Date, T., Oshima, T., Konishi, J., Denda, K. & Yoshida, M. (1989). Evolution of the vacuolar H +-ATPase: implications for the origin of eukaryotes. *Proceedings of National Academic Society USA* **86**, 6661-6665.

Gojobori, T., Li, W. H. & Graur, D. (1982). Patterns of nucleotide substitution in pseudogenes and functional genes. *Journal of Molecular Evolution* **18**, 360-369.

Good, R. C. (1992). The genus *Mycobacterium*-Medical. In *The Prokaryotes*, 2nd edition, pp. 1238-1270. A. Balows, H. G. Trüper, M. Dworkin, W. Harder and K. H. Schleifer, eds. Springer-Verlag, New York.

Goodacre, R. & Berkeley, R. C. W. (1990). Detection of small genotypic changes in *Escherichia coli* by pyrolysis mass spectrometry. *FEMS Microbiology Letters* **71**, 133-138.

Goodacre, R. & Kell, D. B. (1993). Rapid and quantitative analysis of bioprocesses using pyrolysis mass spectrometry and neural networks: application to indole production. *Analytica Chimica Acta* **279**, 17-26.

Goodacre, R., Hartmann, A., Beringer, J. E. & Berkeley, R. C. W. (1991). The use of pyrolysis mass spectrometry in the characterization of *Rhizobium meliloti*. *Letters in Applied Microbiology* **13**, 157-160.

Goodacre, R., Kell, D. B. & Bianchi, G. (1992). Neural networks and olive oil. *Nature* **359**, 594.

Goodacre, R., Edmonds, A.N. & Kell, D.B. (1993a). Quantitative analysis of the pyrolysis-mass spectra of complex mixtures using artificial neural networks: Application to amino acids in glycogen. *Journal of Analytical and Applied Pyrolysis* **26**, 93-114.

Goodacre, R., Kell, D. B. & Bianchi, G. (1993b). Rapid assessment of virgin olive oils and other seed oils using pyrolysis mass spectrometry and artificial neural networks. *Journal of Science of Food and Agriculture* **63**, 297-307.

Goodacre, R., Neal, M. J. & Kell, D. B. (1994a). Rapid and quantitative analysis of the pyrolysis mass spectra of complex binary and tertiary mixtures using multivariate calibration and artificial neural networks. *Analytical Chemistry* **66**, 1070-1085.

Goodacre, R., Neal, M. J., Kell, D. B., Greenham, L. W., Noble, W. C. & Harvey, R. G. (1994b). Rapid identification using pyrolysis mass spectrometry and artificial neural networks of *Propionibacterium acnes* isolated from dogs. *Journal of Applied Bacteriology* **76**, 124-134.

Goodacre, R., Trew, S., Wrigley-Jones, C., Neal, M. J., Maddock, J., Ottley, T. W., Porter, N. & Kell, D. B. (1994c). Rapid screening for metabolite overproduction in fermentor broths, using pyrolysis mass spectrometry with multivariate calibration and artificial neural networks. *Biotechnology and Bioengineering* **44**, 1205-1216.

Goodfellow, M. (1971). Numerical taxonomy of some nocardioform bacteria. *Journal of General Microbiology* **69**, 33-80.

Goodfellow, M. (1977). Numerical taxonomy. In *CRC Handbook of Microbiology*, vol. 1, pp. 579-597. A. I. Laskin & H. A. Lechevalier, eds. CRC Press: Ohio.

Goodfellow, M. (1984). Reclassification of *Corynebacterium fascians* (Tilford) Dowson in the genus *Rhodococcus* as *Rhodococcus fascians* comb. nov. *Systematic and Applied Microbiology* **5**, 225-229.

Goodfellow, M. (1989a). Suprageneric classification of actinomycetes. In *Bergey's Manual of Systematic Bacteriology*, pp. 2333-2339. S. T. Williams, M. E. Sharpe & J. G. Holt, eds. Williams and Wilkins: Baltimore.

Goodfellow, M. (1989b). The genus *Rhodococcus* Zopf 1891, 28^{AL}. In *Bergey's Manual of Systematic Bacteriology*, pp. 2362-2371. S. T. Williams, M. E. Sharpe & J. G. Holt, eds. Williams and Wilkins: Baltimore.

Goodfellow, M. (1992). The family *Nocardiaceae.* In *The Prokaryotes*, 2nd edition, pp. 1188-1213. A. Balows, H. G. Trüper, M. Dworkin, W. Harder and K. H. Schleifer, eds. Springer-Verlag, New York.

Goodfellow, M. (1995a). Numerical taxonomy. In *Biology of the prokaryotes.* H. G. Schlegel, & J. W. Lengeler. Georg Thieme Verlag, Sttutgart (in press).

Goodfellow, M. (1995b). Actinomycetes: *Actinomyces, Actinomadura, Nocardia, Streptomyces* and related genera. In *Mackie and McCarney Practical Medical Microbiology*, J. G. Collee, J. P. Duguid, A. G. Fraser, B. P. Marmin & A. Simmons, eds. 14th edition, Churchill Livingstone, Edinburgh (In press).

Goodfellow, M. (1995c). Inter-strain comparison of pathogenic microorganisms by pyrolysis mass spectrometry. *Binary* **7**, 54-59.

Goodfellow, M. & Alderson, G. (1977). The actinomycete-genus *Rhodococcus*: a home for the '*rhodochrous*' complex. *Journal of General Microbiology* **100**, 99-122.

Goodfellow, M. & Cross, T. (1974). Actinomycetes. In *Biology of Plant Litter Decomposition.* pp. 269-302. C. H. Dickinson & G. J. F. Pugh, eds. Academic Press: London.

Goodfellow, M. & Cross, T. (1984). Classification. In *The Biology of the Actinomycetes*, pp. 7-164. M. Goodfellow, M. Mordarski & S. T. Williams, eds. Academic Press: London.

Goodfellow, M. & Dickinson, C. H. (1985). Delineation and description of microbial populations using numerical methods. In *Computer-Assisted Bacterial Systematics,* pp. 165-225. M. Goodfellow, D. Jones & F. G. Priest, eds. Academic Press: London.

Goodfellow, M. & James, A. L. (1994). Rapid enzyme tests in the characterization and identification of microorganisms. In *The Identification and Characterization of Pest Organisms*, pp. 289-301. D. L. Hawksworth, ed. CAB International: Wallingford.

Goodfellow, M. & Lechevalier, M. P. (1989). The genus *Nocardia* Trevisan 1889, 9^{AL}. In *Bergey's Manual of Systematic Bacteriology*, pp. 2350-2361. S. T. Williams, M. E. Sharpe & J. G. Holt, eds. Williams and Wilkins: Baltimore.

Goodfellow, M. & Minnikin, D. E. (1980). Definition of the genus *Mycobacterium* vis-á-vis other taxa. In *Twenty-five Years of Mycobacterial Taxonomy.* pp. 115-130. G. P. Kubica, L. G. Wayne & I. S. Good, eds. CDC Press: Atlanta.

Goodfellow, M. & Minnikin, D. E. (1981a). Introduction of coryneform bacteria. In *The Prokaryotes: A Handbook of habitats, Isolation and Identification of Bacteria, Volume II.* pp. 1811-1826. M. P.Star, H. Stolp, H. G. Trüper, A. Balows & H. G. Schlegel, eds. Springer Verlag: Berlin.

Goodfellow, M. & Minnikin, D. E. (1981b). The genera *Nocardia* and *Rhodococcus.* In *The Prokaryotes: A Handbook of habitats, Isolation and Identification of Bacteria, Volume II.* pp. 2016-2027. M. P.Star, H. Stolp, H. G. Trüper, A. Balows & H. G. Schlegel, eds. Springer Verlag: Berlin.

Goodfellow, M. & Minnikin, D. E. (1981c). Classification of nocardioform bacteria. *Zentralblatt für Bakteriologie, Parasitenfunde, Infektionskrankheiten und Hygiene. I. Abteilung Supplement* **11**, 7-16.

Goodfellow, M. & O'Donnell (1993). Roots of Bacterial Systematics. In *Handbook of New Bacterial Systematics*, pp. 3-54. M. Goodfellow & A. G. O'Donnell eds. Academic Press: London.

Goodfellow, M. & O'Donnell, A. G. (1994). Chemosystematics: current state and future prospects. In *Chemical Methods in Prokaryotic Systematics*, pp. 1-20. M. Goodfellow & A. G. O'Donnell, eds. John Wiley & Sons: Chichester.

Goodfellow, M. & Orchard, V. A. (1974). Antibiotic sensitivity of some nocardioform bacteria and its value as a criterion for taxonomy. *Journal of General Microbiology* **83**, 375-387.

Goodfellow, M. & Pirouz, T. (1982). Numerical classification of sporoactinomycetes containing *meso*-diaminopimelic acid in the cell wall. *Journal of General Microbiology* **128**, 503-527.

Goodfellow, M. & Wayne, L. G. (1982). Taxonomy and nomenclature. In *The Biology of the Mycobacteria. Physiology, Identification and Classification*, vol. 1, pp. 471-521. C. Ratledge & J. S. Stanford, eds. Academic Press: London.

Goodfellow, M., Austin, B. & Dickinson, C. H. (1976). Numerical taxonomy of some yellow-pigmented bacteria isolates from plants. *Journal of General Microbiology* **97**, 219-233.

Goodfellow, M., Orlean, P. A. B., Collins, M. D. & Alshamaony, L. (1978). Chemical and numerical taxonomy of strains received as *Gordona aurantiaca*. *Journal of General Microbiology* **109**, 57-68.

Goodfellow, M., Minnikin, D. E., Todd, C., Anderson, G., Minnikin, S. M. & Collins, M. D. (1982a). Numerical and chemical classification of *Nocardia amarae*. *Journal of General Microbiology* **128**, 1283-1297.

Goodfellow, M., Weaver, C. R. & Minnikin, D. E. (1982b). Numerical classification of some rhodococci, corynebacteria and related organisms. *Journal of General Microbiology* **128**, 731-745.

Goodfellow, M., Stanton, L. J., Simpson, K. E. & Minnikin, D. E. (1990). Numerical and chemical classification of *Actinoplanes* and some related actinomycetes. *Journal of General Microbiology* **136**, 19-36.

Goodfellow, M., Zakrzewska-Czerwinska, J., Thomas, E. G., Mordarski, M., Ward, A. C. & James, A. L. (1991). Polyphasic taxonomic study of the genera *Gordona* and *Tsukamurella* including the description of *Tsukamurella wratislaviensis* sp. nov. *Zentralblatt für Bakteriologie* **275**, 162-178.

Goodfellow, M., Chun, J., Stubbs, S. & Toboli, A. S. (1994). Transfer of *Nocardia amarae* Lechevalier and Lechevalier 1974 to the genus *Gordona* as *Gordona amarae* comb. nov. *Letters in Applied Microbiology* **19**, 401-405.

Gordon, R. E. & Mihm, J. E. (1957). A comparative study of some new strains received as nocardiae. *Journal of Bacteriology* **73**, 15-27.

Gordon, R. E. & Mihm, J. E. (1962). Identification of *Nocardia caviae* (Erikson) nov. comb. *Annals of the New York Academy of Sciences* **98**, 628-636.

Gordon, R. E. & Smith, M. M. (1953). Rapidly growing acid fast bacteria. I. Species descriptions of *Mycobacterium phlei* Lehman and Neumann and *Mycobacterium smegmatis* (Trevisan) Lehman and Neumann. *Journal of Bacteriology* **66**, 41.

Gordon, R. E., Barnett, D. A., Handerhan, J. E. & Pang, C. H. N. (1974). *Nocardia coeliaca*, *Nocardia autotrophica*, and the Nocardin strain. *International Journal of Systematic Bacteriology* **24**, 54-63.

Gottlieb, D. (1973). General consideration and implications of the *Actinomycetales*. In *Actinomycetales: Characteristics and Practical Importance*. pp. 1-10. G. Skyes & F. A. Skinner, eds. Academic Press: London.

Gottlieb, D. (1974). *Actinomycetales*. In *Bergey's Manual of Determinative Bacteriology*. pp. 657-659. R. E. Buchnan & N. E. Gibbons, eds. The Williams and Wilkins Company, Baltimore.

Gottlieb, P., LaFauci, G. & Runder, R. (1985). Alterations in the number of rRNA operons within the *Bacillus subtilis* genome. *Gene* **33**, 259-268.

Gould, F. K., Freeman, R., Sisson, P. R., Cookson, B. D. & Lightfoot, N. F. (1991). Inter-strain comparison by pyrolysis mass spectrometry in the investigation of *Staphylococcus aureus* nosocomial infection. *Journal of Hospital Infection* **19**, 41-48.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**, 325-338.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics* **27**, 857-874.

Gray, P. H. H. (1928). The formation of indigotin from indol by soil bacteria. *Proceedings of the Royal Society Series B* **102**, 263-280.

Gray, P. H. H. & Thornton, H. G. (1928). Soil bacteria that decompose certain aromatic compounds. *Zentralblatt für Bakteriologie, Parasitenfunde, Infektionskrankheiten und Hygiene. Abteilung II.* **73**, 74-96.

Gunderson, J. H., Sogin, M. L., Wollett, G., Hollingdale, M., De La Cruz, V. F., Waters, A. P. & McCutchan, T. F. (1987). Structurally distinct, stage-specific ribosomes occur in *Plasmodium. Science* **238**, 933-937.

Gutell, R. R. (1993). Collection of small subunit (16S- and 16S-like) ribosomal RNA structures. *Nucleic Acids Research* **21**, 3051-3054.

Gutell, R. R., Gray, M. W. & Schnare, M. N. (1993). A compliation of large subunit (23S and 23S-like) ribosomal RNA structures: 1993. *Nucleic Acids Research* **21**, 3055-3074.

Gutteridge, C. S. (1987). Characterization of microorganisms by pyrolysis mass spectrometry. *Methods in Microbiology* **19**, 227-272.

Gutteridge, C. S. & Norris, J. R. (1979). The application of pyrolysis techniques to the identification of microorganisms. *Journal of Applied Bacteriology* **47**, 5-43.

Gutteridge, C. S., Vallis, L. & MacFie, H. J. H. (1985). Numerical methods in the classification of micro-organisms by pyrolysis mass spectrometry. In *Computer-Assisted Bacterial Systematics,* pp. 369-401. M. Goodfellow, D. Jones & F. G. Priest, eds. Academic Press: London.

Gyllenberg, H. G. (1963). General method for deriving determination schemes for random collections of microbial isolates. *Ann. Acad. Sci. Fenn., Ser. A, IV Biol.,* no 69, 1-23.

Gyllenberg, H. G. (1965). A model for computer identification of microorganisms. *Journal of General MicroBiology* **39**, 401-405.

Haack, S. K., Garchow, H., Odelson, D. A., Forney, L. & Klug, M. J. (1994). Accuracy, reproducibility, and interpretation of fatty acid methyl ester profiles of model bacterial communities. *Applied Environmental Microbiology* **60**, 2483-2493.

Haas, F. & König, H. (1987). Characterization of an anaerobic symbiont and the associated aerobic bacterial flora of *Pyrrhocoris apterus* (Heteroptera: Pyrrhocoridae). *FEMS Microbiology Ecology* **45**, 99-106.

Haas, F. & König, H. (1988). *Coriobacterium glomerans* gen. nov., sp. nov. from the intestinal tract of the red soldier bug. *International Journal of Systematic Bacteriology* **38**, 382-384.

Haddadin, H. P., Stirland, R. M., Preston, N. W. & Collard, P. (1973). Identification of *Vibrio cholerae* by pyrolysis gas-liquid chromatography. *Applied Microbiology* **25**, 40-43.

Haft, R. F., Shapiro, C. L., Gantz, N. M., Christenson, J. C. & Wallace, R. J., Jr. (1992). Pseudoinfection due to *Tsukamurella paurometabolum. Clinical Infectious Diseases* **15**, 883.

Hägglblom, M. M., Nohynek, L. J., Palleroni, J., Kronqvist, K., Nurmiaho-Lassila, E. L., Salkinoja-Salonen, M. S., Klatte, S. & Kroppenstedt, R. M. (1994). Transfer of polychlorophenol-degrading *Rhodococcus chlorophenolicus* (Apajalahti et al. 1896) to the genus *Mycobacterium* as *Mycobacterium chlorophenolicum* comb, nov. *International Journal of Systematic Bacteriology* **44**, 485-493.

Hahn, D., Lechevalier, M. P., Fischer, A. & Stackebrandt, E. (1989). Evidence for close phylogenetic relationship between members of the genera *Frankia, Geodermatophilus,* and "*Blastococcus*" and emendation of the family *Frankiaceae. Systematic and Applied Microbiology* **11**, 236-242.

Hall, R. M. & Ratledge, C. (1985). Equivalence of mycobactins from *Mycobacterium senegalense, Mycobacterium farcinogenes* and *Mycobacterium fortuitum. Journal of General Microbiology* **131**, 1691-1696.

Hall, R. M. & Ratledge, C. (1986). Distribution and application of mycobactins for the characterization of species within the genus *Rhodococcus. Journal of General Microbiology* **132**, 853-856.

Hamid, M. E. (1994). *Classification and Identification of Actinomycetes Associated with Bovine Farcy.* Ph.D. Thesis, University of Newcastle upon Tyne.

Hamid, M. E., Minnikin, D. E. & Goodfellow, M. (1993). A simple chemical test to distinguish mycobacteria from other mycolic-acid-containing actinomycetes. *Journal of General Microbiology* **139**, 2203-2213.

Hamid, M. E., Chun, J., Magee, J. G., Minnikin, D. E. & Goodfellow, M. (1994). Rapid characterisation and idenitification of *Mycobacterium* using fluorgenic enzyme tests. *Zentralblatt für Bakteriologie* **280**, 476-487.

Hancock, I. C. (1994). Analysis of cell wall constitients of Gram-positive bacteria. In *Chemical methods in Prokaryotic Systematics,* pp. 63-84. M. Goodfellow & A. G. O'Donnell, eds. John Wiley and Sons: Chichester.

Harasawa, R., Mizusawa, H., Nozawa, K., Nakagawa, T., Asada, K. & Kato, I. (1993). Detection and tentative identification of dominant mycoplasma species in cell cultures by restriction analysis of the 16S-23S rRNA intergenic spacer regions. *Research in Microbiology* **144**, 489-493.

Hartford, T. & Sneath, P. H. A. (1988). Distortion of taxonomic structure from DNA relationships due to different choice of reference strains. *Systematic and Applied Microbiology* **10**, 241-250.

Hartford, T. & Sneath, P. H. A. (1993). Optical DNA-DNA homology in the genus *Listeria. International Journal of Systematic Bacteriology* **43**, 26-31.

Hartmans, S. & De Bont, A. M. (1992). The genus *Mycobacterium*-Nonmedical. In *The Prokaryotes,* 2nd edition, pp. 1214-1237. A. Balows, H. G. Trüper, M. Dworkin, W. Harder and K. H. Schleifer, eds. Springer-Verlag, New York.

Hasegawa, M. & Yano, T. (1984). Maximum likelihood method of phylogenetic inference from DNA sequence data. *Bulltin of Biometrical Society of Japan* **5**, 1-7.

Haverkamp, J., Meuzelaar, H. L. C., Beuvery, E. C., Boonekamp, P. M. & Tiesjema, R. H. (1980). Characterization of *Neisseria meningitidis* capsular polysaccharides containing sialic acid by pyrolysis mass spectrometry. *Analytical Biochemistry* **104**, 407-418.

Hebb, D. (1949). *Organization of Behaviour.* John Wiley & Sons: New York.

Hecht-Nielson, R. (1988). Applifications of counterpropagation networks. *Neural Networks* **1**, 131-140.

Hedges, S. B. (1992). The number of replications needed for accurate estimation of the bootstrap P value in phylogenetic studies. *Molecular Biology and Evolution* **9**, 366-369.

Hefferan, M. (1904). A comparative and experimental study of bacilli producing red pigment. *Zentralblatt für Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene. Abteilung II* **11**, 397-404.

Heinonin, T. Y. K., Schnare, M. N. & Gray, M. W. (1990). Sequence heterogeneity in the duplicated large subunit ribosomal RNA genes of *Tetrahymena pyriformis* mitochondrial DNA. *Journal of Biological Chemistry* **265**, 22336-22341.

Helm, D.; Labischinski, H.; Schallehn, G. & Naumann, D. (1991). Classification and identification of bacteria by Fourier-transform infrared spectroscopy. *Journal of General Microbiology* **137**: 69-79.

Helmke, E. & Weyland, H. (1984). *Rhodococcus marinonascens* sp. nov., an actinomycete from the sea. *International Journal of Systematic Bacteriology* **34**, 127-138.

Hendriks, L., Goris, A., Van de Peer, Y., Neefs, J. M., Vancanneyt, M., Kersters, K., Hennebert, G. L. & De Wachter, R. (1991). Phylogenetic analysis of five medically important *Candida* species as deduced on the basis of small ribosomal subunit RNA sequences. *Journal of General Microbiology* **137**, 1223-1230.

Hennecke, H., Kaluza, K., Thöny, B., Fuhrmann, N., Ludwig, W. & Stackebrandt, E. (1985). Concurrent evolution of nitrogenase genes and 16S rRNA in *Rhizobium* species and other nitrogen fixing bacteria. *Archives of Microbiology* **142**, 342-348.

Hernandez, M., Jenkins, D. & Beaman, B. L. (1994). Mass and vaiability estimations of *Nocardia* in activated sludge and anaerobic digesters using conventional stains and immunofluorescent methods. *Water Science and Technology* **29**, 249-259.

Hertz, J., Krogh, A. & Palmer, R.G. (1991). *Introduction to the Theory of Neural Computation.* Addison-Wesley: Redwood City, CA.

Higgins, D. G. & Sharp, P. M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73**, 237-244.

Higgins, D. G. & Sharp, P. M. (1989). Fast and sensitive multiple sequence alignments on a microcomputer. *Computer Applications in Biological Sciences* **5**, 151-153.

Higgins, M. L. & Lechevalier, M. P. (1969). Poorly lytic bacteriophage from *Dactylosporangium thailandensis. Journal of Virology* **3**, 210.

Hill, I. D. (1985). Algorithm AS66: the normal integral. In *Applied Statistics Algorithms.* pp. 126-129. P. Griffiths & I. D. Hill, eds. Royal Statistical Society, Ellis Horwood Ltd.: Chichester.

Hill, L. R. (1974). Theoretical aspects of numerical identificaiton. *International Journal of Systematic Bacteriology* **24**, 494-499.

Hillis, D. M. & Bull, J. J. (1993). An impirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* **42**, 182-192.

Hillis, D. M. & Dixon, M. T. (1991). Ribosomal RNA: molecular evolution and phylogenetic inference. *Quarterly Review of Biolology* **66**, 411-453.

Hillis, D. M., Allard, W. & Miyamoto, M. M. (1993). Analysis of DNA sequence data: phylogenetic inferrence. *Methods in Enzymology* **224**, 456-487.

Hindmarch, J. M., Magee, J. T., Hadfield, M. A. & Duerden, B. I. (1990). A pyrolysis-mass spectrometry study of *Corynebacterium* spp. *Journal of Medical Microbiology* **31**, 137-149.

Hiraishi, A. & Ueda, Y. (1994). *Rhodoplanes* gen. nov., a genus of phototrophic bacteria including *Rhodopseudomonas rosea* as *Rhodoplanes roseus* comb. nov. and *Rhodoplanes elegans* sp. nov. *International Journal of Systematic Bacteriology* **44**, 665-673.

Hirsch, P. (1960). Einige, weitere, von Luftverunreinigungen lebende Actinomyceten und ihre Klassifizierung. *Archiv für Mikrobiologie* **35**, 391-414.

Hirsch, P. (1961). Wasserstoffaktivierrung und Chemoautotrophie bei Actinomyceten. *Archiv für Mikrobiologie* **39**, 360-373.

Hirsch, P. & Engel, H. (1956). Uber oligocarbophile Actinomyceten. *Bericht der Deutschen botanischen Gesellschaft* **69**, 441-454.

Hoefer Scientific Instruments. (1992). Basic techniques and exercises. In *Hoefer Scientific Instruments 1992-1993*, pp. 121-158. . Hoefer Scientific Instruments, San Francisco.

Hori, H. & Osawa, S. (1986). Evolutionary change in 5S rRNA secondary structure and a phylogenetic tree of 352 5S rRNA species. *BioSystems* **19**, 163-172.

Howarth, O. W., Grund, E., Kroppenstedt, R. M. & Collins, M. D. (1986). Structural determination of a new naturally occurring cyclic vitamin K. *Biochemical and Biophysical Research Communication* **140**, 916-923.

Hucker, G. J. & Conn, H. J. (1923). Methods of Gram straining. In *Technical Bulletin of the New York State of Agricultural and Experimental Station* No. 93

Hudson, R., Williams, C., Seviour, R. J. & Soddell, J. A. (1994). Variations in phenotypic characters of type 021N from activated sludge systems. *Water Science and Technology* **29**, 143-147.

Huelsenbeck, J. P. & Hillis, D. M. (1993). Success of phylogenetic methods in the four-taxon case. *Systematic Biology* **42**, 247-264.

Huis I'nt Veld, J. H. J., Meuzelaar, H. L. C. & Tom, A. (1973). Anaysis of streptococcal cell wall fractions by Curie-point pyrolysis gsa-liquid chromatography. *Applied Microbiology* **26**, 92-97.

Huss, V. A. R., Festl, H. & Schleifer, K. H. (1983). Studies on the spectrophotometric determination of DNA hybridization from renaturation rates. *Systematic and Applied Microbiology* **4**, 184-192.

Hyman, I. S. & Chaparas, S. D. (1977). A comparative study of the '*rhodochrous*' complex and related taxa by delayed type skin reactions on guinea pigs and by polyarylamide gel electrophoresis. *Journal of General Microbiology* **100**, 363-371.

Imaeda, T., Broslawski, G. & Imaeda, S. (1988). Genomic relatedness among mycobacterial species by nonisotopic blot hybridization. *International Journal of Systematic Bacteriology* **38**, 151-156.

Irwin, W. J. (1982). *Analytical Pyrolysis: A Comprehensive Guide.* Marcel Dekker: New York.

Iwabe, N., Kuma, K. I., Hasegawa, M., Osawa, S. & Miyata, T. (1989). Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proceedings of National Academic Society USA* **86**, 9355-9359.

Jacin, H. & Mishkin, A. R. (1965). Separation of carbohydrates on borate impregnated silica gel G plates. *Journal of Chromatography* **18**, 170-173.

Jagoueix, S., Bové, J. M. & Garnier, M. (1994). The phleom-limited bacterium of greening disease of citrus is a member of the α subdivision of the *Proteobacteria. International Journal of Systematic Bacteriology* **44**, 379-386.

James, B. D., Olsen, G. J. & Pace, N. R. (1989). Phylogenetic comparative analysis of RNA secondary structure. *Methods in Enzymology* **180**, 227-239.

Jenkins, D., Richard, M. G. & Daigger, G. T. (1993). *Manual on the Causes and Control of Activated Sludge Bulking and Foaming*, 2nd edition. Lewis Publisher: Boca Raton.

Jervis, B. W., Saatchi, M. R., Lacey, A., Roberts, T., Allen, E. M., Hudson, N. R. , Oke, S. & Grimsley, M. (1994). Artificial neural network and spectrum analysis-methods for detecting brain diseases from the CNV response in the electroencephalogram. *IEE Proceedings-Science Measurement and Technology* **141**, 432-440.

Ji, Y. E., Colston, M. J. & Cox, R. A. (1994a). Nucleotide sequence and secondary structures of precursor 16S rRNA of slow-growing mycobacteria. *Microbiology* **140**, 123-132.

Ji, Y. E., Colston, M. J. & Cox, R. A. (1994b). The ribosomal RNA (*rrn*) operons of fast-growing mycobacteria: primary and secondary structures and thier relation to *rrn* operons of pathogenic slow-growers. *Microbiology* **140**, 2829-2840.

Jin, L. & Nei, M. (1990). Limitations of the evolutionary parsimony method of phylogenetic analysis. *Molecular Biology and Evolution* **7**, 82-102.

Johnson, J. L. (1991). DNA reassociation experiments. In *Nucleic Acid Techniques in Bacterial Systematics*, pp. 21-44. E. Stackebrandt & M. Goodfellow, eds. John Wiley & Sons: New York.

Johnson, J. L. & Francis, B. S. (1975). Taxonomy of the clostridia: ribosomal ribonucleic acid homologies among the species. *Journal of General Microbiology* **88**, 229-244.

Johnson, N. L. & Welch, B. L. (1939). Applications of the non-central *t*-distribution. *Biometrika* **31**, 362-389.

Jones, D. (1975). A numerical taxonomic study of coryneform and related taxa. *Journal of General Microbiology* **87**, 52-96.

Jones, D. & Sackin, M. J. (1980). Numerical methods in the classification and identification of bacteria with special reference to *Enterobacteriaceae*. In *Microbiological Classification and Identification*, pp. 73-106. M. Goodfellow & R. G. Board, eds. Academic Press: London.

Jones, K. L. (1949). Fresh isolates of actinomycetes in which the presence of sporogenous aerial mycelia is a fluctuating characteristic. *Journal of Bacteriology* **57**, 141-145.

Jones, R. S., Fekete, T., Truant, A. L. & Satishchandran, V. (1994). Persistent bacteremia due to *Tsukamurella paurometabolum* in a patient undergoing hemodialysis: case report and review. *Clinical Infectious Diseases* **18**, 830-832.

Jukes, T. H. & Cantor, C. R. (1969). Evolution of protein molecules. In *Mammalian Protein Metabolism*, pp. 21-132. H. N. Munro, ed. Academic Press: New York.

Kajioka, R. & Tang, P. W. (1984). Curie-point pyrolysis mass spectrometry of *Legionella* species. *Journal of Analytical and Applied Pyrolysis* **6**, 59-68.

Kämpfer, P. (1995). Automation and miniaturization of physiological tests-some applications in numerical taxonomy and numerical identification. *Binary* **7**, 42-48.

Kämpfer, P. & Kroppenstedt, R. M. (1991). Probabilistic identification of streptomycetes using miniaturized physiological tests. *Journal of General Microbiology* **137**, 1893-1902.

Kämpfer, P., Dott, W. & Kroppenstadt, R. M. (1990). Numerical classification and identification of some nocardioform bacteria. *Journal of General and Applied Microbiology* **36**, 309-331.

Kämpfer, P., Kroppenstedt, R. M. & Dott, W. (1991). A numerical classification of the genera *Streptomyces* and *Streptoverticillium* using miniaturized physiological tests. *Journal of General Microbiology* **137**, 1831-1891.

Kämpfer, P., Kulies, I. & Dott, W. (1992). Fluorgenic substrates for differentiation of gram-negative nonfermentative and oxidase-positive fermentative bacteria. *Journal of Clinical Microbiology* **30**, 1402-1406.

Karlin, S. & Cardon, L. R. (1994). Computational DNA sequence analysis. *Annual Reviews of Microbiology* **48**, 619-654.

Kay, H. E., Coutinho, H. L. C., Fattori, M., Manfio, G. P., Goodacre, R., Nuti, M. P., Basaglia, M. & Beringer, J. E. (1994). The identification of *Bradyrhizobium japonicum* strains isolated from italian soils. *Microbiology* **140**, 2333-2339.

Kemmerling, C., Gürtler, H., Kroppenstedt, R. M., Toalster, R. & Stackebrandt, E. (1993). Evidence for the phylogenetic heterogeneity of the genus *Streptosporangium*. *Systematic and Applied Microbiology* **16**, 369-372.

Kennedy, M. J. & Thakur, M. S. (1993). The use of neural networks to aid in microorganism identification: A case study of *Haemophilus* species identification. *Antonie van Leeuwenhoek* **63**, 35-38.

Kilpper-Bälz, R. (1991). DNA-rRNA hybridization. In *Nucleic acid techniques in bacterial systematics*, pp. 45-68. E. Stackebrandt & M. Goodfellow, eds. John Wiley and Sons: New York.

Kim, H. J. (1993). *Selective Isolation, Characterisation and Identification of Streptosporangia.* Ph. D. Thesis, University of Newcastle upon Tyne.

Kim, J. (1993). Improving the accuracy of phylogenetic estimation by combining different methods. *Systematic Biology* **42**, 331-340.

Kim, S. B. (1995). *Polyphasic Taxonomic Study on the Genus Saccharomonospora and related taxa.* Ph. D. Thesis, University of Newcastle upon Tyne.

Kim, S. B., Yoon, J. H., Kim, H., Lee, S. T., Park, Y. H. & Goodfellow, M. (1995). A phylogenetic analysis of the genus *Saccharomonospora* conducted with 16S rRNA gene sequences. *International Journal of Systematic Bacteriology* **45**, 351-356.

Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**, 111-120.

Kimura, M. (1983). *The Natural Theory of Molecular Evolution.* Cambridge Univeristy Press: Cambridge.

Kirschner, P., Teske, A., Schröder, K. H., Kroppenstedt, R. M., Wolters, J. & Böttger, E. C. (1992). *Mycobacterium confluentis* sp. nov. *International Journal of Systematic Bacteriology* 42, 257-262.

Kishino, H. & Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *Journal of Molecular Evolution* **29**, 170-179.

Kjaer, M. & Poulsen, F. M. (1991). Identification of *2D H'-NMR antiphase cross peaks using* a neural network. *Journal of Magnetic Resonance* **94**, 659-663.

Klatte, S., Jahnke, K. D., Kroppenstedt, R. M., Rainey, F. A. & Stackebrandt, E. (1994a). *Rhodococcus luteus* is a later subjective synonym of *Rhodococcus fascians. International Journal of Systematic Bacteriology* **44**, 627-630.

Klatte, S., Kroppenstedt, R. M. & Rainey, F. A. (1994b). *Rhodococcus opacus* sp. nov., an unusual nutritionally versatile *Rhodococcus*-species. *Systematic and Applied Microbiology* **17**, 355-360.

Klatte, S., Rainey, F. A. & Kroppenstedt, R. M. (1994c). Transfer of *Rhodococcus aichiensis* Tsukamura 1982 and *Nocardia amarae* Lechevalier and Lechevalier 1974 to the genus *Gordona aichiensis* comb. nov. and *Gordona amarae* comb. nov. *International Journal of Systematic Bacteriology* **44**, 769-773.

Klawun, C. & Wilkins, C. L. (1995). Neural network assisted rapid screening of large infrared spectral databases. *Analytical Chemistry* **67**, 374-378.

Klenk, H. P. & Zillig, W. (1994). DNA-dependent RNA polymerase subunit B as a tool for phylohenetic reconstructions: Branching topology of the archeal domain. *Journal of Molecular Evolution* **38**, 420-432.

Klenk, H. P., Haas, B., Schwass, V. & Zillig, W. (1986). Hybridization homology: a new parameter for the analysis of phylogenetic relationships, demonstrated with the urkingdom of the archaebacteria. *Journal of Molecular Evolution* **24**, 167-173.

Klugbauer, N., Ludwig, W., Bäuerlein, E. & Schleifer, K. H. (1992). Subunit β of adenosine triphosphate synthetase of *Pectinatus frisingensis* and *Lactobacillus casei. Systematic and Applied Microbiology* **15**, 323-330.

Kluge, A. G. & Farris, F. S. (1969). Quantitative phyletics and the evolution of anurans. *Systematic Zoology* **18**, 1-32.

Koch, C., Rainey, F. A. & Stackebrandt, E. (1994). 16S rDNA studies on members of *Arthrobacter* and *Micrococcus*: an aid for their future taxonomic restructuring. *FEMS Microbiology Letters* **123**, 167-172.

Kocianova, E., Forster, C. F. & Foot, R. J. (1994). What is *Microthrix parvicella* ?. *Journal of Applied Bacteriology* **76**, 301-306.

Kohonen, T. (1989). Self-Organization and Associative Memory. Springer-Verlag: Berlin.

Korn-Wendisch, F., Rainey, F. A., Kroppenstedt, R. M., Kempf, A., Majazza, A., Kutzner, H. J. & Stackebrandt, E. (1995). *Thermocrispum* gen. nov., a new genus of the order *Actinomycetales*, and description of *Thermocrispum municipale* sp. nov. and *Thermocrispum agreste* sp. nov. *International Journal of Systematic Bacteriology* **45**, 67-77.

Krassilnikov, N. A. (1938). *Ray Fungi and Related Organisms-Actinomycetales.* Akademii Nauk SSSR Moscow.

Krassilnikov, N. A. (1949). *Opredelitelv Bakterii i Actinomicetov.* Akademii Nauk SSSR Moscow.

Kroppenstedt, R. M. (1985). Fatty acid and menaquinone analysis of actinomycetes and related organisms. In *Chemical Methods in Bacterial Systematics*, pp. 173-199. M. Goodfellow & D. E. Minnikin, eds. Academic Press: London.

Kroppenstedt, R. M., Stackebrandt, E. & Goodfellow, M. (1990). Taxonomic revision of the actinomycete genera *Actinomadura* and *Microtetraspora. Systematic and Applied Microbiology* **13**, 148-160.

Kruscal, J. B. (1964a). Multidimensional scaling by optimising goodness of fit to a non-metric hypothesis. *Psychometrika* **29**, 1-27.

Kruscal, J. B. (1964b). A numerical method. *Psychometrika* **29**, 115-129.

Kudo, T., Hatai, K. & Seino, A. (1988). *Nocardia seriolae* sp. nov. causing nocardiosis of cultured fish. *International Journal of Systematic Bacteriology* **38**, 173-178.

Kuhner, M. & Felsenstein, J. (1994). A simulation comparision of phylogeny algorithms under equal and unequal evolutionary rates. *Molecular Biology and Evolution* **11**, 459-468.

Labeda, D. P. (1995). *Amycolatopsis coloradensis* sp. nov., the avoparin (LL-AV290)-producing strain. *International Journal of Systematic Bacteriology* **45**, 124-127.

Labeda, D. P. & Lyons, A. J. (1991). Deoxyribonucleic acid relatedness among species of the "*Streptomyces cyaneus*" cluster. *Systematic and Applied Microbiology* **14**, 158-164.

Labeda, D. P., Testa, R. T., Lechevalier, M. P. & Lechevalier, H. A. (1984). *Saccharothrix*: a new genus of the *Actinomycetales* related to *Nocardiopsis. International Journal of Systematic Bacteriology* **34**, 426-431.

Lai, K. K. (1993). A cancer patient with central venous catheter-related sepsis caused by *Tsukamurella paurometabolum* (*Gordona aurantiaca*). *Clinical Infectious Diseases* **17**, 285-287.

Lake, J. A. (1987a). Prokaryotes and archaebacteria are not monophyletic: rate invariant analysis of rRNA genes indicates that eukaryotes and eocytes from a monophyletic taxon. *Cold Spring Habor Symposium on Quantitative Biology* **52**, 839-846.

Lake, J. A. (1987b). A rate-independent technique for analysis of nucleic acid sequences: evolutionary parsimony. *Molecular Biology and Evolution* **4**, 167-191.

Lamfrom, H., Sarabhai, A. & Abelson, J. (1978). Cloning of *Beneckea* genes in *Escherichia coli. Journal of Bacteriology* **133**, 354-363.

Lane, D. J. (1991). 16S/23S rRNA sequencing. In *Nucleic acid techniques in bacterial systematics*, pp. 115-175. E. Stackebrandt & M. Goodfellow, eds. John Wiley and Sons: New York.

Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L. & Pace, N. R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analysis. *Proceedings of National Academic Society USA* **82**, 6955-6959.

Lane, D. J., Field, K. G., Olsen, G. & Pace, N. R. (1988). Reverse transcriptase sequencing of ribosomal RNA for phylogenetic analysis. *Methods in Enzymology* **167**: 138-144.

Lane, D. J., Harrison, A. P., Stahl, D., Pace, B., Giovannoni, S. J., Olsen, G. J. & Pace, N. R. (1992). Evolutionary relationships among sulfur- and iron-oxidizing eubacteria. *Journal of Bacteriology* **174**, 269-278.

Langham, C. D., Williams, S. T., Sneath, P. H. A. & Mortimer , A. M. (1989a). New probability matrices for identification of *Streptomyces*. *Journal of General Microbiology* **135**, 121-133.

Langham, C. D., Sneath, P. H. A., Williams, S. T. & Mortimer, A. M. (1989b). Detecting aberrant strains in bacterial groups as an aid to constructing databases for computer identification. *Journal of Applied Bacteriology* **66**, 339-352.

Lapage, S. P., Bascomb, S., Willcox, W. R. & Curtis, M. A. (1970). Computer identification of bacteria. In *Automation Mechanisation and Data Handling in Microbiology*, pp. 1-22. A. Baillie & R. J. Gilbert, eds. Academic Press: London.

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.

Larsen, N., Olsen, G. J., Maidak, B. L., McCaughey, M. J., Overbeek, R., Macke, T. J., Marsh, T. L. & Woese, C. R. (1993). The ribosomal database project. *Nucleic Acids Research* **21**, 3021-3023.

Lasker, B. A., Brown, J. M. & McNeil, M. M. (1992). Identification and epidermiological typing of clinical and environmental isolates of the genus Rhodococcus with use of a digoxigenin-labeled rDNA gene probe. Clinical Infectious Diseases 15, 223-233.

Lechevalier, H. A. & Lechevalier, M. P. (1970). A critical evaluation of the genera of aerobic actinomycetes. In *The* Actinomycetales, pp. 393-405. H. Prauser, ed. Veb Gustav Fischer Verlag: Jena.

Lechevalier, H. A., Solotorovsky, M. & McDurmont, C. I. (1961). A new genus of *Actinomycetales*: *Micropolyspora* gen. nov. *Journal of General Microbiology* **26**, 11-18.

Lechevalier, H. A., Lechevalier, M. P., Wyszkowski, P. E. & Mariat, F. (1976). Actinomycetes found in sewage-treatment plants of the activated sludge type. In *Actinomycetes: The Boundary Organisms*, pp. 227-247. T. Arai ed. Toppan Company: Tokyo.

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.

Lechevalier, M. P. & Lechevalier, H. A. (1974). *Nocardia amarae* sp. nov., an actinomycete common in foaming activated sludge. *International Journal of Systematic Bacteriology* **24**, 278-288.

Lechevalier, M. P., De Bièvre, C. & Lechevalier, H. A. (1977). Chemotaxonomy of aerobic actinomycetes: phospholipid composition. *Biochemical Systematics and Ecology* **5**, 249-260.

Lechevalier, M. P., Stern, A. E. & Lechevalier, H. A. (1981). Phospholipids in the taxonomy of actinomycetes. In *Actinomycetes*, pp. 111-116. K. P. Schaal & G. Pulverer Gustav Fischer Verlag: Stuttgart.

Lechevalier, M. P., Prauser, H., Labeda, D. P. & Ruan, J. S. (1986). Two new genera of nocardioform actinomycetes: *Amycolata* gen. nov. and *Amycolatopsis* gen. nov. *International Journal of Systematic Bacteriology* **36**, 29-37.

Lehmann, K. B., & Neumann, R. (1896). *Atlas und Grundriss der Bakteriologie und Lehrbuch der speciellen bakteriologischen Diagnostik*, Teil II, J.F. Lehmann, München.

Lehmann, K. B., & Neumann, R. (1907). *Lehmann,s Medzin, Handatlanten X. Atlas und und Grundriss der Bakteriologie und Lehrbuch der speciellen bakteriologischen Diagnostik. 4 Auflage.* J. F. Lehman: Munchen.

Lemmer, H. & Kroppenstedt, R. M. (1984). Chemotaxonomy and physiology of some actinomycetes isolated from scumming activated sludge. *Systematic and Applied Microbiology* 5, 124-135.

Leonard, J. & Kramer, M. A. (1990). Improvement of the backpropagation algorithm for the training neural networks. *Computers in Chemical Engineering* **14**, 337-341.

Lévy-Frébault, V., Grimont, F., Grimont, P. A. D. & David, H. L. (1986). Deoxyribonucleic acid relatedness study of the *Mycobacterium fortuitum-Mycobacterium chelonae* complex. *International Journal of Systematic Bacteriology* **36**, 458-460.

Lewis, I. R., Daniel, N. W., Chaffin, N. C. & Griffiths, P. R. (1994). Raman-spectrometry and neural networks for the classification of wood types. *Spectrochimica Acta Part A-Molecular Spectroscopy* **50**, 1943-1958.

Li, W. H. & Gouy, M. (1990). Statistical tests of molecular phylogenies. *Methods in Enzymology* **183**, 645-659.

Li, W. H., Wu, C. I. & Luo, C. C. (1984). Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *Journal of Molecular Evolution* **21**, 58-71.

Li, W. H., Wolfe, K. H., Sourdis, J. & Sharp, P. M. (1987). Reconstruction of phylogenetic trees and estimation of divergence times under nonconstant rates of evolution. *Cold Springs Harbor Symp. Quant. Biol.* **52**, 847-856.

Liao, D. & Dennis, P. P. (1994). Molecular phylogenies based on ribosomal protein L11, L1, L10, and L12 sequences. *Journal of Molecular Evolution* **38**, 405-419.

Liebl, W. (1992). The genus *Corynebacterium*-Nonmedical. In *The Prokaryotes*, 2nd edition, pp. 1157-1171. A. Balows, H. G. Trüper, M. Dworkin, W. Harder and K. H. Schleifer, eds. Springer-Verlag, New York.

Liesack, W. & Stackebrandt, E. (1992). Occurrence of novel groups of the domain *Bacteria* as revealed by analysis of genetic material isolated from an Australian terrestrial environment. *Journal of Bacteriology* **174**, 5072-5078.

Liesack, W., Pitulle, C., Sela, S. & Stackerbrandt, E. (1990). Nucleotide sequence of the 16S rRNA from *Mycobacterium leprae*. *Nucleic Acids Research* **18**, 5558-5558.

Liesack, W., Weyland, H. & Stackebrandt, E. (1991). Potential risks of gene amplification by PCR as determined by 16S rDNA analysis of a mixed-culture of strict barophilic bacteria. *Microbial Ecology* **21**, 191-198.

Lind, A. & Ridell, M. (1976). Serological relationships between *Nocardia*, *Mycobacterium*, *Corynebacterium* and the 'rhodochrous' taxon. In *The Biology of Nocardiae*, pp. 220-235. M. Goodfellow, G. H. Brownell & J. A. Serrano, eds. Academic Press: London.

Lindenberg, A. (1909). Un nouveau mycétome. *Archives de Parasitologie* **13**, 265-282.

Liston, J., Weibe, W. & Colwell, R. R. (1963). Quantitative approach to the study of bacterial species. *Journal of Bacteriology* **85**, 1061-1070.

Locci, R. (1981). Micromorphology and development of actinomycetes. *Zentralblatt für Bakteriologie, Parasitenfunde, Infektionskrankheiten und Hygiene. I. Abteilung Supplement* **11**, 119-130.

Locci, R., Williams, S. T., Schofield, G. M., Vickers, J. C., Sneath, P. H. A. & Mortimer, A. M. (1986). A probabilistic approach to the identification of *Streptoverticillium* species. In *Biological, Biochemical and Biomedical Aspectes of Actinomycetes*, pp. 507-516. L. Ortiz-Oritiz, L. F. Bajalil & V. Yakoleff, eds. Academic Press: Orlando.

Logan, N. A. (1994). *Bacterial systematics*. Blackwell Scientific Publications: Oxford.

Logan, N. A. & Berkeley, R. C. W. (1981). A classification of the genus *Bacillus* based upon the API system. In *The Aerobic Endospore-forming Bacteria*, pp. 105-140. R. C. W. Berkeley & M. Goodfellow, eds. Academic Press: London.

Low, J. C., Chalmers, R. M., Donachie, W., Freeman, R., McLauchlin, J. & Sisson, P. R. (1992). Pyrolysis mass spectrometry of *Listeria monocytogenes* isolates from sheep. *Research in Vertebrate Science* 53, 64-67.

Ludwig, W. (1991). DNA sequencing in bacterial systematics In *Nucleic acid techniques in bacterial systematics*, pp. 69-94. E. Stackebrandt & M. Goodfellow, eds. John Wiley and Sons: New York.

Ludwig, W. & Schleifer, K. H. (1994). Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. *FEMS Microbioloy Reviews* **15**, 155-173.

Ludwig, W., Seewaldt, E., Schleifer, K. H. & Stackebrandt, E. (1981). The phylogenetic status of *Kurthia zopfi*. *FEMS Microbiology Letters* **10**, 193-197.

Ludwig, W., Weizenegger, M., Bertzl, D., Liedel, E., Lenz, T., Ludvigsen, A., Möllenhoff, D., Wenzig, P. & Schleifer, K. H. (1990). Complete nucleotide sequence of seven eubacterial genes coding for the elongation factor *Tu*: functional, structural and phylogenetical evaluations. *Archives of Microbiology* **153**, 241-247.

Ludwig, W., Kirchhof, G., Klugbauer, N., Weizenegger, M., Betzl, D., Ehrmann, M., Hertel, C., Jilg, S., Tatzel, R., Zitzelsberger, H., Liebl, S., Hochberger, M., Shah, J., Lane, D., Wallnofer, P. R. & Schleifer, K. H. (1992). Complete 23s ribosomal RNA sequences of Gram-positive bacteria with a low DNA G+C content. *Systematic and Applied Microbiology* **15**, 487-501.

Ludwig, W., Neumaier, J., Klugbauer, N., Brockmann, E., Roller, C., Jilg, S., Reetz, K., Schachtner, I., Ludvigsen, A., Bachleitner, M., Fischer, U. & Schleifer, K. H. (1993). Phylogenetic relationships of *Bacteria* based on comparative sequence analysis of elongation factor Tu and ATP-synthase ß-subunit genes. *Antonie van Leeuwenhoek* **64**, 285-305.

Luquin, M., Lopez, F. & Ausina, V. (1989). Capillary gas chromatographic analysis of mycolic acid cleavage products, cellular fatty acids, and alcohols of *Mycobacterium xenopi*. *Journal of Clinical Microbiology* **27**, 1403-1406.

MacDonnell, M. T. & Colwell, R. R. (1985). The contribution of numerical taxonomy to the systematics of Gram-negative bacteria. In *Computer-Assisted Bacterial Systematics*, pp. 107-135. M. Goodfellow, D. Jones & F. G. Priest, eds. Academic Press: London.

Maden, B. E., Dent, C. L., Farrell, T. E., Garde, J., McCallum, F. & Wakeman, J. (1987). Human rRNA heterogeneity. *Biochemical Journal* **246**, 519-527.

Magee, J. T. (1993). Whole-organism fingerprinting. In *Handbook of New Bacterial Systematics*, pp. 383-427. M. Goodfellow & A. G. O'Donnell eds. Academic Press: London.

Magee, J. T. (1994). Analytical fingerprinting methods. In *Chemical methods in Prokaryotic Systematics*, pp. 523-553. M. Goodfellow & A. G. O'Donnell eds. Chichester: John Wiley and Sons.

Magee, J. T., Hindmarch, J. M. & Meechan, D. F. (1983). Identification of staphylococci by pyrolysis gas-liquid chromatography. *Journal of Medical Microbiology* **16**, 483-495.

Magee, J. T., Hindmarch, J. M., Bennett, K. W., Duerden, B. I. & Aries, R. E. (1989a). A pyrolysis mass spectrometry study of fusobacteria. *Journal Medical Microbiology* **28**, 227-236.

Magee, J. T., Hindmarch, J. M., Burnett, I. A. & Pease, A. (1989b). Epidemiological typing of *Streptococcus pyogenes* by pyrolysis mass spectrometry. *Journal Medical Microbiology* **30**, 273-278.

Magee, J. T., Hindmarch, J. M. & Nicol, C. D. (1991). Typing of *Staphylococcus pyogenes* by pyrolysis mass spectrometry. *Journal of Medical Microbiology* **35**, 304-306.

Magee, J. T., Randle, E. A., Gray, S. J. & Jackson, S. K. (1993). Pyrolysis mass spectrometry characterisation and numerical taxonomy of *Aeromonas* spp. *Antonie van Leeuwenhoek* **64**, 315-323.

Magnusson, H. (1923). Spezifische infektiose Pseumonie beim Fohlen. Ein neuer Entreneger beim Pferde. *Archiv für Wissenschaftliche und Praktische Tierheikunde* **50**, 22-38.

Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Science of India* **2**: 49-55.

Maltsev, I. I., Kalinovsky, A. I., Zgurskaya, H. I. & Evtsushenko, L. I. (1992). Tyvelose in *Agromyces* cell walls. *Systematic and Applied Microbiology* **15,** 187-190.

Manafi, M., Kneifel, W. & Bascomb, S. (1991). Fluorogenic and chromogenic susbstrates used in bacterial diagnostics. *Microbiological Reviews* **55,** 335-348.

Manfio, G. P. (1995). *Towords minimal standards for the description of Streptomyces species.* Ph. D. Thesis, University of Newcastle upon Tyne.

Manly, B. F. J. (1986). *Multivariate Statistical Methods. A Primer.* Chapman & Hall, London.

Margalith, P. & Beretta, G. (1960). Rifomycin XI. Taxonomic study on *Streptomyces mediterranei* nov. sp. *Mycopathologia et Mycologia Applicata* **12**, 189-195.

Mathers, K., Freeman, R., Sisson, P. R. & Lightfoot, N. F. (1995). Differentiation between bacterial species and sub-species by pyrolysis mass spectrometry of extracted DNA. *Zentralblatt für Bakteriologie* (in press).

Maynard Smith, J. (1990). The evolution of prokaryotes: does sex matter ?. *Annual Reviews of Ecology and Systematics* **21**, 1-12.

McCarthy, A. J. & Cross, T. (1984). A taxonomic study of *Thermomonospora* and other monosporic actinomycetes. *Journal of General Microbiology* **130**, 5-25.

McClung, N. M. (1974). Genus *Nocardia* Trevisan 1989, 9. In *Bergey's Manual of Determinative Bacteriology, Volume 8.* pp. 726-746. R. E. Buchnan & N. E. Gibbons, eds. The Williams and Wilkins Company, Baltimore.

McCulloch, W. S. & Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* **5**, 115-133.

McNeil, M. & Brown, J. (1994). The medically important actinomycetes: epidermiology and microbiology. *Clinical Microbiological Reviews* **7**, 357-417.

McVeigh, H., Munro, J. & Embley, T. M. (1994). The phylogenetic position of *Pseudoamycolata halophobica* (Akimov *et al.*) and a proposal to reclassify it as *Pseudonocardia halophobica*. *International Journal of Systematic Bacteriology* **44**, 300-302.

Meier, A., Kirschner, P., Schroeder, K. H., Wolters, J., Kroppenstedt, R. M. & Böttger, E. C. (1993). *Mycobacteriun intermedium* sp. nov. *International Journal of Systematic Bacteriology* **43**, 204-209.

Metcalf & G. & Brown, M. (1957). Nitorgen fixation by new species of *Nocardia*. *Journal of General Microbiology* **17**, 567-572.

Meuzelaar, H. L. C. & Kistemaker, P. G. (1973). A technique for fast and reproducible fingerprinting of bacteria by pyrolysis mass spectrometry. *Analytical Chemistry* **45**, 587-590.

Meuzelaar, H. L. C., Kistemaker, P. G., Eshuis, W. & Engel, H. W. B. (1976). Progress in automated and computerised characterisation of micro-organisms by pyrolysis mas spectrometry. In *Rapid Methods and Automation in Microbiology*, pp. 225-230. H. H. Johnston & S. W. B. Newsom, eds. Learned Information: Oxford.

Meuzelaar, H. L. C., Haverkamp, J. & Hileman, F. D. (1982). *Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials.* Compendium and Atlas. Elsevier: Amsterdam.

Meyer, B., Hansen, T., Nute, D., Albersheim, P., Darvill, A., York, W. & Sellers, J. (1991). Identification of the ^{1}H-NMR spectra of complex oligosaccharides with artificial neural networks. *Science* **251**, 542-544.

Miller, R. G. (1974). The jackknife - a review. *Biometrika* **61**, 1-15.

Millership, S. (1993). Use of a neural network for analysis of bacterial whole cell protein fingerprints. *Binary* **5**, 126-131.

Milligan., G. W. (1989). A validation study of a variable weighting algorithm for cluster analysis. *Journal of Classification* **6**, 53-71.

Minnikin, D. E. & O'Donnell, A. G. (1984). Actinomycete envelope lipid and peptidoglycan composition. In *The Biology of the Actinomycetes*, pp. 337-388. M. Goodfellow, M. Mordarski & S. T. Williams, eds. Academic Press: London.

Minnikin, D. E., Hutchinson, I. G., Caldicott, A. B. & Goodfellow, M. (1980). Thin-layer chromatography of methanolysates of mycolic acid-containg bacteria. *Journal of Chromatography* **188**, 221-233.

Minnikin, D. E., O'Donnell, A. G., Goodfellow, M., Alderson, G., Athayle, M., Schaal, A. & Parlett, J. H. (1984). An integrated procedure for the extraction of isoprenoid quinones and polar lipids. *Journal of Microbiological Methods* **2**, 233-241.

Minsky, M. & Papert, S. (1969). *Perceptrons.* MIT Press: Cambridge.

Mittermayr, C. R., Drouen, A. C. J. H. , Otto, M. & Grasserbauer, M. (1994). Neural networks for library search of ultraviolet spectra. *Analytica Chimica Acta* **294**, 227-242.

Montague, G. & Morris, J. (1994). Neural network contributions in biotechnology. *Trends in Biotechnology* **12**, 312-324.

Moore, L. B. & McCarthy, B. (1967). Comparative study of ribosomal ribonucleic acid cistrons in enterobacteria and mycobacteria. *Journal of Bacteriology* **94**, 1066-1074.

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., Tkacz, A., Pulverer, G. & Schaal, K. P. (1980a). Ribosomal ribonucleic acid similarities in the classification of *Rhodococcus* and related taxa. *Journal of General Microbiology* **118**, 313-319.

Mordarski, M., Goodfellow, M., Szyba, K., Tkacz, A., Pulverer, G. & Schaal, K. P. (1980b). Deoxyribonucleic acid reassociation in the classification of the genus *Rhodococcus*. *International Journal of Systematic Bacteriology* **30**, 521-527.

Mordarski, M., Tkacz, A., Goodfellow, M., Schaal, K. P. & Pulverer, G. (1981). Ribosomal ribonucleic acid similarities in the classification of actinomycetes. *Zentralblatt für Bakteriologie, Mikrobiologie und Hygiene. I. Abteilung, Supplement* **11**, 79-85.

Morden, C. E., Delwiche, C. F., Kuhsel, M. & Palmer, J. D. (1992). Gene phylogenies and the endosymbiotic origin of plastids. *BioSystems* **28**, 75-90.

Morris, A. J. , Montague, G. A. & Willis, M. J. (1994). Artificial neural networks studies in process modeling and control. *Chemical Engineering Research & Design* **72,** 3-19.

Morris, C. W. & Boddy, L. (1993). Some neural network packages: A comparative review. *Binary* **5**, 7-9.

Morris, C. W., Boddy, L. & Allman, R. (1992). Identification of basidiomycete spores by neural network analysis of flow cytometry data. *Mycological Research* **96**, 697-701.

Munk, M. E., Madison, M. S. & Robb, E. W. (1991). Neural network models for infrared spectrum interpretation. *Mikrochimika Acta (Wien)* **2**, 505-514.

Murray, R. G. E. (1989). The higher taxa, or a place for everything. In *Bergey's Manual of Systematic Bacteriology*, pp. 2329-2332. S. T. Williams, M. E. Sharpe & J. G. Holt, eds. Williams and Wilkins: Baltimore.

Murray, R. G. E., Brenner, D. J., Colwell, R. R., De Vos, P., Goodfellow, M., Grimont, P. A. D., Pfennig, N., Stackebrandt, E. & Zavarzin, G. A. (1990). Report of the *ad hoc* committee on approaches to taxonomy with proteobacteria. *International Journal of Systematic Bacteriology* **40**, 213-215.

Mylvaganam, S. & Dennis, P. P. (1992). Sequence heterogeneity between the two genes encoding 16S rRNA from the halophlic archaebacterium *Haloarcula marismortui*. *Genetics* **130**, 399-410.

Nagylaki, T. (1984). Evolution of multigene families under interchromosomal gene conversion. *Proceedings of National Academic Society USA* **81**, 3796-3800.

Nagylaki, T. & Petes, T. D. (1982). Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. *Genetics* **100**, 315-337.

Naumann, D., Helm, D. & Schultz, C. (1994). Characterization and identification of microorganisms by FT-IR spectroscopy and FT-IR microscopy. In *Bacterial Diversity and Systematics*, pp. 67-86. F. G. Priest, A. Ramos-Cormenzana & B. J. Tindall, eds. Plenum Press: New York.

Neefs, J. M., Van de Peer, Y., De Rijk, P., Chapelle, S. & De Wachter, R. (1993). Compilation of small ribosomal subunit RNA structures. *Nucleic Acids Research* **21**, 3025-3049.

Nelder, J.C. (1979). *Genstat Reference Manual.* Scientific and Social Service Program Library, University of Edinburgh: Endinburgh.

Nesterenko, O. A., Nogina, T. M., Kasumova, S. A. & Kvasnikov, E. I. & Batrakov, S. G. (1982). *Rhodococcus luteus* nom. nov. and *Rhodococcus maris* nom. nov. *International Journal of Systematic Bacteriology* **32**, 1-14.

Niemalä, S. I., Hopkins, J. W. & Quadling, C. (1968). Selecting an ecological binary test battery for a set of microbial cultures. *Canadian Journal of Microbiology* **14**, 271-279.

Noller, H. F. (1984). Structure of ribosomal RNA. *Annual Reviews of Microbiology* **53**, 119-162.

Nolof, G. & Hirsch, P. (1962). Nocardia hydrocarbonoxydans n. spec.: ein oligocarbophiler Actinomycet. *Archic für Mikrobiologie* **44**, 266-277.

Nomura, M., Traub, P. & Bechmann, H. (1968). Hybrid 3DS ribosomal particles reconstituted from components of different bacterial origins. *Nature* **219**, 793-799.

Normand, P., Cournoyer, B., Simonet, P. & Nazaret, S. (1992). Analysis of a ribosomal RNA operon in the actinomycete *Frankia*. *Gene* **111**, 119-124.

O'Brien, M. & Colwell, R. R. (1987). Characterization tests for numerical taxonomy studies. *Methods in Microbiology* **19**, 69-104.

O'Donnell, A. G. (1985). Numerical analysis of chemotaxonomic data. In *Computer-Assisted Bacterial Systematics*, pp. 403-414. M. Goodfellow, D. Jones & F. G. Priest, eds. Academic Press: London.

O'Donnell, A. G., MacFie, H. J. H. & Norris, J. R. (1980). An investigation of the relationship between *Bacillus cereus, Bacillus thuringiensis* and *Bacillus mycoides* using pyrolysis gas-liquid chromatography. *Journal of General Microbiology* **119**, 189-194.

O'Donnell, A. G., Goodfellow, M. & Hawksworth, D. L. (1994). Theoretical and practical aspects of the quantification of biodiversity among microorganisms. *Phylosophical Transactions of the Royal Society of London, Series B* **345**, 65-73.

Ochi, K. (1995). Amino acid sequence analysis of ribosomal protein AT-L30 from members of the family *Pseudonocardiaceae*. *International Journal of Systematic Bacteriology* **45**, 110-115.

Ochi, K. & Hiranuma, H. (1994). A taxonomic review of the genera *Kitasatosporia* and *Streptoverticillium* by analysis of ribosomal protein AT-L30. *International Journal of Systematic Bacteriology* **44**, 285-292.

Ochi, K., Satoi, E. & Shima, J. (1993). Amino acid sequemce analysis of ribosomal protein AT-L30 from *Streptosporangium corrugatum* and *Kibdelosporangium aridum*. *Systematic and Applied Microbiology* **16**, 13-16.

Ochman, H. & Wilson, A. C. (1987). Evolution in bacteria: Evidence for a universal substitution rate in cellular genomes. *Journal of Molecular Evolution* **26**, 74-86.

Olsen, G. J. (1988). Phylogenetic analysis using ribosomal RNA. *Methods in Enzymology* **164**, 793-812.

Olsen, G. J. & Woese, C. R. (1993). Ribosomal RNA: a key to phylogeny. *FASEB Journal* **7**, 113-123.

Olsen, G. J., Woese, C. R. & Overbeek, R. (1994a). The winds of (evolutionary) change: breathing new life into microbiology. *Journal of Bacteriology* **176**, 1-6.

Olsen, G. J., Matsuda, H., Hagstrom, R. & Overbeek, R. (1994b). FastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Computer Applications in Biological Sciences* **10**, 41-48.

Oren, A., Lau, P. P. & Fox, G. E. (1988). The taxonomic status of "*Halobacterium marismortui* from the dead sea: a comparison with *Halobacterium vallismortis*. *Systematic and Applied Microbiology* **10**, 251-258.

Orr, K., Gould, F. K., Sisson, P. R., Lightfoot, N. F., Freeman, R. & Burdess, D. (1991). Rapid interstrain comparison by pyrolysis mass spectrometry in nosocomial infection with *Xanthomonas maltophilia*. *Journal of Hospital Infection* **17**, 187-195.

Orr, K. E., Lightfoot, N. F., Sisson, P. R., Harkis, B. A., Tweddle, J. L., Boyd, P., Carroll, A., Jackson, C. J., Wareing, D. R. A. & Freeman, R. (1995). Direct milk excretion of *Campylobacter jejuni* in a dairy cow causing cases of human enteritis. *Epidemiology and Infection* **114**, 15-24.

Orso, S., Gouy, M., Navarro, E. & Normand, P. (1994). Molecular phylogenetic analysis of *Nitrobacter* spp. *International Journal of Systematic Bacteriology*, **44**, 83-86.

Palleroni, N. J., Kunisawa, R., Contopoulou, R. & Doudoroff, M. (1973). Nucleic acid homologies in the genus *Pseudomonas*. *International Journal of Systematic Bacteriology* **23**, 333-339.

Pao, Y. H. (1989). *Adaptive Pattern Recognition and Neural Networks.* Addison-Wesley: Reading, Massachusetts.

Park, Y. H., Hori, H., Suzuki, K. I., Osawa, S. & Komagata, K. (1987). Phylogenetic analysis of the coryneform bacteria by 5S rRNA sequences. *Journal of Bacteriology* **169**, 1801-1806.

Park, Y. H., Yim, D. G., Kim, E., Kho, Y. H., Mheen, T. I., Lonsdale, J. & Goodfellow, M. (1991). Classification of acidophilic, neutrotolerant and neutrophilic streptomycetes by nucleotide sequencing of 5S ribosomal RNA. *Journal of General Microbiology* **137**, 2265-2269.

Park, Y. H., Suzuki, K. I., Yim, D. G., Lee, K. C., Kim, E., Yoon, J. S., Kim, S. J., Kho, Y. H., Goodfellow, M. & Komagata, K. (1993). Suprageneric classification of peptidoglycan group B actinomycetes by nucleotide sequencing of 5S ribosomal RNA. *Antonie van Leeuwenhoek* **64**, 307-313.

Parker, D. (1982). *Learning Logic*, Invention Report, S81-64, File 1, Office of Technology Licencing, Stanford University.

Perelson, A. S. & Bell, G. I. (1977). Mathematical models for the evolution multigene families by unequal crossing over. *Nature* **265**, 304-310.

Petes, T. D. (1980). Unequal meiotic recombination within tandem arrays of yeast ribosomal DNA genes. *Cell* **19**, 765-774.

Pijper, A. & Pullinger, B. D. (1927). South African nocardioses. *Journal of Tropical Medicine and hygiene* **30**, 153-156.

Pike, E. B. (1972). Aerobic bacteria. In *Ecological Aspects of Used Water Treatment,* pp. 1-63. C. R. Curds & H. A. Hawkes, eds. Academic Press: New York.

Pinoy, E. (1913). Actinomycoses et mycétomes. *Bulletin de l'Institut Pasteur* **11**, 929-938.

Pitcher, D. G., Saunders, N. A. & Owen, R. J. (1989). Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Letters in Applied Microbiology* **8**, 151-156.

Pitt, P. A. & Jenkins, D. (1990). Causes and control of *Nocardia* in activated sludge. *Research Journal of Water Pollution Control Federation* **62**, 143.

Pittenger, R. C. & Brigham, R. B. (1956). *Streptomyces orientalis* n. sp., the source of vancomycin. *Antibiotics and Chemotheraphy* **6**, 642-647.

Pitulle, C., Dorsch, M., Kazda, J., Wolters, J. & Stackebrandt, E. (1992). Phylogeny of rapidly growing members of the genus *Mycobacterium*. *International Journal of Systematic Bacteriology* **42**, 337-343.

Polz, M. F., Distel, D. L., Zarda, B., Amann, R., Felbeck, H., Ott, J. A. & Cavanaugh, C. M. (1994). Phylogenetic analysis of a highly specific association between ectosymbiotic, sulfur-oxidizing bacteria and a marine nematode. *Applied Environmental Microbiology* **60**, 4461-4467.

Postic, D., Assous, M. V., Grimont, P. A. D. & Baranton, G. (1994). Diversity of *Borellia burdorferi sensu* Lato evidenced by restriction fragment length polymorphism of *rrf*(5S)-*rrl*(23S) intergenic spacer amplicons. *International Journal of Systematic Bacteriology* **44**, 743-752.

Pot, B., Vandamme, P. & Kersters, K. (1994) Analysis of electrophoretic whole-organism protein fingerprints. In *Chemical methods in Prokaryotic Systematics*, pp. 493-521. M. Goodfellow & A. G. O'Donnell, eds. Chichester: John Wiley and Sons.

Prauser, H. (1970). Characters and genera arrangement in the *Actinomycetales*. In *The Actinomycetales*. pp. 407-418. H. Prauser, ed. Gustav Fischer Verlag: Jena.

Prauser, H. (1978). Considerations on taxonomic relations among Gram-positive, branching bacteria. *Zentralblatt für Bakteriologie, Parasitenfunde, Infektionskrankheiten und Hygiene. I. Abteilung Supplement* **6**, 3-12.

Prauser, H. (1981). Nocardioform orgainsms: general charcaterization and taxonomic relationships. *Zentralblatt für Bakteriologie, Parasitenfunde, Infektionskrankheiten und Hygiene. I. Abteilung Supplement* **11**, 17-24.

Pridham, T. G. & Lyons, A. J. (1969). Progress in the clarification of the taxonomic and nomenclatural status of some problem actinomycetes. *Developments in Industrial Mcirobiology* **10**, 183-221.

Priest, F. G. & Alexander, B. (1988). A frequency matrix for probabilistic identification of some bacilli. *Journal of General Microbiology* **134**, 3011-3018.

Priest, F. G. & Williams, S. T. (1993). Computer-assisted identification. In *Handbook of New Bacterial Systematics*, pp. 361-381. M. Goodfellow & A. G. O'Donnell eds. Academic Press: London.

Pühler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H. P., Lottspeich, F., Garrett, R. A. & Zillig, W. (1989). Archaebacterial DNA-dependent RNA polymerases testify the evolution of the eukaryotic nuclear genome. *Proceedings of National Academic Society USA* **86**, 4569-4573.

Qian, N. & Sejnowski, T. S. (1988). Predicting the secondary structure of globular proteins using neural network models. *Journal of Molecular Biology* **202**, 865-884.

Qu, L. H., Michot, B. & Bachellerie, J. P. (1983). Improved methods for structure probing in large RNA's: a rapid "heterogenous" sequencing approach is coupled to the direct mapping of nuclease accessible sites: application to the 5' terminal domain of eukaryotic 28S rRNA. *Nucleic Acids Research* **11**, 5903-5919.

Radomski, J. P., Vanhalbeek, H. & Meyer, B. (1994). Neural network based recognition of oligosaccharide H-1-NMR spectra. *Nature Structural Biology* 1, 217-218.

Rainey, F. A., Schumann, P., Prauser, H., Toalster, R. & Stackebrandt, E. (1993). *Sporichthya polymorpha* represents a novel line of descent within the order Actinomycetales. *FEMS Microbiology Letters* **109**, 263-268.

Rainey, F. A., Weiss, N. & Stackebrandt, E. (1994a). *Coriobacterium* and *Atopobium* are phylogenetic neighbors within the actinomycetes line of descent. *Systematic and Applied Microbiology* **17**, 202-205.

Rainey, F. A., Weiss, N., Prauser, H. & Stackebrandt, E. (1994b). Further evidence for the phylogenetic coherence of actinomycetes with group B-peptidoglycan and evidence for the phylogenetic intermixing of the genera *Microbacterium* and *Aureobacterium* as determined by 16S rDNA analysis. *FEMS Microbiology Letters* **118**, 135-140.

Rainey, F. A., Burghardt, J., Kroppenstedt, R. M., Klatte, S. & Stackebrandt, E. (1995a). Phylogenetic analysis of the genera *Rhodococcus* and *Nocardia* and evidence for the evolutionary origin of the genus *Nocardia* from within the radiation of *Rhodococcus* species. *Microbiology* **141**, 523-528.

Rainey, F. A., Burghardt, J., Kroppenstedt, R. M., Klatte, S. & Stackebrandt, E. (1995b). Polyphasic evidence for the transfer of *Rhodococcus roseus* to *Rhodococcus rhodochrous*. *International Journal of Systematic Bacteriology* **45**, 101-103.

Rainey, F. A., Klatte, S., Kroppenstedt, R. M. & Stackebrandt, E. (1995c). *Dietzia*, a new genus including *Dietzia maris* comb. nov., formerly *Rhodococcus maris*. *International Journal of Systematic Bacteriology* **45**, 32-36.

Ramsey, E. D., Collins, M. D. & Mellon, F. A. (1988). Tandem mass spectrometric studies of menaquinones. Potential for rapid location of points of saturation. *Rapid Communications in Mass Spectrometry* **2**, 273-274.

Rastogi, N., David, H. L. & Frehel, C. (1986). Ultrastructure of "*Gordona aurantiaca*" Tsukamura 1971. *Current Microbiology* **13**, 51-56.

Rataj, T. & Schindler, J. (1991). Identification of bacteria by a multilayer neural network. *Binary* **3**, 159-164.

Ratzke, U. (1965). Untersuchungen an Proactinomyceten (*Nocardia*). Verbreitung von Proactinomyceten in den Böden eines Dauerdüngungs- und Fruchtfolgeversuches. *Zentralblatt für Mikrobiologie II Abt.* **119**, 601-611.

Ravin, A. W. (1961). The genetics of transformation. *Advances in Genetics* **10**, 61-163.

Reiner, E. (1965). Identification of bacterial strains by pyrolysis gas liquid chromatography. *Nature* **206**, 1272-1274.

Reiner, E. & Ewing, W. J. (1968). Chemotaxonomic studies of some Gram-negative bacteria by means of pyrolysis gas-liquid chromatography. *Nature* **217**, 191-194.

Reiner, E., Hicks, J. J. & Sulzer, C. R. (1973). Leptospiral taxonomy by pyrolysis gas-liquid chromatography. *Canadian Journal of Microbiology* **19**, 1203-1206.

Relman, D. A., Schmidt, T. M., MacDermott, R. P. & Falkow, S. (1992). Identification of the uncultured bacillus of Whipple's disease. *New England Journal of Medicine* **327**, 293-301.

Ridell, M. & Goodfellow, M. (1983). Numerical classification of *Mycobacterium farcinogenes*, *Mycobacterium senegalense* and related taxa. *Journal of General Microbiology* **129**, 599-611.

Ridell, M., Goodfellow, M. & Minnikin, D. E. (1985). Immunodiffusion and lipid analyses in the classification of "*Mycobacterium album*" and the "*aurantiaca*" taxon. *Zentralblatt für Bakteriologie* **259**, 1-10.

Riegel, P., Kamne-Fotso, M. V., De Briel, D., Prévost, G., Jehl, F., Piémont, Y. & Monteil, H. (1994). *Rhodococcus chubuensis* Tsukamura 1982 is a later subjective synonym of *Gordona sputi* (Tsukamura 1978) Stackebrandt 1989 comb. nov. *International Journal of Systematic Bacteriology* **44**, 764-768.

Rivera, M. C. & Lake, J. A. (1992). Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* **257**, 74-76.

Robb, E. W. & Munk, M. E. (1990). A neural network approach to infrared spectrum interpretation. *Mikrochimica Acta* **1**, 131-155.

Robertson, J. A., Howard, L. A., Zinner, C. L. & Stemke, G. W. (1994). Comparison of 16S rRNA genes within the T960 and Parvo biovars of ureaplasmas isolated from humans. *International Journal of Systematic Bacteriology* **44**, 836-838.

Rogall, T., Wolters, J., Flohr, t. & Böttger, E. C. (1990). Towards a phylogeny and definition of species at the molecular level within the genus *Mycobacterium*. *International Journal of Systematic Bacteriology* **40**, 323-330.

Rogers, H. J., Perkins, H. R. & Ward, J. B. (1980). *Microbial Cell Walls and Membranes.* Chapman & Hall: London.

Rohlf, F. J. & Fisher, D. R. (1970). Tests for hierarchical structure in random data sets. *Systematic Zoology* **17**, 407-412.

Roller, C., Ludwig, W. & Schleifer, K. H. (1992). Gram-positive bacteria with a high DNA G+C content are characterized by a common insertion within their 23S rRNA genes. *Journal of General Microbiology* **138**, 167-175.

Rosenblatt, F. (1957). *The perceptron: a percieving and recognizing automation (project PARA).* Cornell Aeronautical Laboratory Report, VG-1196-G-1.

Rossi Doria, T. (1891). Su di alcune specie di '*Streptothrix*' trovate nell'aria studate in rapporto a quelle giá note a specialmente all' '*Actinomyces*'. *Annali dell' Istituto d'Igiene Sperimentale, Universita Roma* **1**, 399-438.

Rowbotham, T. J. & Cross, T. (1977). Ecology of *Rhodococcus coprophilus* sp. nov.: an aerobic nocardioform actinomycete belonging to the '*rhodochrous*' complex. *Journal of General Microbiology* **100**, 123-138.

Ruimy, R., Breittmayer, V., Elbaze, P., Lafay, B., Boussemart, O., Gauthier, M. & Christen, R. (1994a). Phylogenetic analysis and assessment of the genera *Vibrio, Photobacterium, Aeromonas*, and *Plesiomonas* deduced from small-subunit rRNA sequences. *International Journal of Systematic Bacteriology* **44**, 416-426.

Ruimy, R., Boiron, P., Boivin, V. & Christen, R. (1994b). A phylogeny of the genus *Nocardia* deduced from the analysis of small-subunit ribosomal DNA sequences, including transfer of *Nocardia amarae* to the genus *Gordona* as *Gordona amarae* comb. nov. *FEMS Microbiology Letters* **123**, 261-268.

Rumelhart, D.E., McClelland, J.L & the PDP Research Group (1986). *Parallel Distributed Processing.* MIT Press: Cambridge.

Saccone, C., Lanave, C., Pesole, G. & Preparata, G. (1990). Influence of base composition on quantitative estimates of gene evolution. *Methods in Enzymology* **183**, 570-583.

Sackin, M. J. (1981). Vigour and pattern as applied to multistate quantitative characters in taxonomy. *Journal of General Microbiology* **122**, 247-254.

Sackin, M. J. (1987). Computer programs for classification and identification. *Methods in Microbiology* **19**, 459-494.

Sackin, M. J. & Jones, D. (1993). Computer-assisted classification. In *Handbook of New Bacterial Systematics*, pp. 282-313. M. Goodfellow & A. G. O'Donnell eds. Academic Press: London.

Saddler, G. S., O'Donnell, A. G., Goodfellow, M. & Minnikin, D. E. (1987). SIMCA pattern recognition in the analysis of streptomycete fatty acids. *Journal of General Microbiology* **133**, 1137-1147.

Saddler, G.S., Falconer, C. & Sanglier, J. J. (1989). Preliminary experiments for the selection and identification of actinomycetes by pyrolysis mass spectrometry. *Actinomycetologia* **2**, S3-S4.

Saddler, G. S., Tavecchia, P., Lociuro, S., Zanol, M., Colombo, L. & Selva, E. (1991). Analysis of madurose and other actinomycete whole cell sugars by gas chromatography. *Journal of Microbiological Methods* **14**, 185-191.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuch, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988). Primer-directed enzymatic amplication of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491.

Saitou, N. (1990). Maximum likelihood methods. *Methods in Enzymology* **183**, 584-598.

Saitou, N. & Imanishi, T. (1989). Relative efficiencies of the Fitch-Margoliash, maximum-evolution and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. *Molecular Biology and Evolution* **6**, 514-525.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406-425.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual.* 2nd ed. Cold Spring Harbor Laboratory Press: New York.

Sanger, F., Brownlee, G. G. & Barrel, B. G. (1965). A two dimensional fractionation procedure for radioactive nucleotides. *Journal of Molecular Biology* **13**, 373-398.

Sanglier, J. J., Whitehead, D., Saddler, G. S., Ferguson, E. V. & Goodfellow, M. (1992). Pyrolysis mass spectrometry as a method for the classification, identification and selection of actinomycetes. *Gene* **115**, 235-242.

Santibanez, M. & Rohde, K. (1987). A multiple alignment program for protein sequences. *Computer Applications in Biological Sciences* **3**, 111-114.

Schaal, K. P. (1985). Identification of clinically significant actinomycetes and related bacteria using chemical techniques. In *Chemical Methods in Bacterial Systematics*, pp. 359-381. M. Goodfellow & D. E. Minnikin, eds. Academic Press: London.

Schaal, K. P. & Lee, H. J. (1992). Actinomycete infections in humans - a review. *Gene* **151**, 201-211.

Schindler, J., Paryzek, P. & Farmer, J. (1994). Identification of bacteria by artificial neural networks. *Binary* **6**, 191-196.

Schleifer, K. H. & Kandler, O. (1972). Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriological Review* **36**, 407-477.

Schleifer, K. H. & Seidl, P. H. (1985). Chemical composition and structure of murein. In *Chemical Methods in Bacterial Systematics*, pp. 201-219. M. Goodfellow & D. E. Minnikin, eds. Academic Press: London.

Schleifer, K. H., Ludwig, W. & Amman, R. (1993). Nucleic acid probes. In *Handbook of New Bacterial Systematics*, pp. 463-510. M. Goodfellow & A. G. O'Donnell eds. Academic Press: London.

Schmidt, T. M., DeLong, E. F. & Pace, N. R. (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *Journal of Bacteriology* **173**, 4371-4378.

Schneider, G. & Wrede, P. (1994). The rational design of amino-acid-sequences by artificial neural networks and simulated molecular evolution - *de novo* design of an idealized leader peptidase cleavage site. *Biophysical Journal* **66**, 335-344.

Schulten, H. R. & Lattimer, R. P. (1984). Applications of mass spectrometry to polymer analysis. *Mass Spectrometry Review* **3**, 2310-2315.

Schulze, H. G, Blades, M. W., Bree, A. V., Gorzalka, B. B., Greek, L. S.& Turner, R. F. B. (1994). Characteristics of backpropagation neural networks employed in the identification of neurotransmitter Raman spectra. *Applied Spectroscopy* **48**, 50-57.

Schuppler, M., Mertens, F., Schön, G. & Göbel, U. B. (1995). Molecular characterization of nocardioform actinomycetes in activated sludge by 16S rRNA analysis. *Microbiology* **141**, 513-521.

Sela, S., Clark-Curtiss, J. E. & Bercovier, H. (1989). Characterization and taxonomic implications of the rRNA genes of *Mycobacterium leprae*. *Journal of Bacteriology* **171**, 70-73.

Sellers, J., York, W., Albersheim, P., Darvill, A. & Meyer, B. (1990). Identification of the mass spectra of partially methylated alditol acetates by artificial neural networks. *Carbohydrate Research* **207**, c1-c5.

Serrano, J. A., Tablante, R. V., de Serrano, A. A., de San Blas, G. C. & Imaeda, T. (1972). Physiological, chemical and ultrastructural characteristics of *Corynebacterium rubrum*. *Journal of General Microbiology* **70**, 339-349.

Seviour, E. M., Williams, C. J., Seviour, R. J., Soddell, J. A. & Lindrea, K. C. (1990). A survey of filamentous bacterial populations from foaming activated sludge plants in Eastern States of Australia. *Water Research* **24**, 493-498.

Sezgin, M., Jenkins, D. & Parker, D. S. (1978). A unified theory of filamentous activated sludge bulking. *Journal of Water Pollution and Control Federation* **50**, 362.

Sezgin, M., Lechecalier, M. P. & Karr, P. R. (1988). Isolation and identification of actinomycetes present in activated sludge scum. *Water Science and Technology* **20**, 257-263.

Shapiro, C. L., Haft, R. F., Gantz, N. M., Doern, G. V., Christenson, J. C., O'Brien, R., Overall, J. C., Brown, B. A. & Wallace, R. J., Jr. (1991). *Tsukamurella paurometabolum*: A novel pathogen causing catheter-related bacteremia in patients with cancer. *Clinical Infectious Diseases* **14**, 200-203.

Shaw, N. (1968). The detection of lipids on thin-layer chromatograms with the periodate-Schiff reagent. *Biochimica et Biophysica Acta* **164**, 435-436.

Shute, L. A., Gutteridge, C. S., Norris, J. R. & Berkeley, R. C. W. (1984). Curie-point pyrolysis mass spectrometry applied to characterization and identification of selected *Bacillus* species. *Journal of General Microbiology* **130**, 343-355.

Shute, L. A., Berkeley, R. C. W., Norris, J. R. & Gutteridge, C. S. (1985). Pyrolysis mass spectrometry in bacterial systematics. In *Chemical Methods in Bacterial Systematics*, pp. 95-114. M. Goodfellow & D. E. Minnikin eds. Academic Press: London.

Shute, L. A., Gutteridge, C. S., Norris, J. R. & Berkeley, R. C. W. (1988). Reproducibility of pyrolysis mass spectrometry: effect of growth medium and instrument stability on the differentiation of selected *Bacillus* species. *Journal of Applied Bacteriology* **64**, 79-88.

Siebler, M., Rose, G., Sitzer, M., Bender, A. & Steinmetz, H. (1994). Real-time identification of cerebral microemboli with US feature detection by a neural network. *Radiology* **192**, 739-742.

Silvestri, L. G., Turri, M., Hill, L. R. & Gilardi, E. (1962). A quantitative approach to the systematics of actinomycetes based on overall similarity. In *Microbial Classification*, pp. 333-360. G. C. Ainsworth & P. H. A. Sneath, eds. Cambridge University Press: Cambridge.

Simmonds, P. G. (1970). Whole microorganisms studied by pyrolysis-gas chromatography-mass spectrometry: significance for extraterrestrial life detection systems. *Applied Microbiology* **20**, 567-572.

Simpson, E. H. (1949). Measurement of diversity. *Nature (London)* **163**:688.

Simpson, P. K. (1990). Artificial Neural Systems. Pergamon Press: Oxford.

Sisson, P. R., Freeman, R., Magee, J. G. & Lightfoot, N.F. (1991a). Differentiation between mycobacteria of the *Mycobacterium tuberculosis* complex by pyrolysis mass spectrometry. *Tubercle* **72**, 206-209.

Sisson, P. R., Freeman, R., Lightfoot, N. F. & Richardson, I. R. (1991b). Incrimination of an environmental source of a case of Legionnaires' disease by pyrolysis mass spectrometry. *Epidermiology and Infection* **107**, 127-132.

Sisson, P. R., Freeman, R., Gould, F. K. & Lightfoot, N. F. (1991c). Strain differentiation of nosocomial isolates of *Pseudomonas aeruginosa* by pyrolysis mass spectrometry. *Journal of Hospital Infection* **19**, 137-140.

Sisson, P. R., Lightfoot, N. F. & Freeman, R. (1992a). Pyrolysis mass spectrometry of micro-organisms. *PHLS Microbiology Digest* **9**, 65-68.

Sisson, P. R., Freeman, R., Magee, J. G. & Lightfoot, N. F. (1992b). Rapid differentiation of *Mycobacterium xenopi* from mycobacteria of the *Mycobacterium avium-intracellulare* complex by pyrolysis mass spectrometry. Journal of Clinical Pathology **45**, 355-357.

Skerman, V. B. D. (1968). A new type of micromanipulator and microforge. *Journal of General Microbiology* **54**, 287-297.

Skerman, V. B. D., McGowan, V. & Sneath, P. H. A. (1980). Approved lists of bacterial names. *International Journal of Systematic Bacteriology* **30**, 225-420.

Smits, J. R. M., Breedveld, L. W., Derksen, M. W. J., Kateman, G., Balfoort, H. W., Snoek, J. & Hofstraat, J. W. (1992). Pattern classification with artificial neural networks - classification of algae, based upon flow cytometer data. *Analytica Chimica Acta* **258**, 11-25.

Sneath, P. H. A. (1957a). The application of computers to taxonomy. *Journal of General Microbiology* **17**, 201-226.

Sneath, P. H. A. (1957b). Some thoughts on bacterial classification. *Journal of General Microbiology* **17**, 184-200.

Sneath, P. H. A. (1962). Construction of taxonomic groups. *Symposium of Society for General Microbiology* **12**, 287-332.

Sneath, P. H. A. (1968). Vigour and pattern in taxonomy. *Journal of General Microbiology* **54**, 1-11.

Sneath, P. H. A. (1970). Application of numerical taxonomy to *Actinomycetales*: problems and prospects. In *The Actinomycetales*. pp. 371-377. H. Prauser, ed. Gustav Fischer Verlag: Jena.

Sneath, P. H. A. (1972). Computer taxonomy. *Methods in Microbiology* **4**, 29-98.

Sneath, P. H. A. (1974a). Phylogeny of microorganisms. *Symposium of the Society for General Microbiology* **24**, 1-39.

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

Sneath, P. H. A. (1975). Cladistic representation of reticulate evolution. *Systematic Zoology* **24**, 360-368.

Sneath, P. H. A. (1977). A method for testing the distinctness of clusters: a test of disjunction of two clusters in Euclidean space as measured by their overlap. *Mathematical Geology* **9**, 123-143.

Sneath, P. H. A. (1978). Classification of micro-organisms. In *Essays in Microbiology*, pp. 1-31. J. R. Norris & M. H. Richmond, eds. John-Wiley and Sons, Chichester.

Sneath, P. H. A. (1979a). Numerical taxonomy and automated identification: some implications for geology. *Computers and Geosciences* **5**, 41-46.

Sneath, P. H. A. (1979b). BASIC program for a significance test for two clusters in euclidean space as measured by their overlap. *Computers and Geosciences* **5**, 143-155.

Sneath, P. H. A. (1979c). BASIC program for nonparametric significance of overlap between a pair of clusters using the Kolmogorov-Smirnov test. *Computers and Geosciences* **5**, 173-188.

Sneath, P. H. A. (1979d). BASIC programs 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. (1979e). BASIC program for character separation indices from an identification matrix of percent positive characters. *Computers and Geosciences* **5**, 349-357.

Sneath, P. H. A. (1980a). 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. (1980b). BASIC programs for determining the best identification scores possible from the most typical examples when compared with an identification matrix of percent positive characters. *Computers and Geosciences* **6**, 27-34.

Sneath, P. H. A. (1980c). 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. (1983). Distortions of taxonomic structure from incomplete data on a restricted set of reference strains. *Journal of General Microbiology* **129**, 1045-1073.

Sneath, P. H. A. (1986). Estimating uncertainty in evolutionary trees from Manhattan-distance triads. *Systematic Zoology* **35**, 470-488.

Sneath, P. H. A. (1989a). Analysis and interpretation of sequence data for bacterial systematics: the view of a numerical taxonomist. *Systematic and Applied Microbiology* **12**, 15-31.

Sneath, P. H. A. (1989b). The use of computers for the identification of bacterial isolates from natural habitats. *Binary* **1**, 89-91.

Sneath, P. H. A. (1993). Evidence from *Aeromonas* for genetic crossing-over in ribosomal sequences. *International Journal of Systematic Bacteriology* **43**, 626-629.

Sneath, P. H. A. (1995). The potential of dead bacterial specimens for systematic studies. *International Journal of Systematic Bacteriology* **45**, 188-189.

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. & Langham, C. D. (1989). OUTLIER: a BASIC program for detecting outlying members of multivariate clusters based on presence-absence data. *Computers and Geosciences* **15**, 939-964.

Sneath, P. H. A. & Sokal, R. R. (1973). Numerical taxonomy: the principles and practice of numerical classification. W. H. Freeman, Baltimore.

Snijders, N. L. (1924). Cavia-scheefkopperij, een nocardiose. *Geneeskundig Tijdschrift voor Nederlandsch-Indie* **64**, 85-87.

Snyder, E. E. & Stormo, G. D. (1993). Identification of coding regions in genomic DNA sequences: An application of dynamic programming and neural networks. *Nucleic Acids Research* **21**, 607-613.

Sobel, E. & Martinez, H. M. (1986). A multiple seqeunce alignment program. *Nucleic Acid Research* **14**, 363-374.

Sober, E. (1993). Experimental tests of phylogenetic inference methods. *Systematic Biology* **42**, 85-89.

Sobieski, J. M., Chen, K. N., Filiatreau, J. C., Pickett, M. H. & Fox, G. E. (1984). 16S rRNA oligonucleotide catalog data base. *Nucleic Acids Research* **12**, 141-148.

Soddell, J. A. & Seviour, R. J. (1990). Microbiology of foaming in activated sludge plants. *Journal of Applied Bacteriology* **65**, 145-176.

Soddell, J. A. & Seviour, R. J. (1994). Incidence and morphological variability of *Nocardia pinensis* in Australian activated sludge. *Water Research* **28**, 2343-2351.

Soddell, J., Knight, G., Strachan, W. & Seviour, R. (1992). Nocardioforms, not *Nocardia* foams. *Water Science and Technology* **26**, 455-460.

Sokal, R. R. & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *Kansas University Science Bulletin* **38**, 1409-1438.

Sokal, R. R. & Rolf, F. J. (1962). The comparison of dendrograms by objective methods. *Taxon* **XI**, 33-40.

Sourdis, J. & Nei, M. (1988). Relative efficiencies of the maximum parsimony and distance-matrix methods in obtaining the correct phylogenetic tree. *Molecular Biology and Evolution* **5**, 298-311.

Spring, S., Amann, R., Ludwig, W., Schleifer, K. H., Schüler, S., Poralla, K. & Petersen, N. (1994). Phylogenetic analysis of uncultured magnetotactic bacteria from the α-subclass *Proteobacteria*. *Systematic and Applied Microbiology* **17**, 501-508.

St-Laurent, L., Bousquet, J., Simon, L. & Lalonde, M. (1987). Seperation of various *Frankia* strains in the *Alnus* and *Elaeagnus* host specificity groups using sugar analysis. *Canadian Journal of Microbiology* **33**, 764-772.

Stack, M. V., Donoghue, H. D. & Tyler, J. E. (1978). Discrimination between oral streptococci by pyrolysis gas-liquid chromatography. *Applied Environmental Microbiology* **35**, 45-50.

Stack, M. V., Donoghue, H. D. & Tyler, J. E. (1981). Differentiation of *Streptococcus mutans* serotypes by discriminant analysis of pyrolysis-gas chromatography data. *Journal of Analytical and Applied Pyrolysis* **3**, 221-227.

Stackebrandt, E. (1992). Unifying phylogeny and phenotypic diversity. In *The Prokaryotes*, 2nd edition, pp. 19-47. A. Balows, H. G. Trüper, M. Dworkin, W. Harder and K. H. Schleifer, eds. Springer-Verlag, New York.

Stackebrandt, E. & Goebel, B. M. (1994). A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic Bacteriology* **44**, 846-849.

Stackebrandt, E. & Liesack, W. (1993). Nucleic acids and classification. In *Handbook of new bacterial systematics*, pp. 152-194. M. Goodfellow & A. G. O'Donnell, eds. Academic Press: London.

Stackebrandt, E. & Ludwig, W. (1994). The importance of using outgroup reference organisms in phylogenetic studies: the *Atopobium* case. *Systematic and Applied Microbiology* **17**, 39-43.

Stackebrandt, E. & Scheifer, K. H. (1984). Molecular systematics of actinomycetes and related organisms. In *Biological, Biochemical and Biomedical Aspectes of Actinomycetes*, pp. 485-504. L. Ortiz-Oritiz, L. F. Bajalil & V. Yakoleff, eds. Academic Press: Orlando.

Stackebrandt, E. & Woese, C. R. (1979). A phylogenetic dissection of the family *Micrococcaceae*. *Current Microbiology* **2**, 317-322.

Stackebrandt, E. & Woese, C. R. (1981a). The evolution of prokaryotes. In *Molecular and Cellular Aspects of Microbial Evolution*, pp. 1-31. M. J. Carlile, J. F. Collins & B. E. B. Moseley, eds. Cambridge University Press: Cambridge.

Stackebrandt, E. & Woese, C. R. (1981b). Towards a phylogeny of the actinomycetes and related organisms. *Current Microbiology* **5**, 197-202.

Stackebrandt, E., Lewis, B. J. & Woese, C. R. (1980a). The phylogenetic structure of the coryneform group of bacteria. *Zentralblatt für Bakteriologie, Parasitenfunde, Infektionskrankheiten und Hygiene. I. Abteilung, Originale Reihe* **C1**, 137-149.

Stackebrandt, E., Häringer, M. & Schleifer, K. H. (1980b). Molecular evidence for the transfer of *Oerkovia* species into the genus *Cellulomonas*. *Archives of Microbiology* **127**, 179-185.

Stackebrandt, E., Wunner-Füssl, B., Fowler, V. J. & Schleifer, K. H. (1981). Deoxyribonucleic acid homologies and ribosomal ribonucleic acid similarities among the sporeforming members of the order *Actinomycetales*. *International Journal of Systematic Bacteriology* **31**, 420-431.

Stackebrandt, E., Ludwig, W., Seewaldt, E. & Schleifer, K. H. (1983). Phylogeny of sporeforming members of the order *Actinomycetales*. *International Journal of Systematic Bacteriology* **33**, 173-180.

Stackebrandt, E., Murray, R. G. E. & Trüper, H. G. (1988a). *Proteobacteria* classis nov., a name for the phylogenetic taxon that includes the purple bacteria and their relatives. *International Journal of Systematic Bacteriology* **38**, 321-325.

Stackebrandt, E., Smida, J. & Collins, M. D. (1988b). Evidence of phylogenetic heterogeneity within the genus *Rhodococcus*: revival of the genus *Gordona* (Tsukamura). *Journal of General and Applied Microbiology* **34**, 341-348.

Stackebrandt, E., Liesack, W. & Witt, D. (1992). Ribosomal RNA and rDNA sequence analyses. *Gene* **115**, 255-260.

Stackebrandt, E., Liesack, W. & Goebel, B. M. (1993). Bacterial diversity in a soil sample from a subtropical Australian environment as determined by 16S rRNA analysis. *FASEB Journal* **7**, 232-236.

Stahl, D. A. & Amman, R. (1991). Development and application of nucleic acid probes. In *Nucleic Acid Techniques in Bacterial Systematics*, pp. 205-248. E. Stackebrandt & M. Goodfellow eds. John Wiley and Sons: Chichester.

Stahl, D. A. & Urbance, J. W. (1990). The division between fast- and slow-growing species corresponds to natural relationships among the mycobacteria. *Journal of Bacteriology* **172**, 116-124.

Stahl, D. A., Lane, D. J., Olsen, G. J. & Pace, N. R. (1984). Analysis of hydrothermal vent-associated symbionts by ribosomal RNA sequences. *Science* **224**, 409-411.

Stahl, D. A., Key, R., Flesher, B. & Smit, J. (1992). The phylogeny of marine and freshwater caulobacters reflects their habitat. *Journal of Bacteriology* **174**, 2193-2198.

Stanier, R. Y. & van Niel, C. B. (1941). The main outlines of bacterial classification. *Journal of Bacteriology* **42**, 437-466.

Stead, D. E. (1992). Grouping of plant-pathogenic and some other *Pseudomonas* spp. by using cellular fatty acid profiles. *International Journal of Systematic Bacteriology* **42**, 281-295.

Steel, M. A., Lockhart, P. J. & Penny, D. (1993). Confidence in evolutionary trees from biological sequence data. *Nature* **364**, 440-442.

Steinhaus, E. A. (1941). A study of the bacteria associated with thirty species of insects. *Journal of Bacteriology* **42**, 757-790.

Stoecker, M. A., Herwig, R. P. & Staley, J. T. (1994). *Rhodococcus zopfii* sp. nov., a toxicant-degrading bacterium. *International Journal of Systematic Bacteriology* **44**, 106-110.

Stroustrup, B. J. (1991). The C++ Programming Language. Prentice Hall: New York.

Suzuki, K., Goodfellow, M. & O'Donnell, A. G. (1993). Cell envelopes and classification. In *Handbook of New Bacterial Systematics*, pp. 195-250. M. Goodfellow & A. G. O'Donnell eds. Academic Press: London.

Suzuki, Y., Nagata, A., Ono, Y. & Yamada, T. (1988a). Complete nucleotide sequence of the 16S rRNA gene of *Mycobacterium bovis* BCG. *Journal of Bacteriology* **170**, 2886-2889.

Suzuki, Y., Ono, Y., Nagata, A. & Yamada, T. (1988b). Molecular cloning and characterization of an rRNA operon in *Streptomyces lividans* TK21. *Journal of Bacteriology* **170**, 1631-1636.

Swofford, D. L. & Olsen, G. J. (1990). Phylogenetic reconstruction. In *Molecular systematics*, pp. 411-501. D. Hillis & C. Moritz, eds. Sinauer Associates, Sunderland (USA).

Tabor, W. A. (1976). Wastewater microbiology. *Annual Reviews of Microbiology* **30**, 263-277.

Tago, Y. & Aiba, K. (1977). Exocellular mucopolysaccharide closely related to bacterial floc formation. *Appiled and Environmental Microbiology* **34**, 308-314.

Tajima, F. & Nei, M. (1984). Estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution* **1**, 269-285.

Takamiya, A. & Tubaki, K. (1956). A new form of streptomycetes capable of growing autotrophically. *Archic für Mikrobiologie* **25**, 58-64.

Takeuchi, M. & Yokota, A. (1994). Phylogenetic analysis of *Kineococcus aurantiacus* based on 16S rRNA gene sequences. *FEMS Microbiology Letters* **116**, 7-12.

Takewaki, S. I., Okuzumi, K., Manabe, I., Tanimura, M., Miyamura, K., Nakahara, K. I., Yazaki, Y., Ohkubo, A. & Nagai, R. (1994). Nucleotide sequence comparison of the mycobacterial *dnaJ* gene and PCR-restriction fragment length polymorphism analysis for identification of mycobacterial species. *International Journal of Systematic Bacteriology* **44**, 159-166.

Tamaoka, J. (1994). Determination of DNA base composition. In *Chemical Methods in Prokaryotic Systematics*, pp. 463-470. M. Goodfellow & A. G. O'Donnell, eds. John Wiley & Sons: Chichester.

Tamura, T. & Yokota, A. (1994). Transfer of *Nocardioides fastidiosa* Collins and Stackebrandt 1989 to the genus *Aeromicrobium* as *Aeromicrobium fastidiosum* comb. nov. *International Journal of Systematic Bacteriology* **44**, 608-611.

Tanabe, K., Tamura, T. & Uesaka, H. (1992). Neural network system for the identification of infrared spectra. Applied *Spectroscopy* **46**, 807-810.

Tanner, R., Stackebrandt, E, Fox, G. E. and Woese, C. R. (1981). A phylogenetic analysis of *Acetobacterium woodii, Clostridium barkeri, Clostridium butyricum, Clostridium lituseburense, Eubacterium limosum,* and *Eubacterium tenue. Current Microbiology* **5**, 35-38.

Taschke, C., Klinkert, M. Q., Wolters, J. & Herrmann, R. (1986). Organization of the ribosomal RNA genes in *Mycoplasma hyopneumoniae*: the 5S rRNA gene is separated from the 16S and 23S rRNA genes. **205**, 428-433.

Taylor, E. B., Beacham, T. D. & Kaeriyama, M. (1994). Population structure and identification of north pacific-ocean chum salmon (*Oncorhynchus-keta*) revealed by an analysis of minisatellite DNA variation. *Canadian Journal of Fisheries and Aquatic Sciences* **51**, 1430-1442.

Taylor, W. R. (1987). Multiple sequence alignment by a pairwise algorithm. *Computer Applications in Biological Sciences* **3**, 81-87.

Thiemann, J. E., Zucco, G. & Pelizza, G. (1969). A proposal for the transfer of *Streptomyces mediterranei* Margalith and Beretta 1960 to the genus *Nocardia* as *Nocardia mediterranei* (Margalith and Beretta) comb. nov. *Archic für Mikrobiologie* **67**, 147-155.

Tilford, P. E. (1936). Fasciation of sweet peas caused by *Phytomonas fascians* n. sp. *Journal of Agricultural Research* **53**, 383-394.

Tillier, E. R. M. (1994). Maximum likelihood with multiparameter models of substitution. *Journal of Molecular Evolution* **39**, 409-417.

Tindall, B. J., Wray, V., Huber, R. & Collins, M. D. (1991). A novel, fully saturated cyclic menaquinone in the archaebacterium *Pyrobaculum organotrophum. Systematic and Applied Microbiology* **14**, 218-221.

Tomiyasu, I. & Yano, I. (1984). Separation and analysis of novel polyunsaturated mycolic acids from psychrophilic, acid-fast bacterium, *Gordona aurantiaca*. **139**, 173-180.

Tomiyasu, I., Yoshinaga, J. Y., Kurano, F., Kato, Y., Kaneda, K., Imaizumi, S. & Yano, I. (1986). Occurrence of a novel glycolipid, 'trehalose 2,3,6'-trimycolate' in a psychrophilic, acid-fast bacterium, *Rhodococcus aurantiacus* (*Gordona aurantiaca*). *FEBS Letters* **203**, 239-242.

Trevisan, V. (1889). *Generiele Specie delle Batteriacee*. Zanaboni and Gabuzzi: Milano.

Trujillo, M. E. (1994). *Taxonomic Revision of the Genus Actinomadura and Related Taxa Using Rapid Methods*. Ph.D. Thesis, University of Newcastle upon Tyne.

Tsukamura, M. (1971). Proposal of a new genus, *Gordona*, for slightly acid-fast organisms occuring in sputa of patients with pulmonary disease and in soil. *Journal of General Microbiology* **68**, 15-26.

Tsukamura, M. (1974). A further numerical taxonomic study of the *rhodochrous* group. *Japanese Journal of Microbiology* **18**, 37-44.

Tsukamura, M. (1978). Numerical classification of *Rhodococcus* (formerly *Gordona*) organisms recently isolated from sputa of patients: Description of *Rhodococcus sputi* Tsukamura sp. nov. *International Journal of Systematic Bacteriology* **28**, 169-181.

Tsukamura, M. (1982). Numerical analysis of the taxonomy of nocardiae and rhodococci. Division of *Nocardia asteroides sensu stricto* into two species and desciptions of *Nocardia paratuberculosis* sp. nov. Tsukamura (formerly Kyoto-I group of Tsukamura), *Nocardia nova* sp. nov. Tsukamura, *Rhodococcus aichiensis* sp. nov. Tsukamura, *Rhodococcus chubuensis* sp. nov. Tsukamura, and *Rhodococcus obuensis* sp. nov. Tsukamura. *Microbiology Immunology* **26**, 1101-1119.

Tsukamura, M. (1991). Some addenda on the species *Gordona sputi* (Tsukamura 1978) Stackbrandt *et al.* 1989 comb. nov. *International Journal of Systematic Bacteriology* **41**, 569-570.

Tsukamura, M. & Mizuno, S. (1971). A new species, *Gorodna aurantiaca*, occurring in sputa of patients with pulmonary disease. *Kekkaku* **46**, 93-98 (In Japanese).

Tsukamura, M. & Yano, I. (1985). *Rhodococcus sputi* sp. nov., nom. rev. and *Rhodococcus aurantiacus* sp. nov. nom. rev. **35**, 364-368.

Turfitt, G. E. (1944). Microbiological agencies in the degradation of steroids. I. The Cholesterol-decomposing organisms in soil. *Journal of Bacteriology* **47**, 487-493.

Uchida, K. & Aida, K. (1977). Acyl type of bacterial cell wall: its simple identification by colorimetric method. *Journal of General and Applied Microbiology* **23**, 249-260.

Uchida, K. & Aida, K. (1984). An improved method for the glycolate test for simple identification of the acyl type of bacterial cell walls. *Journal of General and Applied Microbiology* **30**, 131-134.

Uemori, T., Asada, K., Kato, I. & Harasawa, R. (1992). Amplification of the 16S-23S spacer region in rRNA operons of mycoplasmas by the ploymerase chain reaction. *Systematic and Applied Microbiology* **15**, 181-186.

Ulbrich, N., Kumagai, I. & Erdmann, V. A. (1984). The number of ribosomal RNA genes in *Thermus thermophilus* HB8. *Nucleic Acids Research* **12**, 2055-2060.

Vainshtein, M., Hippe, H. & Kroppenstedt, R. M. (1992). Cellular fatty acid composition of *Desulfovibrio* species and its use in classification of sulfate-reducing bacteria. *Systematic and Applied Microbiology* **15**, 554-566.

Van de Peer, Y., Neefs, J. & De Watcher, R. (1990). Small ribosomal subunit RNA sequences, evolutionary relationships among different life forms, and mitochondrial origins. *Journal of Molecular Evolution* **30**, 463-476.

Van de Peer, Y., Hendriks, L., Goris, A., Neefs, J. M., Vancanneyt, M., Kersters, K., Berny, J. F., Hennebert, G. L. & De Wachter, R. (1992). Evolution of basidiomycetous yeasts as deduced from small ribosomal subunit RNA sequences. *Systematic and Applied Microbiology* **15**, 250-258.

Van de Peer, Y. & De Wachter, R. (1993). TREECON: a software package for the construction and drawing of evolutionary trees. *Computer Applications in Biological Sciences* **9**, 177-182.

Van de Peer, Y., Neefs, J. M., De Rijk, P., De Vos, P. & De Wachter, R. (1994). About the order of divergence of the major bacterial taxa during evolution. *Systematic and Applied Microbiology* **17**, 32-38.

Van Den Eynde, H., De Baere, R., Shah, H. N., Gharbia, S. E., Fox, G. E., Michalik, J., Van De Peer, Y. & De Watcher, R. (1990). 5S ribosomal ribonucleic acid sequences of *Bacteroides* and *Fusobacterium*: evolutionary relationships within these genera and among eubacteria. *International Journal of Systematic Bacteriology* **39**, 78-84.

Van Niekerk, A. M., Malea, A., Kawahigashi, J. & Reichlin, D. (1987). Foaming in anaerobic digesters: A survey and laboratory investigation. *Journal of Water Pollution Control Federation* **59**, 249.

Van Niel, C. B. (1946). The classification and natural relationships of bacteria. *Cold Spring Harbor Symposium on Quantitative Biology* **11**, 285-301.

Van Wezel, G. P., Krab, I. M., Douthwaite, S., Bibb, M. J., Vijgenboom, E. & Bosch, L. (1994). Transcription analysis of the i Streptomyces coelicolor A3(2) *rrnA* operon. *Microbiology* **140**, 3357-3365.

Vandamme, P., Bernardet, J. F., Segers, P., Kersters, K. & Holmes, B. (1994). New perspectives in the classification of the flavobacteria: description of *Chryseobacterium* gen. nov., *Bergeyella* gen. nov., and *Empedobacter* nom. rev. *International Journal of Systematic Bacteriology* **44**, 827-831.

Vauterin, L. & Vauterin, P. (1992). Computer-aided objective comparison of electrophoresis patterns for grouping and identification of microorganisms. *European Microbiology* **1**, 37-41.

Vauterin, L., Swings, J. & Kersters, K. (1991). Grouping of *Xanthomonas campestris* pathovars by SDS-PAGE of proteins. *Journal of General Microbiology* **137**, 1677-1687.

Viale, A. M., Arakaki, A. K., Soncini, F. C. & Ferreyra, R. G. (1994). Evolutionary relationships among eubacterial groups as inferred from GroEL (chaperonin) sequence comparisions. *International Journal of Systematic Bacteriology* **44**, 527-533.

Vickers, J. C., Williams, S. T. & Ross, G. W. (1984). A taxonomic approach to selective isolation of streptomycetes from soil. In, *Biochemical and Biomedical Aspects of Actinomycetes*, pp. 553-561. L. Ortiz-Ortiz, L. F. Bojalil & V. Yakoleff eds. Academic Press: Orlando.

Vieth, M. & A. Kolinski (1991). Prediction of protein secondary structure by an enhanced neural network. *Acta Biochimica Polonia* **38**, 335-351.

Vohra, G. B., Golding, G. B., Tsao, N. & Pearlman, R. E. (1992). A phylogenetic analysis based on the gene encoding phosphoglycerate kinase. *Journal of Molecular Evolution* **34**, 383-395.

Von Graevenitz, A & Krech, T. (1992). The genus *Corynebacterium*-Medical. In *The Prokaryotes*, 2nd edition, pp. 1172-1187. A. Balows, H. G. Trüper, M. Dworkin, W. Harder and K. H. Schleifer, eds. Springer-Verlag, New York.

Voorhees, K. J., Durfee, S. L. & Updegraff, D. M. (1988). Identification of diverse bacteria grown under diverse conditions using pyrolysis-mass spectrometry. *Journal of Microbiological Methods* **8**, 315-325.

Wagner, H., Horhammer, L. & Wolff, L. (1961). Dunnschicht-chromatographic von Phosphatiden und Glycolipiden. *Biochimische Zeitschrift* **334,** 175-184.

Wagner, M., Amann, R., Kämpfer, P., Assmus, B., Hartmann, A., Hutzler, P., Springer, N. & Schleifer, K. H. (1994a). Identification and *in situ* detecttion of Gram-negative filamentous bacteria in activated sludge. *Systematic and Applied Microbiology* **17**, 405-417.

Wagner, M., Amann, R., Lemmer, H., Manz, W. & Schleifer, K. H. (1994b). Probing of activated sludge with fluoroscently labeled rRNA targeted oligonucleotides. *Water Science and Technology* **29**, 15-23.

Wagner, M., Erhart, R., Manz, W., Amman, R., Lemmer, H., Wedi, D. & Schleifer, K. H. (1994c). Developement of an rRNA-targeted oligonucleotide probe specific for the genus *Acinetobacter* and its application for *in situ* monitoring in activated sludge. *Applied Environmental Microbiology* **60**, 792-800.

Waksman, S. A. (1961). *The Actinomycetes. Classification, identification and descriptions of genera and species.*, vol. 2. Williams and Wilkins Compan: Baltimore.

Waksman, S. A. (1967). *The Actinomycetales: A Summery of Current Knowledge.* The Ronald Press Company: New York.

Waksman, S. A. & Henrici, A. T. (1943). The nomenclature and classification of the actinomycetes. *Journal of Bacteriology* **46**, 337-341.

Waksman, S. A. & Henrici, A. T. (1948). Family II. *Actinomycetaceae* Buchanan. In *Bergey's Manual of Determinative Bacteriology, Volume 6.* pp. 892-928. R. S. Breed, E. G. D. Murray & A. P. Hitchens, eds. Williams and Wilkins Company: Baltimore.

Wallace, R. J., Jr., Brown, B. A., Tsukamura, M., Brown, J. M. & Onyi, G. O. (1991). Clinical and laboratory features of *Nocardia nova*. *Journal of Clinical Microbiology* **29**, 2407-2411.

Wampler, T. P. & Levy, E. J. (1987). Reproducibility in pyrolysis, recent developments. *Journal of Applied and Analytical Pyrolysis* **12**, 75-82.

Warwick, S., Bowen, T., McVeigh, H. & Embley, T. M. (1994). A phylogenetic analysis of the family *Pseudonoacrdiaceae* and the genera *Actinokineospora* and *Saccharothrix* with 16S rRNA sequences and a proposal to combine the genera *Amycolata* and *Pseudonocardia* in an emended genus *Pseudonocardia*. *International Journal of Systematic Bacteriology* **44**, 293-299.

Wayne, L. G., Krichevsky, E. G., Love, L. L., Johnson, R. & Krichevsky, M. I. (1980). Taxonomic probability matrix for use with slow growing mycobacteria. *International Journal of Systematic Bacteriology* **30**, 528-538.

Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., Moore, L. H., Moore, W. E. C., Murray, R. G. E., Stackebrandt, E., Starr, M. P. & Trüper, H. G. (1987). Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics. *International Journal of Systematic Bacteriology* 37: 463-464.

Weigel, U. M. & Herges, R. (1992). Automatic interpretation of infrared spectra - recognition of aromatic substitution patterns using neural networks. J. Chem. Inform. Comp. Sci. 32, 723-731.

Weijman, A. C. M. (1977). The application of Curie-point pyrolysis mass spectrometry in fungal taxonomy. In *Analytical Pyrolysis*, pp. 225-233. C. E. R. Jones & C. A. Cramers, eds. Elsevier, Amsterdam.

Weisburg, W. G., Barns, S. M., Pelletier, D. A. & Lane, D. J. (1991). 16S ribosomal DNA amplification for phylogenetic study. *Journal of Bacteriology* **173**, 697-703.

Welch, B. L. (1947). The generalization of 'Students' probelm when several different population variances are involved. *Biometrika* **34**, 28-35.

Wellington, E. M. H. & Williams, S. T. (1978). Preservation of actinomycete inoculum in frozen glycerol. *Microbios Letters* **6**, 151-159.

Wellington, E. M. H., Stackebrandt, E., Sanders, D., Wolstrup, J. & Jorgensen, N. O. G. (1992). Taxonomic status of *Kitasatosporia*, and proposed unification with *Streptomyces* on the basis of phenotypic and 16S rRNA analysis and emendation of *Streptomyces* Waksman and Henrici 1943, 339AL. *International Journal of Systematic Bacteriology* **42**, 156-160.

Werbos, P. (1974) *Beyond Regression: New Tools for Prediction and Analysis in the Behaviorial Sciences.* Ph. D. Thesis, Harvard University.

Werther, W., Lohninger, H., Varmuza, K. & Stancl, F. (1994). Classification of mass spectra - a comparison of yes/no classification methods for the recognition of simple structural properties. *Chemometrics And Intelligent Laboratory Systems* **22**, 63-76.

Wheeler, W. C. & Honeycutt, R. L. (1988). Paired sequence difference in ribosomal RNAs: evolutionary and phylogenetic implications. *Molecular Biology and Evolution* **5**, 90-96.

White, D., Sharp, R. J. & Priest, F. G. (1993). A polyphasic taxonomic study of thermophilic bacilli from a wide geographical area. *Antonie van Leeuwenhoek* **64**, 357-386.

White, H. (1990). Connectional nonparametric regression: Multilayer feedforward networks can learn arbitrary mappings. *Neural Networks* **3**, 535-549.

Whitham, T. S., Athalye, M., Minnikin, D. E. & Goodfellow, M. (1993). Numerical and chemical classificaiton of *Streptosporangium* and some related actinomycetes. *Antonie van Leeuwenhoek* **64**, 387-429.

Wieten, G., Haverkamp, J., Engel, W. W. B. & Tarnok, I. (1979). Pyrolysis mass spectrometry in mycobacterial taxonomy and identification. In *Twenty-five Years of Mycobacterial Taxonomy.* pp. 171-189. G. P. Kubica, L. G. Wayne & I. S. Good, eds. CDC Press: Atlanta.

Wieten, G., Haverkamp, J., Berwald, L. G., Groothuis, D. G. & Draper, P. (1981). Pyrolysis mass spectrometry: a new method to differentiate between the mycobacteria of the '*tuberculosis*' complex and other mycobacteria. *Journal of General Microbiology* **122**, 109-118.

Wieten, G., Haverkamp, J., Berwald, L. G., Groothuis, D. G. & Draper, P. (1982). Pyrolysis mass spectrometry: its applicability to mycobacteriology, including *Mycobacterium leprae. Annales de Microbiologie* **133B**, 15-27.

Wieten, G., Haverkamp, J., Groothuis, D. G., Berwald, L. G. & David, H. L. (1983). Classification and identification of *Mycobacterium africanum* by pyrolysis mass spectrometry. *Journal of General Microbiology* **129**, 3679-3688.

Wilding, P., Morgan, M. A., Grygotis, A. E., Shoffner, M. A. & Rosato, E. F. (1994). Application of backpropagation neural networks to diagnosis of breast and ovarian-cancer. *Cancer Letters* **77**, 145-153.

Wilkins, M. F., Morris, C. W. & Boddy, L. (1994). A comparison of radial basis function and backpropagation neural networks for identification of marine phytoplankton from multivariate flow cytometry data. *Computer Applications in the Biosciences* **10**, 285-294.

Wilkinson, S. R., Young, M., Goodacre, R., Morris, J. G., Farrow, J. A. E. & Collins, M. D. (1995). Phenotypic and genotypic differences between certain strains of *Clostridium acetobutylicum. FEMS Microbiology Letters* **125**, 199-204.

Willcox, W. R., Lapage, S. P., Bascomb, S. & Curtis, M. A. (1973). Identification of bacteria by computer: theory and programming. *Journal of General Microbiology* **77**, 317-330.

Willems, A. & Collins, M. D. (1994). Phylogenetic placement of *Sarcina ventriculi* and *Sarcina maxima* within group I *Clostridium*, a possible problem for future revision of the genus *Clostridium*. *International Journal of Systematic Bacteriology* **44**, 591-593.

Williams, S. T., Sharples, G. P., Serrano, J. A., Serrano, A. A. & Lacey, J. (1976). The micromorphology and fine structure of nocardioform organisms. In *The Biology of Nocardiae*, pp. 102-140. M. Goodfellow, G. H. Brownell & J. A. Serrano, eds. Academic Press: London.

Williams, S. T., Goodfellow, M., Alderson, G., Wellington, E. M. H., Sneath, P. H. A. & Sackin, M. J. (1983a). Numerical classification of *Streptomyces* and related genera. *Journal of General Microbiology* **129**, 1743-1813.

Williams, S. T., Goodfellow, M., Wellington, E. M. H., Vicker, J. C., s,Alderson, G., Sneath, P. H. A., Sackin, M. J. & Mortimer, A. M. (1983b). A probability matrix for identification of some Streptomycetes. *Journal of General Microbiology* **129**, 1815-1830.

Williams, S. T., Vickers, J. C. & Goodfellow, M. (1985). Application of new theoretical concepts to the identification of streptomycetes. In *Computer-Assisted Bacterial Systematics*, pp. 289-306. M. Goodfellow, D. Jones & F. G. Priest, eds. Academic Press: London.

Williams, S. T., Goodfellow, M. & Alderson, G. (1989). Genus *Streptomyces* Waksman and Henrici, 1943, 339^{AL}. In *Bergey's Manual of Systematic Bacteriology* vol. 4., pp. 2452-2492. S. T. Williams, M. E. Sharpe & J. G. Holt, eds. Williams & Wilkins: Baltimore.

Willis, M. J., Di Massimo, C., Montague, G. A., Tham, M. T. & Morris, A. J. (1991). Artificial neural networks in process engineering. *IEE Proceedings-D* **138**, 256-266.

Wilson, K. H., Blitchington, R., Frothingham, R. & Wilson, J. A. P. (1991). Phylogeny of the Whipple's-disease-associated bacterium. *Lancet* **338**, 474-475.

Wilson, M. E., Oyama, V. & Vango, S. P. (1962). Design features of a lunar gas chromatograph. In *Proceedings of 3rd International Symposium on Gas Chromatography*, pp. 329-338. N. Brenner, ed. Academic Press: New York.

Winder, F. G. & Rooney, S. A. (1970). Effects of nitrogeneous components of the medium on the carbohydrate and nucleic acid contents of *Mycobacterium tuberculosis* BCG. *Journal of General Microbiology* **63**, 29-39.

Windig, W., Kistemaker, P. G. Haverkamp, J., Meuzelaar, H. (1979). The effects of sample preparation, pyrolysis and pyrolysate transfer conditions on pyrolysis mass spctra. *Journal of Applied and Analytical Pyrolysis* **1**, 39-52.

Windig, W., Kistemaker, P. G. & Haverkamp, J. (1980). Factor analysis of the influence of changes in experimental conditions in pyrolysis-mass spectrometry. *Journal of Analytical and Applied Pyrolysis* **2**, 7-18.

Winker, S., Overbeek, R., Woese, C. R., Olsen, G. J. & Pfluger, N. (1990). Structure detection through automated covariance search. *Computer Applications in Biological Sciences* **6**, 365-371.

Winstanley, T. G., Magee, J. T., Limb, D. I., Hindmarch, J. M., Spencer, R. C., Whiley, R. A., Beighton, D. & Hardie, J. M. (1992). A numerical taxonomic study of the *'Streptococcus milleri'* group based upon conventional phenotypic tests and pyrolysis mass spectrometry. *Journal Medical Microbiology* **36**, 149-155.

Wishart, D. (1987). *Clustan User Manual.* Computing Laboratory University of St. Andrews, Scotland, UK.

Woese, C. R. (1987). Bacterial evolution. *Microbiological Review* **51**, 221-272.

Woese, C. R. (1992). Prokaryote systematics: the evolution of a science. In *The Prokaryotes*, 2nd edition, pp. 3-18. A. Balows, H. G. Trüper, M. Dworkin, W. Harder and K. H. Schleifer, eds. Springer-Verlag, New York.

Woese, C. R. & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of National Academic Society USA* **74**, 5088-5090.

Woese, C. R., Fox, G. E., Zablen, L., Uchida, T., Bonen, L., Pechman, K., Lewis, J. B. & Stahl, D. (1975). Conservation of primary structure in 16S ribosomal RNA. *Nature* **254**, 83-86.

Woese, C. R., Gibson, J. & Fox, G. E. (1980). Do genealogical patterns in purple photosynthetic bacteria reflect interspecific gene transfer ?. *Nature* **283**, 212-214.

Woese, C. R., Gutell, R., Gupta, R. & Noller, H. F. (1983). Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiological Review* **47**, 621-669.

Woese, C. R., Yang, D., Mandelco, L. & Steller, K. O. (1990a). The *Flexibacter-Flavobacter* connection. *Systematic and Applied Microbiology* **13**, 161-165.

Woese, C. R., Kandler, O. & Wheelis, M. L. (1990b). Towards a natural system for organisms: proposal for the domains Archaea, Bacteria and Eucarya. *Proceedings of National Academic Society USA* **87**, 4576-4579..

Woese, C. R., Achenbach, L., Rouviere, P. & Mandelco, L. (1991). Archaeal phylogeny: reexamination of the phylogenetic position of *Archaeoglobus fulgidus* in the light of certain composition-induced artifacts. *Systematic and Applied Microbiology* **14**, 364-371.

Wold, S. & Söderström, M. J. (1977). SIMCA: A method for analysing chemical data in terms of similarity and analogy. In *Chemometrics: Theory and Application*, pp. 243-282. B. Kowalski ed. Series no. 52, ACS Symposium Washington, DC: American Chemical Society.

Wu, C. H. (1993). Classification neural networks for rapid sequence annotation and automated database organization. *Computers and Chemistry* **17**, 219-227.

Xia, Y., Embley, T. M. & O'Donnell, A. G. (1994). Phylogenetic analysis of *Azospirillum* by direct sequencing of PCR amplified 16S rDNA. *Systematic and Applied Microbiology* **17**, 197-201.

Yamada, Y., Hou, C. F., Sasaki, J., Tahara, Y. & Yoshioka, H. (1982). The structure of the octahydrogenated isoprenoid side-chain menaquinone with nine isoprene units isolated from *Streptomyces albus*. *Journal of General and Applied Microbiology* **28**, 519-529.

Yamamoto, S. & Harayama, S. (1995). PCR Amplification and direct sequencing of *gyrB* genes with universal primers and their application to the detection and taxonomic analysis of *Pseudomonas putida* strains. *Applied Environmental Microbiology* **61**, 1104-1109.

Yano, I., Imaeda, T. & Tsukamura, M. (1990). Characteriztion of *Nocardia nova*. *International Journal of Systematic Bacteriology* **40**, 170-174.

Yurkov, V., Stackebrandt, E., Holmes, A., Fuerst, J. A., Hugenholtz, P., Golecki, J., Gad'On, N., Gorlenko, V. M., Kompantseva, E. I. & Drews, G. (1994). Phylogenetic positions of novel aerobic, bacteriochlorophyll *a*-containing bacteria and description of *Roseococcus thiosulfatophilus* gen. nov., sp. nov., *Erythromicrobium ramosum* gen. nov., sp. nov., and *Erythrobacter litoralis* sp. nov. *International Journal of Systematic Bacteriology* **44**, 427-434.

Zakrzewska-Czerwinska, J., Mordarski, M. & Goodfellow, M. (1988). DNA base composition and homology values in the classification of some *Rhodococcus* species. *Journal of General Microbiology* **134**, 2807-2813.

Zemany, P. D. (1952). Identification of complex organic materials by mass spectrometric analysis of their pyrolysis products. *Analytical Chemistry* **24**, 1709-1713.

Zillig, W., Klenk, H. P., Palm, P., Pühler, G., Gropp, F., Garrett, R. A. & Leffers, H. (1989). The phylogenetic relations of DNA-dependent RNA polymerase of archaebacteria, eukaryotes and eubacteria. *Canadian Journal of Microbiology* 35, 73-80.

Zopf, W. (1891). Uber Ausscheidung von Fellfarbstoffen (Lipochromen) setens gewisser Spattpilze. *Berichte der Deutschen Botanischen Gesellschaft* **9**, 22-28.

APPENDIX A
CULTURE MEDIA AND BUFFERS

1. *Bennett's Agar, modified* (Jones, 1949)

Yeast extract (Difco)	1.0 g
Lab-Lemco (Oxoid)	0.8 g
Bacto-Casitone (Difco)	2.0 g
Glucose	10.0 g
Agar (Difco)	12.0 g
	pH 7.3

The medium components were dissolved in 1 litre of cold distilled water and the pH adjusted with diluted (0.1 M) NaOH or HCl. The agar was then added and the medium liquefied by steaming prior to autoclaving.

2. *Carbon Source Utilisation Medium (Boiron et al., 1993)*

Ammonium Sulphate	2.64 g
Potassium Dihydrogen Phosphate	0.5 g
Magnesium Sulphate	0.5 g
Agar (Oxoid No. 1)	15.0 g
	pH 7.0

The medium was dissolved in 1 litre of distilled water and pH was adjusted by using 10% KOH solution. The agar was then autoclaved. The carbon source compounds were sterilised separately by using filtration or tyndallisation.

3. *Carbon and Nitrogen Source Utilisation Medium (Boiron et al., 1993)*

Potassium Dihydrogen Phosphate	0.5 g
Magnesium Sulphate	0.5 g
Agar (Oxoid No. 1)	15.0 g
	pH 7.0

The medium was dissolved in 1 litre of distilled water and pH was adjusted by using 10% KOH solution. The agar was then autoclaved. The carbon and nitrogen source compounds were sterilised separately by using filtration or tyndallisation.

4. Glucose Yeast-extract Agar (Gordon & Mihm, 1962)

Glucose	10.0 g
Yeast-extract	10.0 g
Agar (Difco)	18.0 g
	pH 7.0

The medium components were dissolved in 1 litre of cold distilled water and the pH adjusted with diluted (0.1 M) NaOH or HCl. The agar was then added and the medium liquefied by steaming prior to autoclaving.

5. Guanidine-Sarkosyl Solution (Pitcher et al., 1989)

Guanidine thiocyanate (Sigma)	60.0 g
EDTA (0.5 mM) pH 8	20 ml
Deionised water	20 ml

The preparation was heated at 65 °C until dissolved. After cooling, 5 ml of 10%, v/v sarkosyl (GES reagent) were added, the solution was made up to 100ml with deionised water, filtered through a 0.45 µm filter and stored at room temperature.

6. LB Agar (Sambrook et al., 1989)

Bacto-Tryptone	10 g
Bacto-Yeast extract	5 g
NaCl	5 g
Agar	15 g
	pH 7.5

The pH was adjusted with 1 N NaOH prior to autoclaving.

7. Non-Sporulating Agar (Sanglier et al., 1992)

Casaminoacids (Difco)	20.0 g
Soluble starch (BDH)	20.0 g
Yeast extract (Difco)	4.0 g
Agar (Difco)	18.0 g
	pH 6.5

The ingredients, apart from the starch and agar, were dissolved in 900 ml of distilled water and the pH adjusted with diluted (0.1 M) NaOH or HCl. The agar was then added and the preparation liquefied by steaming. The starch, which was made into a paste with 100 ml of cold distilled water, was incorporated into the hot medium with constant stirring prior to autoclaving.

8. Sauton's Broth, modified (Modarska et al., 1972)

L-Asparagine	5.0 g
Casamino acid	2.0 g
Glucose	15.0 g
Sodium Citrate	1.5 g
Potassium Dihydrogen Phosphate	5.0 g
Magnesium Sulphate	0.5 g
Potassium Sulphate	0.5 g
Ferric Ammonium Citrate	trace
	pH 7.2

The medium components were dissolved in 1 litre of cold distilled water and the pH adjusted with diluted (0.1 M) NaOH or HCl. The medium and glucose were autoclaved separately.

9. TFB I buffer

Potassium acetate, 30 mM; $CaCl_2$ $2H_20$, 10 mM; KCl, 100 mM; glycerol, 15% (v/v). Dissolved in 900 ml of water and autoclaved. Made up to 1 litre by adding 100 ml of autoclaved 500 mM $MnCl_2$.

10. TFB II buffer

$CaCl_2$ $2H_20$, 75 mM; KCl, 10 mM; glycerol, 15% (v/v). Dissolved in 900 ml of water and autoclaved. Made up to 1 litre by adding 100 ml of autoclaved 100 mM Na-MOPS (pH 7.0).

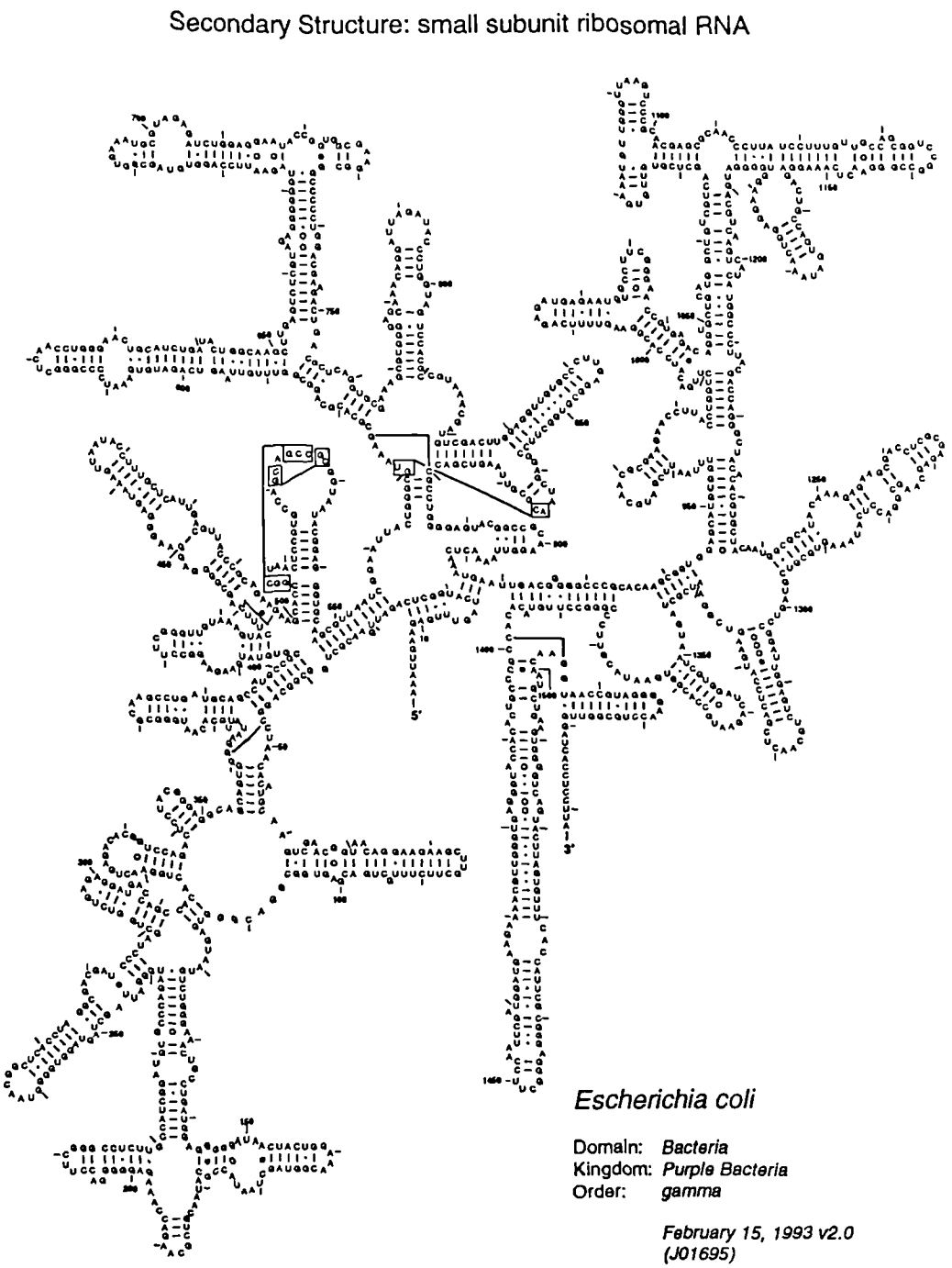
<BLANK>

APPENDIX B.

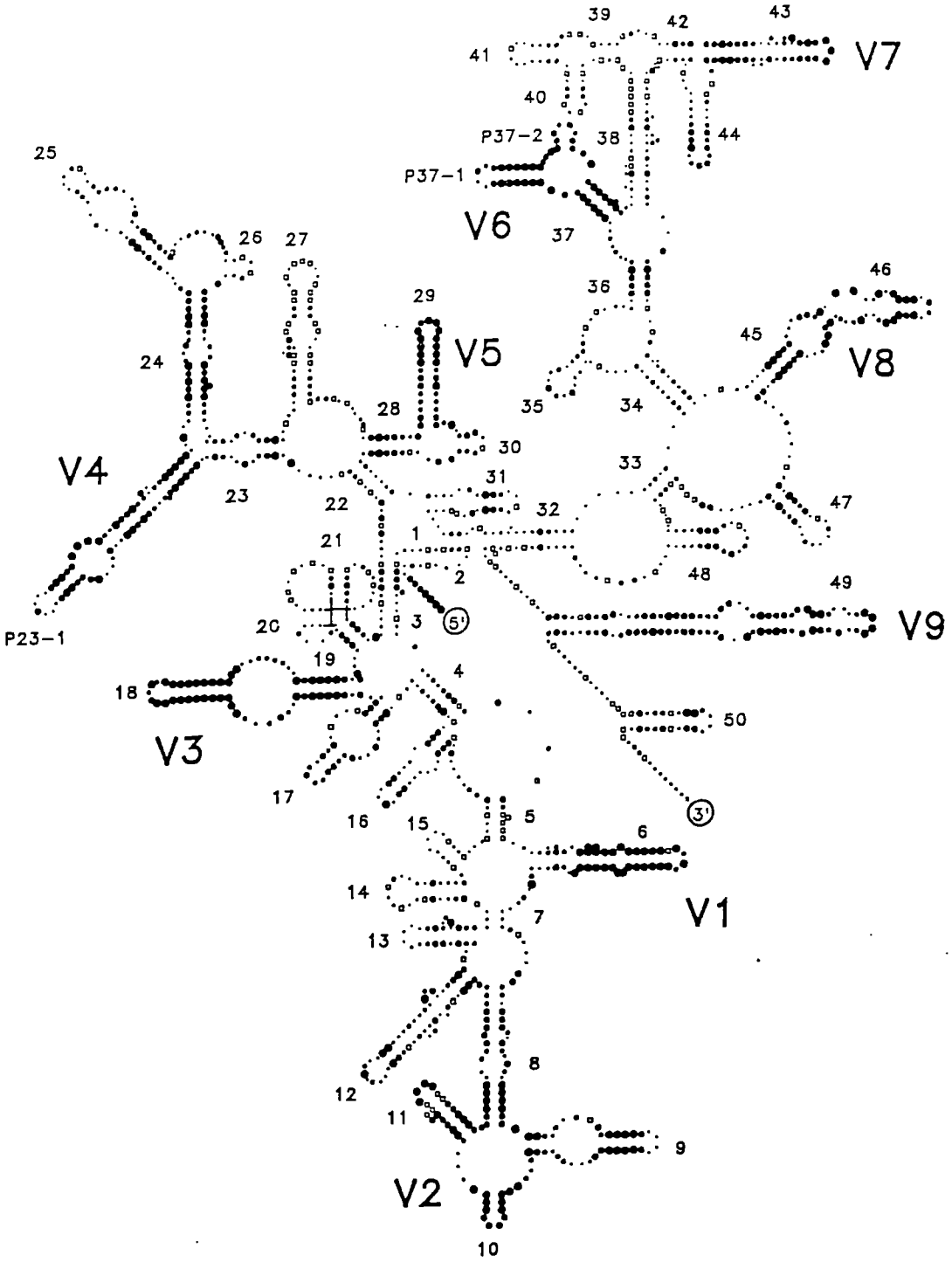
INTERNATIONAL UNION OF BIOCHEMISTRY NUCLEOTIDE CODE

Code	Degeneracy
R	A+G
Y	C+T
M	A+C
K	G+T
S	G+C
W	A+T
H	A+T+C
B	G+T+C
D	G+A+T
V	A+C+G
N	A+C+T+G

APPENDIX C. 16S rRNA NUMBERING SYSTEM BASED ON *ESCHERICHIA COLI* (BROSIUS *et al.*, 1978). OBTAINED FROM THE RIBOSOMAL DATABASE PROJECT (LARSEN *et al.*, 1993)



APPENDIX D. HELIX NOMENCLATURE BASED ON *ESCHERICHIA COLI* (NEEFS *et al.*, 1993).



APPENDIX E. 16S rRNA SEQUENCE SIMILARITY AND CORRESPONDING DNA RELATEDNESS VALUES FOR MYCOLIC ACID-CONTAINING ACTINOMYCETES[1]. THIS DATA WERE USED TO DRAW THE GRAPH SHOWN IN FIGURE 2-1 (PAGE 107).

Taxa compared[2]		16S rRNA	DNA	Reference[3]
Strain 1	Strain 2	Similarity	homology	
M. gastri	*M. kansasii*	100.00	62	Imaeda *et al.* 1988
N. restricta	*R. equi*	99.90	96	Rainey *et al.* 1995b
G. sputi	*G. aichiensis*	99.73	38	Zakrzewska-Czerwinska *et al.* 1988
M. malmoense	*M. szulgai*	99.60	36	Imaeda *et al.* 1988
R. erythropolis	*N. calcarea*	99.50	93	Rainey *et al.* 1995b
M. farcinogenes	*M. senegalense*	99.50	49	Imaeda *et al.* 1988
R. chubuensis	*G. sputi*	99.46	97	Riegel *et al.* 1994
T. paurometabola N663	*T. paurometabola* M334	99.38	44	Goodfellow *et al.* 1991
R. fascians	*R. luteus*	99.30	79	Klatte *et al.* 1994
M. tuberculosis	*M. marinum*	99.24	11	Imaeda, 1985
R. rhodochrous	*R. roseus*	99.20	80	Rainey *et al.* 1995b
M. avium	*M. intracelluare*	99.20	59	Imaeda *et al.* 1988
M. gastri	*M. malmoense*	99.20	46	Imaeda *et al.* 1988
M. kansasii	*M. szulgai*	99.00	45	Imaeda *et al.* 1988
M. marinum	*M. ulcerans*	99.00	86	Imaeda *et al.* 1988
M. chelonae	*M. fortuitum*	98.96	8	Lévy-Frébault *et al.* 1986
M. intracelluare	*M. malmoense*	98.90	48	Imaeda *et al.* 1988
M. intracelluare	*M. scrofulaceum*	98.90	48	Imaeda *et al.* 1988
M. scrofulaceum	*M. szulgai*	98.90	39	Imaeda *et al.* 1988
C. macginleyi	*C. accolens*	98.85	25	Riegel *et al.* 1995
M. asiaticum	*M. gordonae*	98.80	44	Imaeda *et al.* 1988
M. avium	*M. malmoense*	98.80	46	Imaeda *et al.* 1988
M. haemophilum	*M. malmoense*	98.80	55	Imaeda *et al.* 1988
M. malmoense	*M. tuberculosis*	98.80	53	Imaeda *et al.* 1988
M. tuberculosis	*M. kansasii*	98.71	32	Imaeda, 1985
M. avium	*M. gastri*	98.70	53	Imaeda *et al.* 1988
M. gastri	*M. intracelluare*	98.70	41	Imaeda *et al.* 1988
M. gastri	*M. scrofulaceum*	98.70	57	Imaeda *et al.* 1988
M. asiaticum	*M. szulgai*	98.66	30	Imaeda *et al.* 1988
M. gastri	*M. marinum*	98.60	51	Imaeda *et al.* 1988
M. kansasii	*M. scrofulaceum*	98.60	47	Imaeda *et al.* 1988
M. malmoense	*M. scrofulaceum*	98.60	42	Imaeda *et al.* 1988
M. avium	*M. marinum*	98.50	45	Imaeda *et al.* 1988
M. avium	*M. scrofulaceum*	98.50	41	Imaeda *et al.* 1988
M. avium	*M. tuberculosis*	98.40	30	Imaeda *et al.* 1988
C. sp. (CDC group-G)	*C. accolens*	98.35	11	Riegel *et al.* 1995
M. tuberculosis	*M. ulcerans*	98.30	34	Imaeda *et al.* 1988
M. asiaticum	*M. avium*	98.25	49	Imaeda *et al.* 1988

APPENDIX E. CONTINUED.

Taxa compared		16S rRNA	DNA	Reference
Strain 1	Strain 2	Similarity	homology	
M. tuberculosis	*M. intracelluare*	98.20	48	Imaeda, 1985
M. asiaticum	*M. malmoense*	98.20	52	Imaeda *et al.* 1988
M. tuberculosis	*M. asiaticum*	98.11	9	Imaeda, 1985
M. asiaticum	*M. gastri*	98.03	37	Imaeda *et al.* 1988
M. haemophilum	*M. marinum*	98.00	49	Imaeda *et al.* 1988
M. tuberculosis	*M. haemophilum*	97.95	17	Imaeda, 1985
M. tuberculosis	*M. scrofulaceum*	97.94	22	Imaeda, 1985
M. asiaticum	*M. scrofulaceum*	97.90	44	Imaeda *et al.* 1988
M. gordonae	*M. scrofulaceum*	97.80	47	Imaeda *et al.* 1988
M. asiaticum	*M. simae*	97.70	50	Imaeda *et al.* 1988
M. avium	*M. ulcerans*	97.70	62	Imaeda *et al.* 1988
N. asteroides	*N. farcinica*	97.55	28	Kudo *et al.* 1988
R. rhodochrous	*R. ruber*	97.50	39	Rainey *et al.* 1995b
M. asiaticum	*M. ulcerans*	97.50	53	Imaeda *et al.* 1988
N. asteroides	*N. nova*	97.46	39	Yano *et al.* 1990
N. asteroides	*N. nova*	97.46	23	Kudo *et al.* 1988
R. roseus	*R. ruber*	97.40	44	Rainey *et al.* 1995b
M. intracelluare	*M. ulcerans*	97.30	39	Imaeda *et al.* 1988
R. chubuensis	*G. bronchialis*	97.28	7	Riegel *et al.* 1994
M. malmoense	*M. simae*	97.10	59	Imaeda *et al.* 1988
M. chelonae	*M. chelonae ssp. abscessus*	97.00	26	Lévy-Frébault *et al.* 1986
M. haemophilum	*M. ulcerans*	97.00	47	Imaeda *et al.* 1988
M. avium	*M. shimoidei*	96.90	32	Imaeda *et al.* 1988
M. terrae	*M. tuberculosis*	96.90	49	Imaeda *et al.* 1988
C. macginleyi	*C. jeikeium*	96.89	1	Riegel *et al.* 1995
R. luteus	*R. globerulus*	96.86	12	Zakrzewska-Czerwinska *et al.* 1988
M. chelonae ssp. abscessus	*M. fortuitum*	96.84	5	Lévy-Frébault *et al.* 1986
M. gordonae	*M. ulcerans*	96.80	33	Imaeda *et al.* 1988
M. nonchromogenicum	*M. trivale*	96.80	46	Imaeda *et al.* 1988
R. chubuensis	*G. rubropertincta*	96.76	7	Riegel *et al.* 1994
M. malmoense	*M. shimoidei*	96.70	51	Imaeda *et al.* 1988
R. chubuensis	*G. terrae*	96.61	9	Riegel *et al.* 1994
M. asiaticum	*M. nonchromogenicum*	96.60	36	Imaeda *et al.* 1988
M. terrae	*M. trivale*	96.60	35	Imaeda *et al.* 1988
M. asiaticum	*M. triviale*	96.50	56	Imaeda *et al.* 1988
N. farcinica	*N. nova*	96.33	20	Yano *et al.* 1990
R. globerulus	*R. rhodnii*	96.33	18	Zakrzewska-Czerwinska *et al.* 1988
M. tuberculosis	*M. nonchromogenicum*	96.29	53	Imaeda, 1985
M. avium	*M. nonchromogenicum*	96.20	44	Imaeda *et al.* 1988

APPENDIX E. CONTINUED.

Taxa compared		16S rRNA	DNA	Reference
Strain 1	Strain 2	Similarity	homology	
M. malmoense	*M. terrae*	96.20	30	Imaeda *et al.* 1988
R. luteus	*R. rhodnii*	95.90	11	Zakrzewska-Czerwinska *et al.* 1988
M. kansasii	*M. senegalense*	95.90	37	Imaeda *et al.* 1988
M. avium	*M. farcinogenes*	95.80	36	Imaeda *et al.* 1988
M. trivale	*M. tuberculosis*	95.80	49	Imaeda *et al.* 1988
M. gastri	*M. triviale*	95.70	52	Imaeda *et al.* 1988
T. paurometabola N663	*T. wratislaviensis*	95.66	16	Goodfellow *et al.* 1991
M. intracelluare	*M. triviale*	95.60	43	Imaeda *et al.* 1988
M. malmoense	*M. trivale*	95.60	58	Imaeda *et al.* 1988
R. globerulus	*D. maris*	95.03	12	Zakrzewska-Czerwinska *et al.* 1988
C. macginleyi	*C. diphtheriae*	95.02	5	Riegel *et al.* 1995
M. intracelluare	*M. xenopi*	95.00	30	Imaeda *et al.* 1988
C. sp. (CDC group-G)	*C. jeikeium*	94.74	1	Riegel *et al.* 1995
T. paurometabola N663	*M. chlorophenolicum*	94.30	10	Goodfellow *et al.* 1991
G. sputi	*R. rhodnii*	94.29	11	Zakrzewska-Czerwinska *et al.* 1988
R. luteus	*D. maris*	94.11	8	Zakrzewska-Czerwinska *et al.* 1988
T. paurometabola N663	*G. branchialis*	93.80	9	Goodfellow *et al.* 1991
T. paurometabola N663	*G. aichiensis*	93.69	13	Goodfellow *et al.* 1991
G. sputi	*D. maris*	93.27	11	Zakrzewska-Czerwinska *et al.* 1988
R. globerulus	*G. aichiensis*	93.17	11	Zakrzewska-Czerwinska *et al.* 1988
G. sputi	*R. globerulus*	93.06	12	Zakrzewska-Czerwinska *et al.* 1988
R. luteus	*G. aichiensis*	92.67	8	Zakrzewska-Czerwinska *et al.* 1988
R. luteus	*G. sputi*	92.54	9	Zakrzewska-Czerwinska *et al.* 1988

Abbreviations: C., *Corynebacterium*; D, *Dietzia*; G., *Gordona*, M. *Mycobacterium*; N. *Nocardia*; R, *Rhodococcus*; and T., *Tsukamurella*.

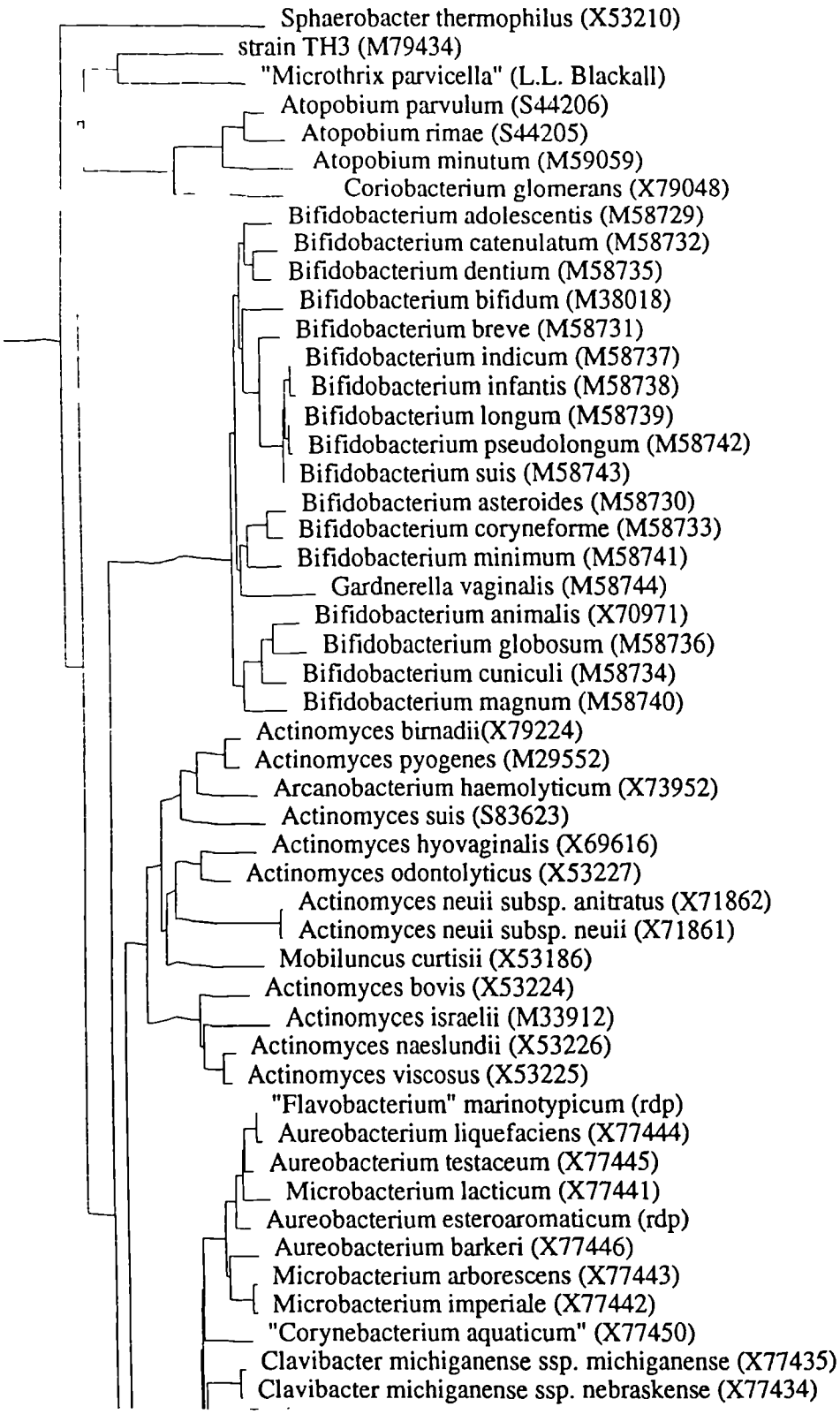
[1], Data were listed in order of 16S rRNA sequence similarity.

[2], Type strains unless specified.

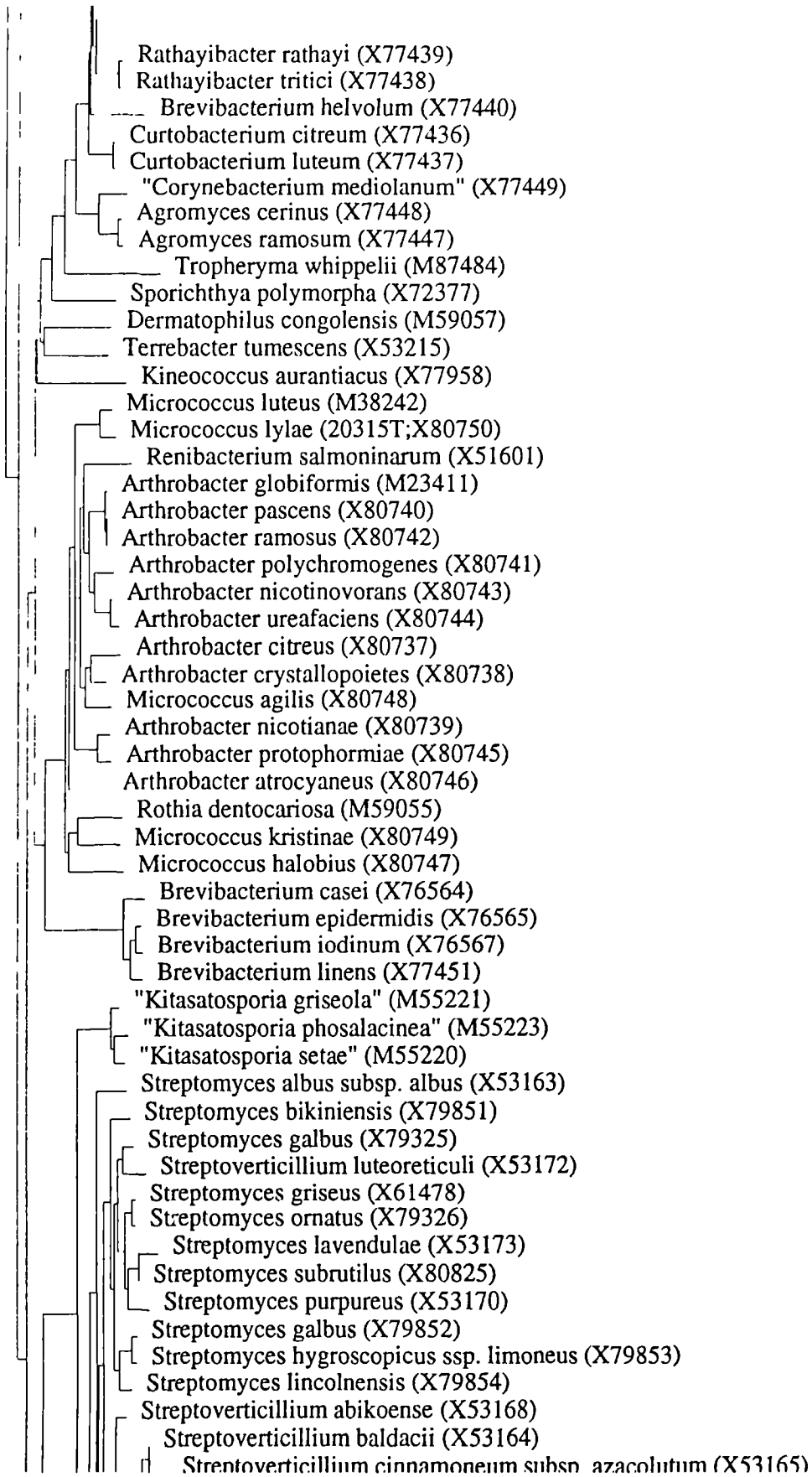
[3], References in which DNA pairing studies were carried out.

<BLANK>

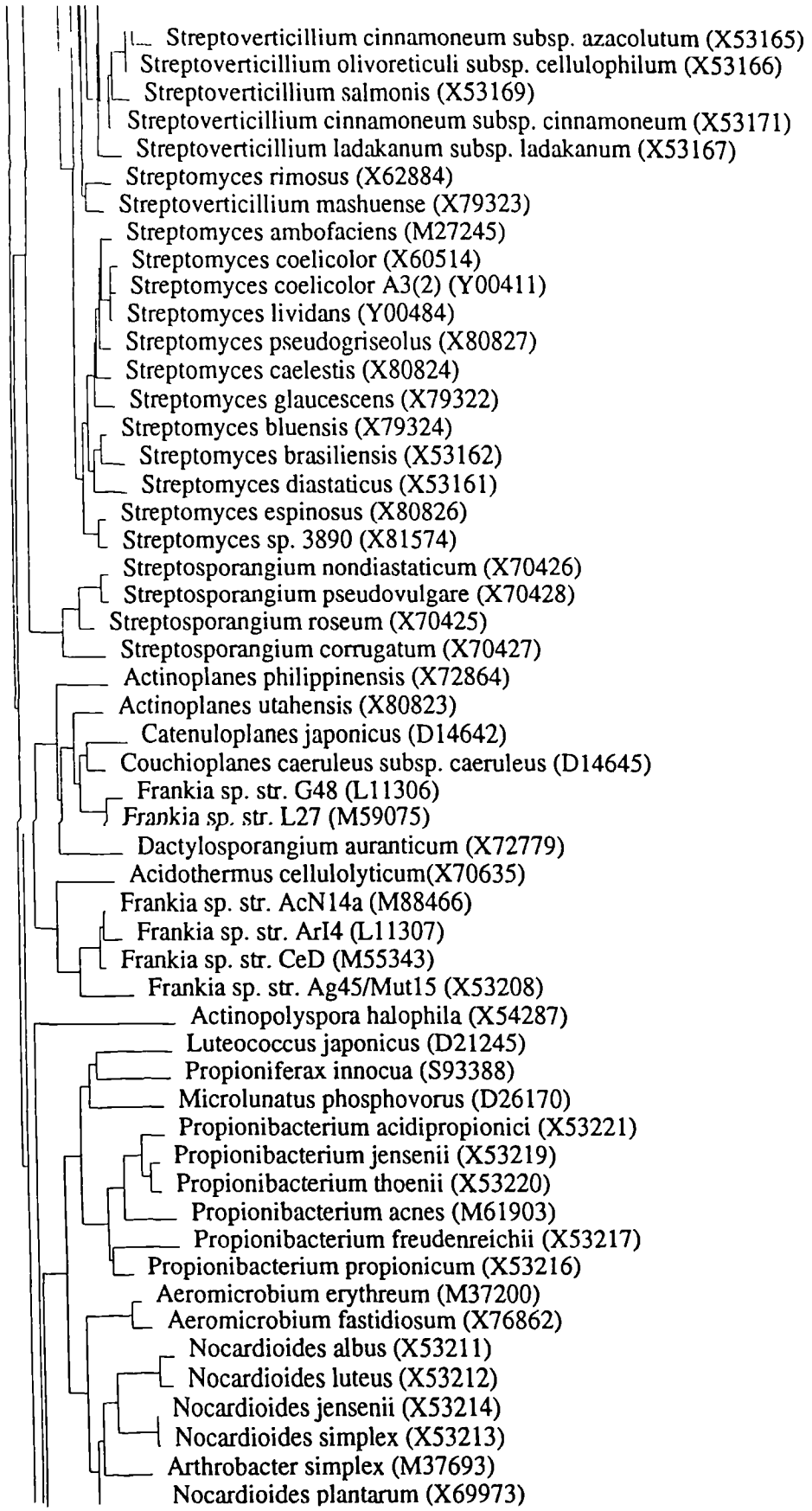
APPENDIX F. A PHYLOGENETIC TREE DEPICTING RELATIONSHIPS BETWEEN REPRESENTATIVES OF ACTINOMYCETES AND RELATED TAXA. TWO HUNDRED NINETY-FIVE SEQUENCES WERE ALIGNED USING THE AL16S PROGRAM, THE EVOLUTIONARY DISTANCES WERE CALCULATED USING THE JUKES AND CANTOR' CORRECTION (JUKES & CANTOR, 1969) AND TREE GENERATED USING THE NEIGHBOR-JOINING METHOD (SAITOU & NEI, 1987). THE TREE WAS BASED ON 1371 NUCLEOTIDE SITES AND *BACILLUS SUBTILIS* (ACCESSION NUMBER K00637) USED AS OUTGROUP. CORRESPONDING ACCESSION NUMBERS ARE GIVEN IN PARENTHESES.



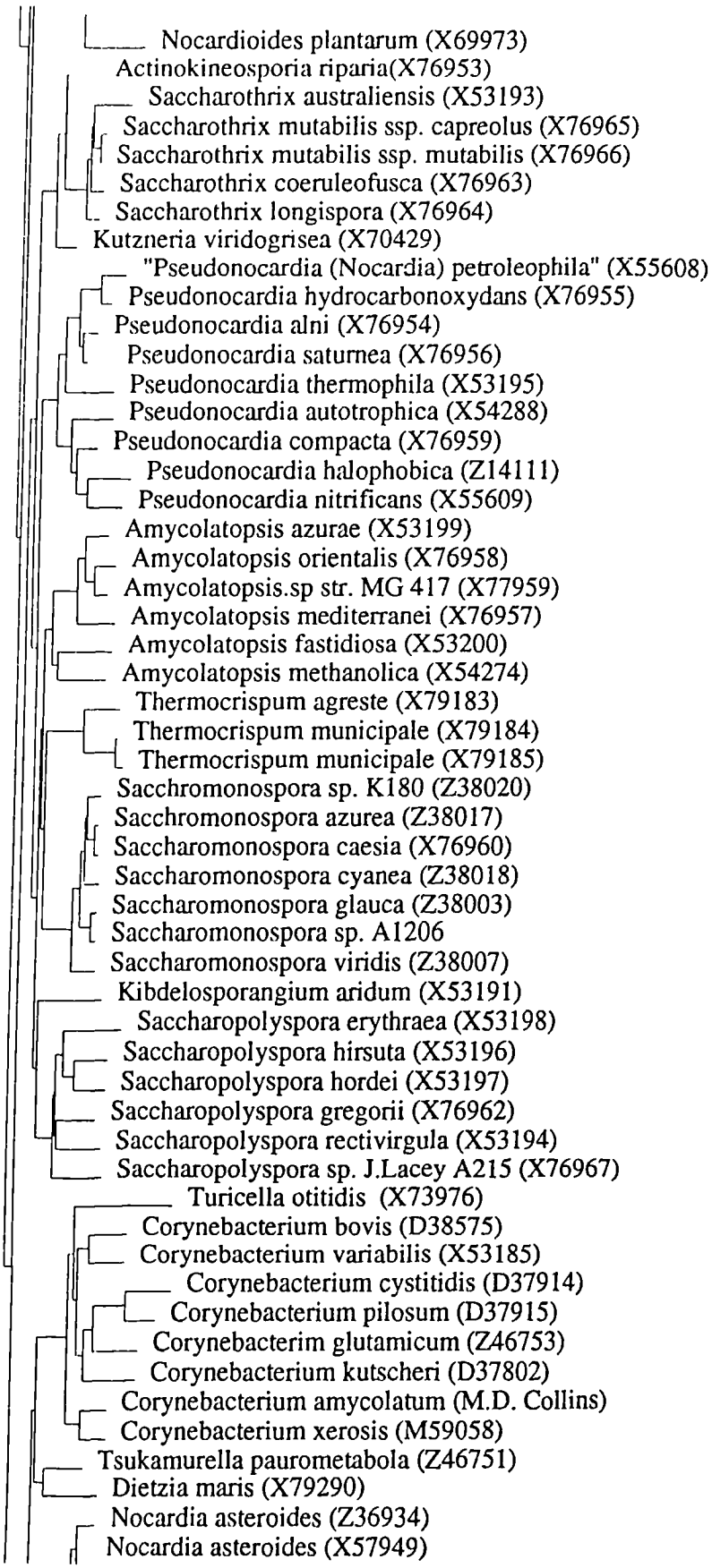
APPENDIX F. CONTINUED.



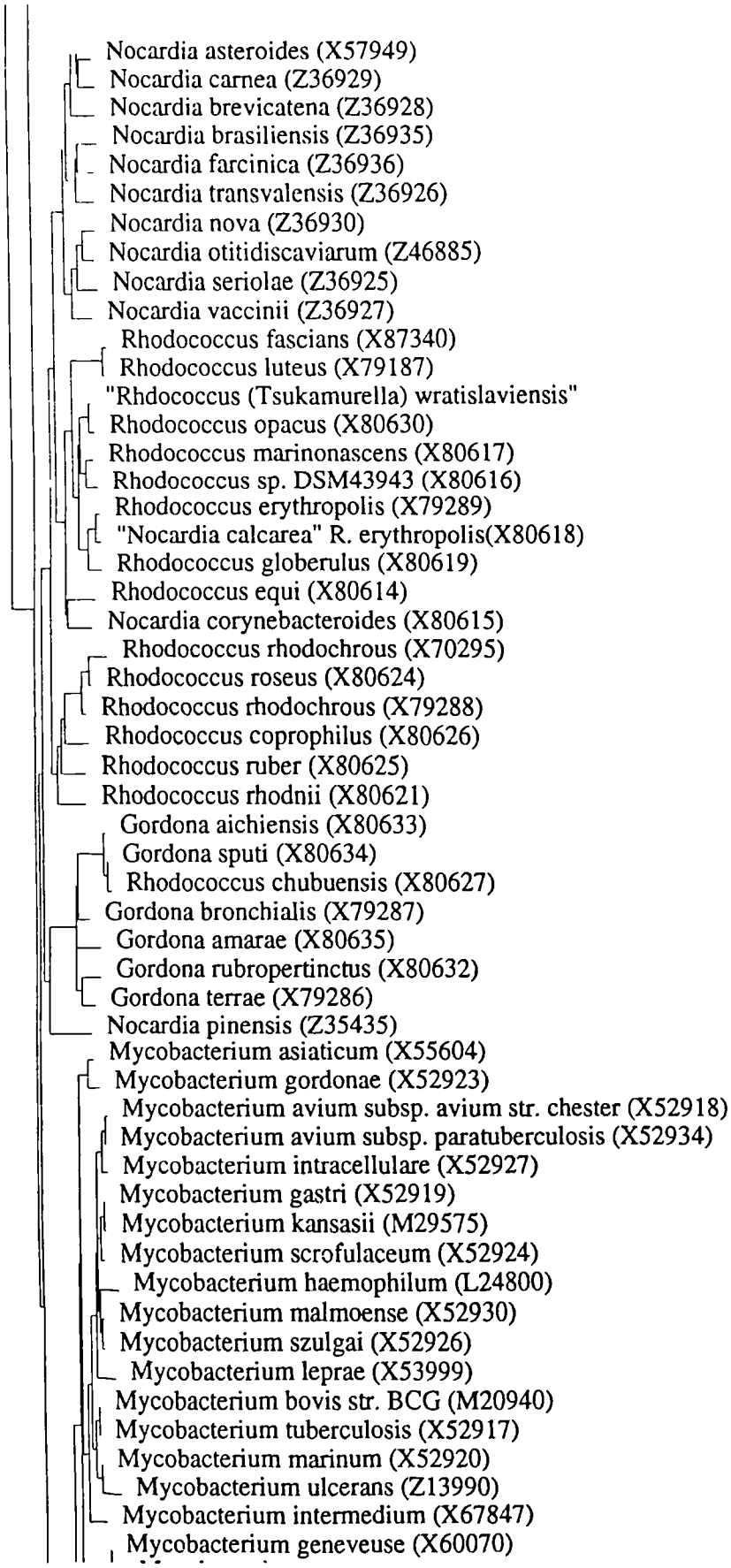
APPENDIX F. CONTINUED.



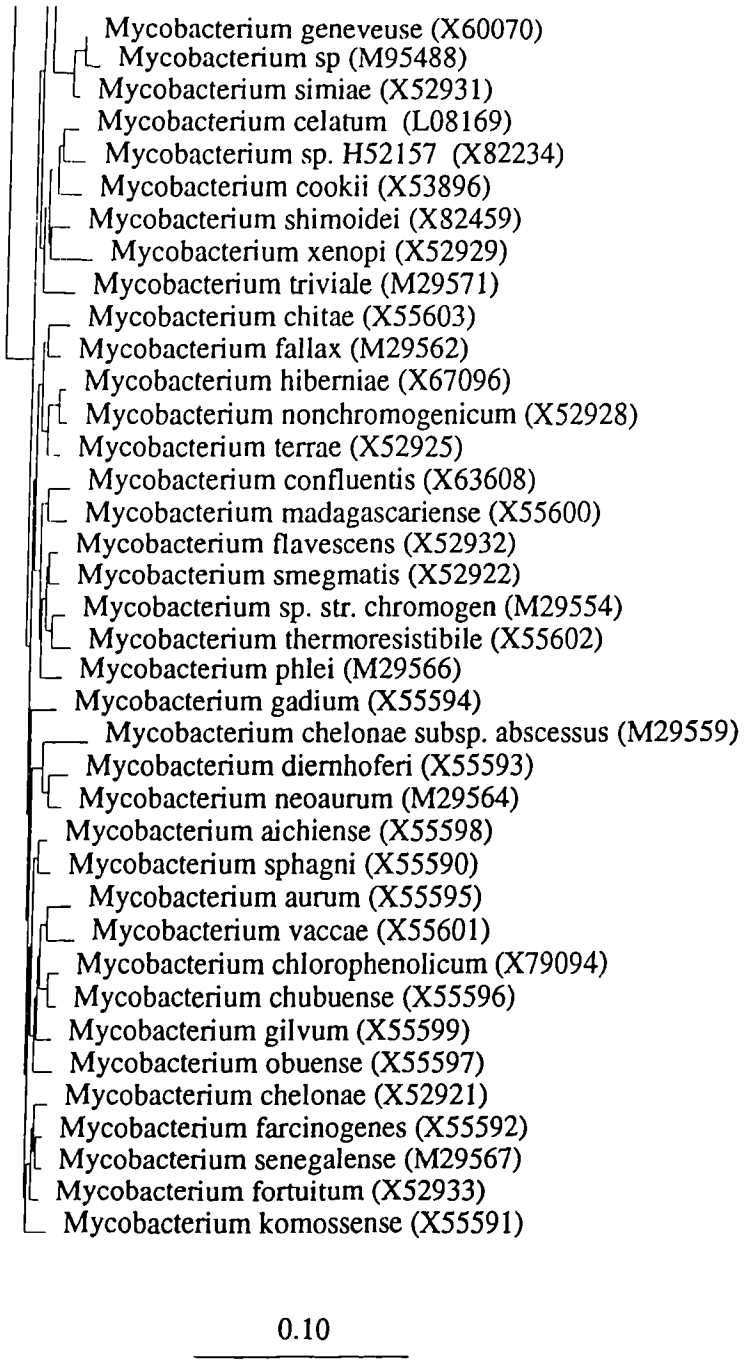
APPENDIX F. CONTINUED.



APPENDIX F. CONTINUED.



APPENDIX F. CONTINUED.



APPENDIX G. ACCESSION NUMBERS OF 16S rRNA SEQUENCES FOR MYCOLIC ACID-CONTAINING ACTINOMYCETES AND RELATED TAXA.

Species[1]	Accession numbers	Reference
Corynebacterium amycolatum	ND	M. D. Collins unpublished
Corynebacterium bovis	D38575	Takahashi *et al.* unpublished
Corynebacterium cystitidis	D37914	Takahashi *et al.* unpublished
Corynebacterium glutamicum	Z46753	This study
Corynebacterium variabilis	X53185	Collins *et al.* 1989b
Corynebacterium xerosis	M59058	Yang & Woese unpublished
Dietzia maris	X79290	Rainey *et al.* 1995c
Gordona aichiensis	X80633	Klatte *et al.* 1994c
Gordona amarae	X80635	Klatte *et al.* 1994c
Gordona bronchialis	X79287	Klatte *et al.* 1994c
Gordona rubropertincta	X80632	Klatte *et al.* 1994c
Gordona sputi	X80634	Klatte *et al.* 1994c
Gordona terrae	X79286;	Klatte *et al.* 1994c
Mycobacterium chlorophenolicum	X79094	Briglia *et al.* 1994
Mycobacterium fortuitum	X52933	Böttger unpublished
Mycobacterium intermedium	X67847	Meier *et al.* 1993
Mycobacterium leprae	X53999	Liesack *et al.* 1990
Mycobacterium simiae	X52931	Rogall *et al.* 1990
Mycobacterium smegmatis	X52922	Rogall *et al.* 1990
Mycobacterium tuberculosis	X52917	Rogall *et al.* 1990
Nocardia asteroides	Z36934	This study
Nocardia asteroides DSM43005	X57949	Rogall *et al.* 1990
Nocardia brasiliensis	Z36935	This study
Nocardia brevicatena	Z36928	This study
Nocardia carnea	Z36929	This study
"Nocardia crassostrae"	Z37989	This study
Nocardia farcinica	Z36936	This study
Nocardia nova	Z36930	This study
Nocardia otitidiscaviarum	Z46885	This study
Nocardia pinensis	Z35435	This study
Nocardia seriolae	Z36925	This study
Nocardia transvalensis	Z36926	This study

APPENDIX G. CONTINUED.

Species	Accession numbers	Reference
Nocardia vaccinii	Z36927	This study
Rhodococcus coprophilus	X80626	Rainey *et al.* 1995a
"Rhodococcus (Nocardia) corynebacteroides"	X80615	Rainey *et al.* 1995a
Rhodococcus equi	X80614	Rainey *et al.* 1995a
Rhodococcus erythropolis	X79289	Rainey *et al.* 1995a
Rhodococcus fascians	X87340	Klatte *et al.* 1994a
Rhodococcus globerulus	X80619	Rainey *et al.* 1995a
Rhodococcus marinonascens	X80617	Rainey *et al.* 1995a
Rhodococcus opacus	X80630	Klatte *et al.* 1994b
Rhodococcus rhodnii	X80621	Rainey *et al.* 1995a
Rhodococcus rhodochrous	X79288	Rainey *et al.* 1995b
Rhodococcus ruber	X80625	Rainey *et al.* 1995a
*Rhodococcus sp.*DSM 43943	X80616	Rainey *et al.* 1995a
Tsukamurella paurometabola	Z46751	This study
Tsukamurella wratislaviensis	Z37138	This study
Turicella otitidis	X73976	Funke *et al.* 1994

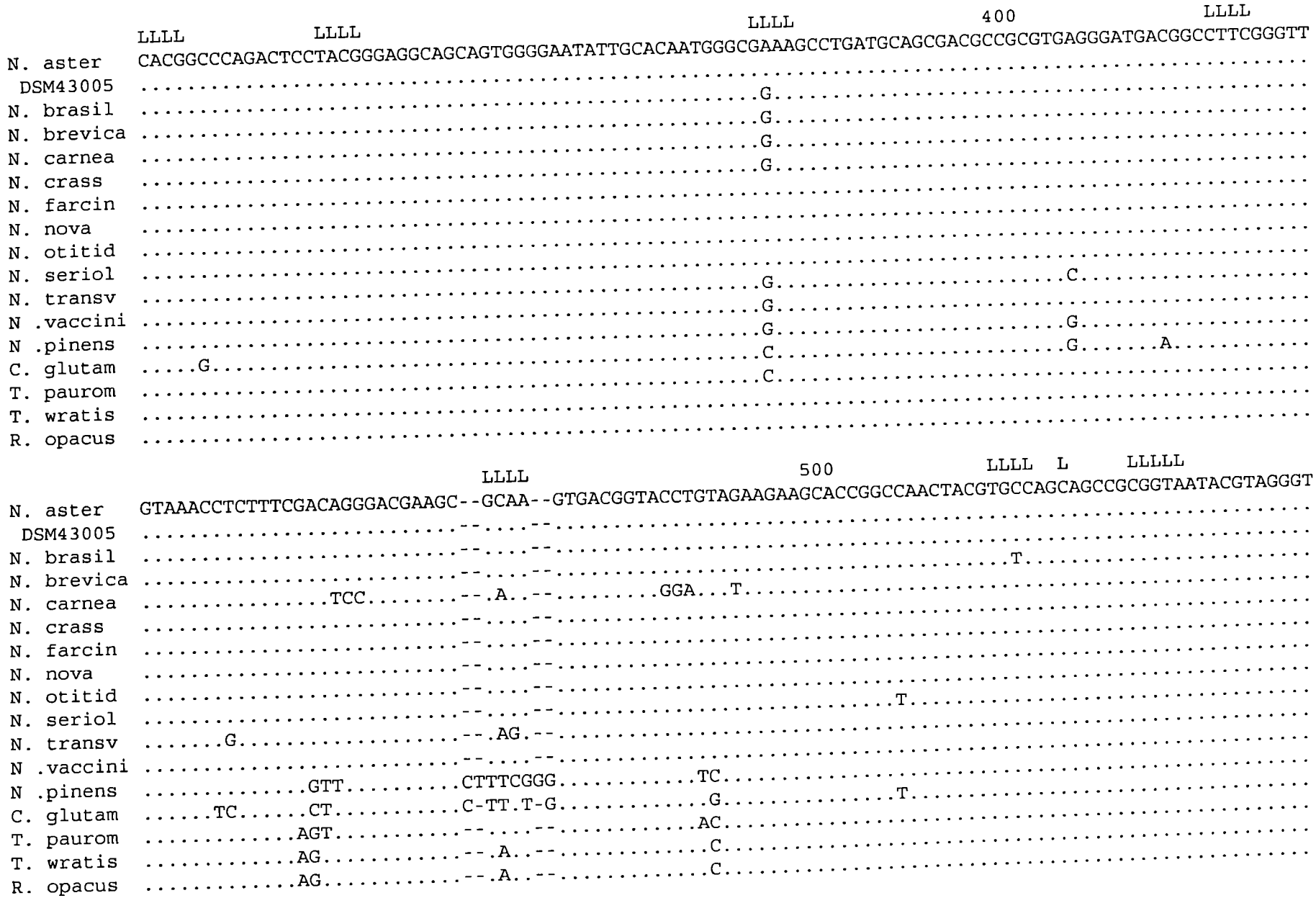
[1], Type strains unless indicated.

<BLANK>

APPENDIX H. ALIGNED 16S rDNA SEQUENCES OF NOCARDIAE AND THE RELATED ACTINOMYCETES DETERMINED IN THE PRESENT STUDY. THE SEQUENCES OF *NOCARDIA ASTEROIDES* DSM43005 AND *RHODOCOCCUS OPACUS* DSM 43205^T WERE OBTAINED FROM ROGALL *et al.* (1990; ACCESSION NUMBER X57949) AND KLATTE *et al.* (1994b; ACCESSION NUMBER X80630), RESPECTIVELY. ABBREVIATIONS: N. ASTER, *NOCARDIA ASTEROIDES*; DSM43005, *NOCARDIA ASTEROIDES* DSM43005; N. BRASIL, *NOCARDIA BRASILIENSIS*; N. BREVICA, *NOCARDIA BREVICATENA*; N. CARNEA, *NOCARDIA CARNEA*; N. CRASS, "*NOCARDIA CRASSOSTRAE*", N. FARCIN, *NOCARDIA FARCINICA*; N. NOVA, *NOCARDIA NOVA*; N. OTITID, *NOCARDIA OTITIDISCAVIARUM*; N. SERIOL, *NOCARDIA SERIOLAE*; N. TRANSV, *NOCARDIA TRANSVALENSIS*; N. VACCINI, *NOCARDIA VACCINII*; N. PINENS, *NOCARDIA PINENSIS*; C. GLUTAM, *CORYNEBACTERIUM GLUTAMICUM*; T. PAUROM, *TSUKAMURELLA PAUROMETABOLA*; T. WRATIS, *TSUKAMURELLA WRATISLAVIENSIS*; AND R. OPACUS, *RHODOCOCCUS OPACUS*. SYMBOLS: L, NUCLEOTIDE POSITION FOUND IN THE LOOP REGION; ?, NUCLEOTIDE THAT WAS NOT DETERMINED; AND -, ALIGNMENT GAP. THE NUMBERS INDICATE THE NUMBERING SYSTEM OF *ESCHERICHIA COLI* (BROSIUS *et al.*, 1978). OUTPUT WAS GENERATED USING THE MODULE '***PROFILE***' OF THE ***AL16S*** PROGRAM.

```
                                                                  LLLL            100
N. aster    GACGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAGCGGTAAGGCCC----TTCG----GGGTACACGAGCGGCGAACGGGTGAGTAACACGTGGGT
 DSM43005   ??????????..........................................----....----..-.....................................
N. brasil   .....................................................----....----.....................................
N. brevica  .....................................................----....----.....................................
N. carnea   .....................................................----....----.....................................
N. crass    .....................................................----....----.....................................
N. farcin   .....................................................----....----.....................................
N. nova     ....................G................................----....----.....................................
N. otitid   .....................................................----....----.....................................
N. seriol   .....................................................----....----.....................................
N. transv   .....................................................----....----.....................................
N .vaccini  .....................................................----....----.....................................
N .pinens   ..................................A.....-..----....----..-.....T.....................................
C. glutam   ..................................A..C.G.AA-..GGAGC.T.CTTT..-.GG.T...T.....................................
T. paurom   ..................................A..........T---.....---G..........T.....................................
T. wratis   .....................................................----....----.....................................
R. opacus   .....................................................----....----.....................................
```

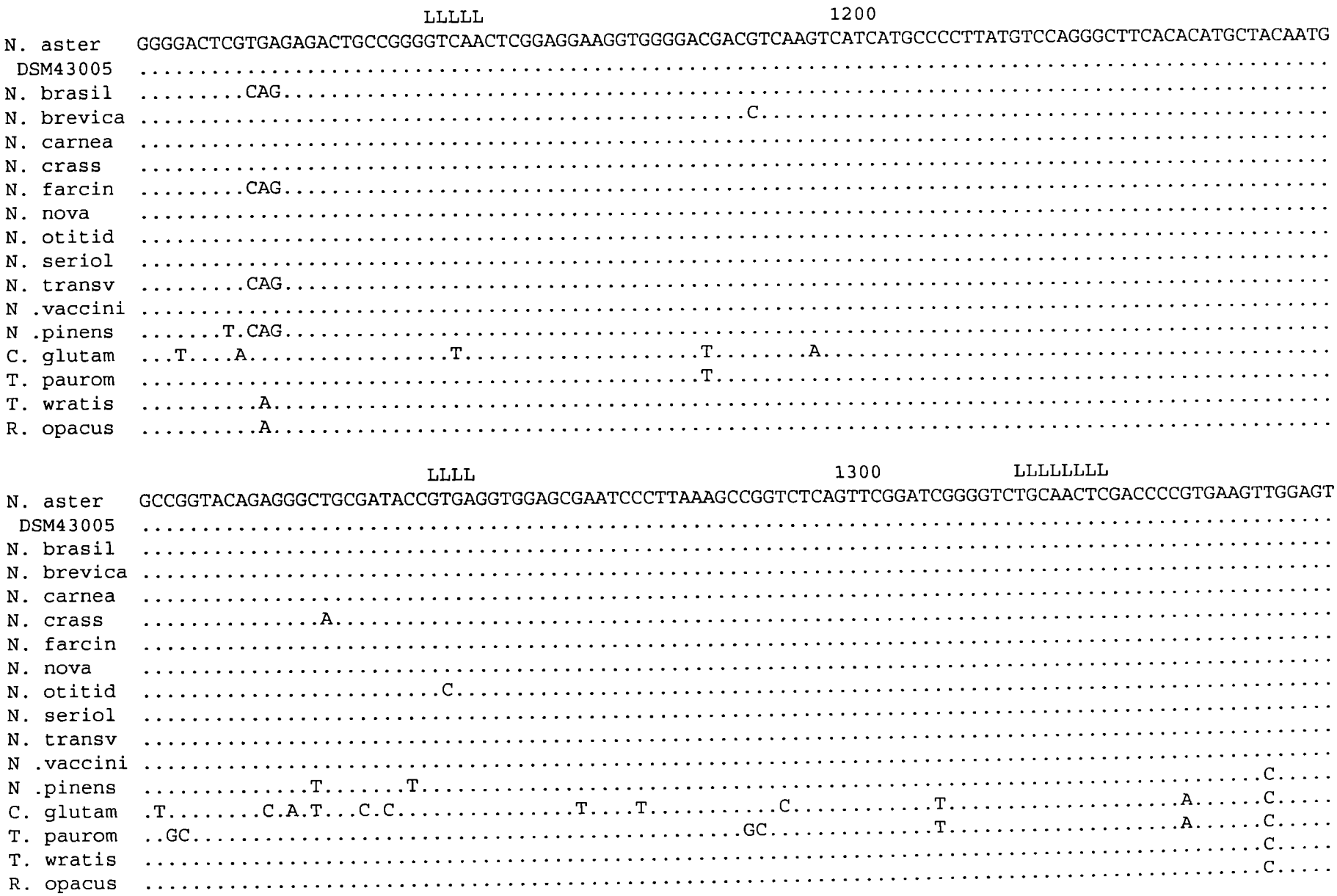
```
                                                                                LLLL                  LLLL
N. aster    GATCTGCCTCGTACTTCGGGATAAGCCTGGGAAACTGGGTCTAATACCGGATATGACCTTCGGATGCATGTCTGAGGGTGGAAAGATTTATCGGTACGAG
 DSM43005   ...........C..CT............................................ACATC.....GTGTTT...................G....
N. brasil   ...........C................................................TCAG......CTGTT....................G....
N. brevica  ...........C..CT.........................................T.CT..TC.....G.C.G....................G....
N. carnea   .........T.C..CT...........................................G.A.T......A..T.....................G....
N. crass    ..............CT..........................................ACGCAC......GTGTGT...................G....
N. farcin   ........CT..................................................ACATC.....GTGTTT.....................AG.
N. nova     ...........C..CT..........................................ACG.ATC.....GT.TGT...................G....
N. otitid   ...........C..CT..........................................ACGAATC.....GT.TGT.........G....CT...G....
N. seriol   .........T.C..CT........................................G.C.GAAC......GT.CG..T.................G....
N. transv   ..............CT..........................................ACAT.TC.....GTGTGT...................G....
N .vaccini  .........T.C..CT..........................................A.T..T......ATC.GT...................G....
N .pinens   ........CT.G..CT..........T...............................A.GCTGC......GGTGT.........C...TG....CT.G.
C. glutam   ........CTAC...T.................................A....TCA.AC.ACCGTAGG-GG..GT.T-......C....G....GT.G.
T. paurom   ..C.....CT.....T.............................................TCC......GGG.TT.........C...TG......AG.
T. wratis   ........CT.C...................................................C......G..............G....CT...G.AG.
R. opacus   ........CT.C...................................................C......G..............G....CT...G.AG.

                                            LLLLLLL                               LLL300                 LLLLLL
N. aster    ATGGGCCCGCGGCCTATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGA
 DSM43005   ....................................................................................................
N. brasil   ............................C.......C....C.......................A..........T.......................
N. brevica  ..A..............................................................................C..................
N. carnea   ..A.................................................................................................
N. crass    ....................................C............................A..........T..T....................
N. farcin   ....................................................................................................
N. nova     .................................................................A..........T.......................
N. otitid   ....................................C...............................................................
N. seriol   ..............................A.....C...TC..........................................................
N. transv   ....................................................................................................
N .vaccini  ....................................................................................................
N .pinens   ............................................................................T.T.....................
C. glutam   ..A...T............................................T........................T.TA.......T............
T. paurom   ....................................................................................................
T. wratis   .................................................................A..........T.......................
R. opacus   .................................................................A..........T.......................
```


LLLL
LLLL
LLLL
400
LLLL
N. aster CACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTT
DSM43005
N. brasil
N. brevica
N. carnea
N. crass
N. farcin
N. nova
N. otitid
N. seriol
N. transv
N .vaccini
N .pinens
C. glutam
T. paurom
T. wratis
R. opacus
LLLL
500
LLLL L LLLLL
N. aster GTAAACCTCTTTCGACAGGGACGAAGC--GCAA--GTGACGGTACCTGTAGAAGAAGCACCGGCCAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGT
DSM43005
N. brasil
N. brevica
N. carnea
N. crass
N. farcin
N. nova
N. otitid
N. seriol
N. transv
N .vaccini
N .pinens
C. glutam
T. paurom
T. wratis
R. opacus

```
                                                                               600                 LLLLL
N. aster    GCGAGCGTTGTCCGGAATTACTGGGCGTAAAGAGCTTGTAGGCGGTTCGTCGCGTCGCTCGTGAAAACTTGGGGCTCAACCCCAAGCTTGCGGGCGATAC
 DSM43005   .............................................C.T.........A..................................NT......
N. brasil   ...............................................T.........TC.........................................
N. brevica  .............................................C.T.........A...........CACA.......TGTG.T.......T......
N. carnea   .............................................C.T.........A...........C.............G.........T......
N. crass    ..............................................CT...A.....A...........G.............C...C.....T......
N. farcin   ...............................................T.........TC................C........................
N. nova     .........................................................A.T...........CA.......TG.........A.T......
N. otitid   ..............................................CT.........A..........G..CA.......TG.GC..C.....T......
N. seriol   ..................G...........................CT...A.....G...........CACA.......TGTGG..C.....T......
N. transv   ...............................................T.........T..........................................
N .vaccini  .........................................................A..........A.ACA.......TGTGT........T......
N .pinens   ....................................C..........T.........A...........C.CA.......TGTGG..G.....T......
C. glutam   ....................................C.....T....T.........TCT......T.CC......T...TT.GG..G...A........
T. paurom   ..A.................................C..........T...A.....TCT........CC.A....T....T.GG..C...A........
T. wratis   ..A.................................C..........T.........TCT.........C.A.........T.G.......A........
R. opacus   ..A.................................C..........T.........TCT.........CANA........T.G.......A........

                                                                   LLLLLL  700                   LLLL
N. aster    GGGCG-GTCTAGAGTACTTCAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGGTCTCTGG
 DSM43005   ....A-.G..T.........................................................................................
N. brasil   ....A-.A..T.........................................................................................
N. brevica  ....A-.G..T.......C..................T..............................................................
N. carnea   ....A-.G..T................................................................................T........
N. crass    ....A-.A............................................................................................
N. farcin   ....A-.A..T.......G.................................................................................
N. nova     .....-.A............................................................................................
N. otitid   ....A-.A............................................................................................
N. seriol   ....A-.A............................................................................................
N. transv   ....A-.A..........G.................................................................................
N .vaccini  .....-.A..........G.................................................................................
N .pinens   ....A-.A..T.......C.................................................................................
C. glutam   ....ATAA..T....G..GT..........................................................AA..........A.........
T. paurom   ....A-.A..T.......GT...............................G.........................T......................
T. wratis   ....A-.A..T.......G.................................................................................
R. opacus   ....A-.A..T.......G.................................................................................
```

```
                                                         LLLLLLLLL    800                                     LL
N. aster     GAAGTAACTGACGCTGAGAAGCGAAAGCGTGGGTAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGGTGGGTACTAGGTGTGGGTTTCCT
 DSM43005    ...................................................................N................................
N. brasil    ....................................................................................................
N. brevica   .C..................................................................................................
N. carnea    ....................................................................................................
N. crass     ....................................................................................................
N. farcin    .C...............................................................................CG.................
N. nova      ....................................................................................................
N. otitid    ....................................................................................................
N. seriol    ....................................................................................................
N. transv    .C..................................................................................................
N .vaccini   .C...............................................................................CG.................
N .pinens    .C................G..............................................................CG.................
C. glutam    .C................G.........A....................................T...............CG........A..GG..-.
T. paurom    .C................G.................................................................................
T. wratis    .C................G..............................................................CG.................
R. opacus    .C................G..................A...........................................CG.................

             LLLLL                 LLLL                               LL900
N. aster     TCCACGGGATCCGTGCCGTAGCTAACGCATTAAGTACCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGC
 DSM43005    ........................................................N................................N..........
N. brasil    ....................................................................................................
N. brevica   .....................T.................................T............................................
N. carnea    ....................................................................................................
N. crass     ....................................................................................................
N. farcin    ..................................CG................................................................
N. nova      ....................................................................................................
N. otitid    ....................................................................................................
N. seriol    ....................................................................................................
N. transv    ....................................................................................................
N .vaccini   ..................................CG................................................................
N .pinens    ..................................CG................................................................
C. glutam    ......ACT..T......C...............CG................................................................
T. paurom    ....................................................................................................
T. wratis    ..................................CG................................................................
R. opacus    .....................T............CG...........................T....................................
```

```
                                  LLLLLLLL                                                1000          LLLL          LLLLL
N. aster    GGCGGAGCATGTGGATTAATTCGATGCAACGCGAAGAACCTTACCTGGGTTTGACATACACCGGAAACCTGCAGAGATGTAGGCCCCCTTGTGGTCGGTG
 DSM43005   ................................................................................................................
N. brasil   ................................................................................................................
N. brevica  ................................................................................................................
N. carnea   ................................................................................................................
N. crass    ..............................................................G.......G.C.T.......CG.C.....................
N. farcin   ................................................................................................................
N. nova     ..............................................................A....G.C.T......ACG.C............T....
N. otitid   ..............................................................G.C.T......ACG.C.....................
N. seriol   ..............................................................G.C.T......ACG.C.....................
N. transv   ..............................................................A....G..............C............T....
N .vaccini  ..............................................................A....G.C.T......ACG.C............T....
N .pinens   ..............................................................A..CG.TG.T......A.CA.TT..........T....
C. glutam   ..........................................................C.......GG......TCGGC.T......ACGTTTT..............T
T. paurom   ..............................................................T.GA...TCG.C..........G.TTTG.......CCTTC.A
T. wratis   ..............................................................T........G.C.T......ACG.C..............A
R. opacus   ..............................................................T........G.C.T......ACC.C..............A

                                                  LLLL          LLLLLL     1100                                      LLLLLL
N. aster    TACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCTTATGTTGCCAGCGCGTAATGGC
 DSM43005   ................................................................................................................
N. brasil   ..................................................................................G..C.G..........GA.T...C.
N. brevica  ................................................................................................................
N. carnea   ................................................................................................................
N. crass    ..............................................................................................................C.....
N. farcin   ..................................................................................G..C.G............T.....
N. nova     ................................................................................................................
N. otitid   ................................................................................................................
N. seriol   ................................................................................................................
N. transv   ..................................................................................G..C.G............C.....
N .vaccini  ..............................................................................................................T.....
N .pinens   ..................................................................................G..C.G.A..........T.....
C. glutam   C.............T.....................................................................G.............A.A.T...T
T. paurom   ..................................................................................G...C..........A........T
T. wratis   ..................................................................................G.............A........T
R. opacus   ..................................................................................G.............A........T
```

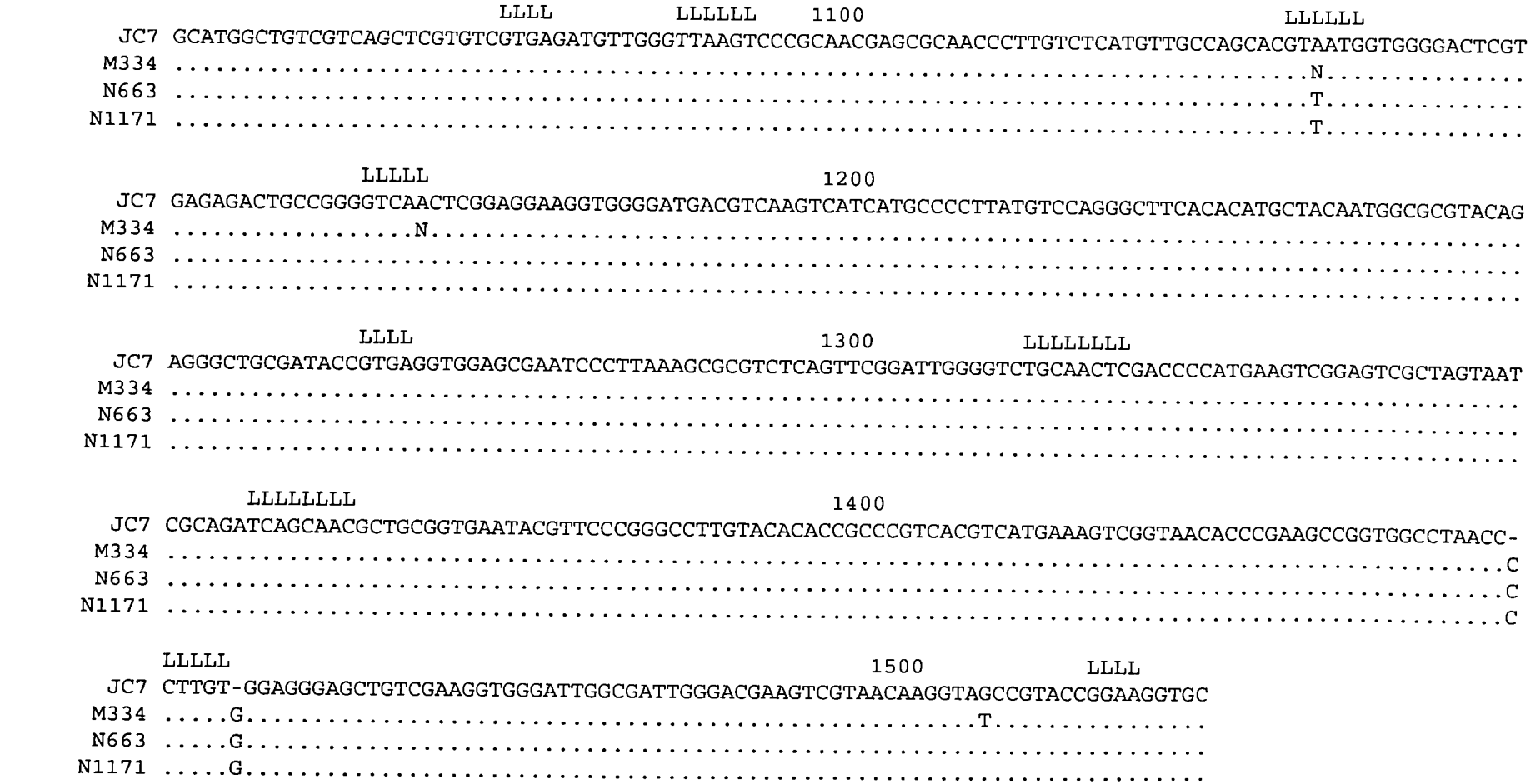

LLLLL
1200
N. aster GGGGACTCGTGAGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGACGACGTCAAGTCATCATGCCCCTTATGTCCAGGGCTTCACACATGCTACAATG
DSM43005
N. brasil .....CAG.....
N. brevica .....C.....
N. carnea
N. crass
N. farcin .....CAG.....
N. nova
N. otitid
N. seriol
N. transv .....CAG.....
N .vaccini
N .pinens .....T.CAG.....
C. glutam ...T....A.....T.....T.....A.....
T. paurom .....T.....
T. wratis .....A.....
R. opacus .....A.....
LLLL
1300
LLLLLLLL
N. aster GCCGGTACAGAGGGCTGCGATACCGTGAGGTGGAGCGAATCCCTTAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAACTCGACCCCGTGAAGTTGGAGT
DSM43005
N. brasil
N. brevica
N. carnea
N. crass .....A.....
N. farcin
N. nova
N. otitid .....C.....
N. seriol
N. transv
N .vaccini
N .pinens .....T.....T.....C.....
C. glutam .T.....C.A.T...C.C.....T....T.....C.....T.....A.....C.....
T. paurom ..GC.....GC.....T.....A.....C.....
T. wratis
R. opacus

```
                           LLLLLLLL                                     1400
N. aster    CGCTAGTAATCGCAGATCAGCAACGCTGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACGTCATGAAAGTCGGTAACACCCGAAGCCGGT
 DSM43005   ....................................................................................................
N. brasil   ....................................................................................................
N. brevica  ....................................................................................................
N. carnea   ....................................................................................................
N. crass    ...................................................................................................C.
N. farcin   .............................................A......................................................
N. nova     ....................................................................................................
N. otitid   ....................................................................................................
N. seriol   ....................................................................................................
N. transv   ....................................................................................................
N .vaccini  ....................................................................................................
N .pinens   .......................T.........................................................................C..
C. glutam   ................................................................................T................A..
T. paurom   ....................................................................................................
T. wratis   ....................................................................................................
R. opacus   ....................................................................................................

                       LLLLL                                                 1500           LLLL
N. aster    GGCCTAACC--CTTGT-GGAGGGAGCCGTCGAAGGTGGGATCGGCGATTGGGACGAAGTCGTAACAAGGTAGCCGTACCGGAAGGTGC
 DSM43005   .........C-..C..G.......................................................????????????????
N. brasil   .........C-.....G........................T..............................................
N. brevica  .........C-.....G.......................................................................
N. carnea   .........--.....-.......................................................................
N. crass    .........C-G.AA-G........G...............T..............................................
N. farcin   .........C-.....G.......................................................................
N. nova     .........--.....-........................T..............................................
N. otitid   .........--..CTG-........................T..............................................
N. seriol   .........--.....-........................T..............................................
N. transv   .........C-.....G........................T..............................................
N .vaccini  .........--.....-.......................................................................
N .pinens   .........CCT....G.........G.............................................................
C. glutam   ....C....--T..TA-..G......T.............................................................
T. paurom   .........--.....-.........T..............T..............................................
T. wratis   .........C-.....G.....................................................................???
R. opacus   .........C-.....G.......................................................................
```

APPENDIX I. ALIGNED 16S rDNA SEQUENCES OF SLUDGE ISOLATE N1171 AND *TSUKAMURELLA PARAUMETABOLA* STRAINS JC7^T, N663 AND M334. SYMBOLS: L, NUCLEOTIDE POSITION FOUND IN THE LOOP REGION; ?, NUCLEOTIDE THAT WAS NOT DETERMINED; AND -, ALIGNMENT GAP. THE NUMBERS INDICATE THE NUMBERING SYSTEM OF *ESCHERICHIA COLI* (BROSIUS *et al.*, 1978). OUTPUT WAS GENERATED USING THE MODULE ***PROFILE*** OF THE ***AL16S*** PROGRAM.

```
                                                                LLLL         100
  JC7 GACGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGGTAAGGCCCTTTCGGGGGTACACGAGTGGCGAACGGGTGAGTAACACGTGGGTGACCTG
 M334 ....................................................................................................
 N663 ....................................................................................................
N1171 ..................................................-....-.....................................................

                                                                          LLLL                 LLLL
  JC7 CCCTGTACTTTGGGATAAGCCTGGGAAACTGGGTCTAATACCGGATATGACCTTCTCCTGCATGGGGGTTGGTGGAAAGCTTTTGCGGTACAGGATGGGC
 M334 ..........C.....................................................................................
 N663 ..........C.....................................................................................
N1171 ..........C.....................................TC...........T.....................................

                                  LLLLLLL                                        LLL300                 LLLLLLLLLL
  JC7 CCGCGGCCTATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGC
 M334 ....................................................................................................
 N663 ....................................................................................................G
N1171 ....................................................................................................

                LLLL                                    LLLL                 400                 LLLL
  JC7 CCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAAC
 M334 ....................................................................................................
 N663 ....................................................................................................
N1171 ..............................................A.....................................................
```

```
                          LLLL                    500             LLLL  L     LLLLL
  JC7 CTCTTTCAGTAGGGACGAAGCGCAAGTGACGGTACCTACAGAAGAAGCACCGGCCAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCAAGCGTTG
 M334 ....................................................................................................
 N663 ..........................................................................C.............G.......
N1171 ..........................................................................C.............G.......

                                                  600             LLLLL
  JC7 TCCGGAATTACTGGGCGTAAAGAGCTCGTAGGCGGTTTGTCACGTCGTCTGTGAAAACCCGAGGCTTAACCTCGGGCCTGCAGGCGATACGGGCA-GACT
 M334 ....................................................................................................-....
 N663 ......T.....................................G.....................................................-....
N1171 ......T.....................................................................................-....

                                          LLLLLL  700                             LLLL
  JC7 TGAGTACTGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGGAATGCGCAGATATCAGGAGGAACACTGGTGGCGAAGGCGGGTCTCTGGGCAGTAACTG
 M334 ..........................................................C.........................................
 N663 ..........................................................C.........................................
N1171 ..........................................................C.........................................

                                      LLLLLLLLL   800                                                 LLLLLLL
  JC7 ACGCTGAGGAGCGAAAGCGTGGGTAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGGTGGGTACTAGGTGTGGGTTTCCTTCCACGGGAT
 M334 ....................................................................................................
 N663 ....................................................................................................
N1171 ....................................................................................................

                  LLLL                          LL900
  JC7 CCGTGCCGTAGCTAACGCATTAAGTACCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCAT
 M334 ....................................................................................................
 N663 ....................................................................................................
N1171 ....................................................................................................

                     LLLLLLLL                          1000          LLLL          LLLLL
  JC7 GTGGATTAATTCGATGCAACGCGAAGAACCTTACCTGGGTTTGACATATAGAGGATCGCCGCAGAGATGTGGTTTGCCTTGTGCCTTCTATACAGGTGGT
 M334 ..........................................................G......TC..........................
 N663 ....................................................................................................
N1171 ....................................................................................................
```


LLLL
LLLLLL
1100
LLLLLL
JC7 GCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCTCATGTTGCCAGCACGTAATGGTGGGGACTCGT
M334 .....N.....
N663 .....T.....
N1171 .....T.....
LLLLL
1200
JC7 GAGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAGTCATCATGCCCCTTATGTCCAGGGCTTCACACATGCTACAATGGCGCGTACAG
M334 .....N.....
N663 .....
N1171 .....
LLLL
1300
LLLLLLLL
JC7 AGGGCTGCGATACCGTGAGGTGGAGCGAATCCCTTAAAGCGCGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTCGCTAGTAAT
M334 .....
N663 .....
N1171 .....
LLLLLLLL
1400
JC7 CGCAGATCAGCAACGCTGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACGTCATGAAAGTCGGTAACACCCGAAGCCGGTGGCCTAACC-
M334 C
N663 C
N1171 C
LLLLL
1500
LLLL
JC7 CTTGT-GGAGGGAGCTGTCGAAGGTGGGATTGGCGATTGGGACGAAGTCGTAACAAGGTAGCCGTACCGGAAGGTGC
M334 .....G.....T.....
N663 .....G.....
N1171 .....G.....

APPENDIX J. TEST STRAINS HELD IN THE NON-TARGET DATABASE.

Laboratory numbers	Taxa	Strain histories
• **Genus *Actinomadura***		
A9	*A. pelletieri*	C. Philpot (London School of Hygiene and Tropical Medicine, Keppel St., London, UK; LSHTM), 368
A11	*A. madurae*	C. Philpot, 393
A10	*A. pelletieri*	C. Philpot, 1065; Nigeria; mycetoma of the arm (=DSM 43382)
A13	*A. pelletieri*	C. Philpot, 388H (=DSM 44039)
A16^T	*A. madurae*	NCTC 5654; mycetoma pedis (=ATCC 19425; CCM 136; DSM 43067; IMRU 1190)
A22	*A. madurae*	F. Mariat (Institut Pasteur, 24 Rue du Dr. Roux, F75015, Paris, France), IP 725 (=DSM 44021)
A31	*A. madurae*	F. Mariat, IP 363; madura foot, Tunis (=DSM 44024)
A32	*A. madurae*	F. Mariat, IP 364; madura foot (=DSM 43380)
• **Genus *Mycobacterium***		
M263^T	*M. senegalense*	NCTC 10956
M259	*M. senegalense*	NCTC 4524
M264	*M. senegalense*	E.H. Runyon, R397; bovine farcy, Dakar, Senegal
M265	*M. senegalense*	E.H. Runyon, R408; bovine farcy, Dakar, Senegal
M266	*M. senegalense*	E.H. Runyon, R409; bovine farcy, Dakar, Senegal
M268	*M. senegalense*	E.H. Runyon, R410; bovine farcy, Dakar, Senegal
M279	*M. senegalense*	E.H. Runyon, R148
M282	*M. senegalense*	E.H. Runyon, R455
M283	*M. senegalense*	E.H. Runyon, R456
N714	*M. senegalense*	M. Ridell, GB045; R.E. Gordon, R1363, bovine farcy, Dakar, Senegal
N717	*M. senegalense*	M. Ridell, GB048; bovine farcy, Dakar, Senegal
N718	*M. senegalense*	M. Ridell, GB049; bovine farcy, Dakar, Senegal
N721	*M. senegalense*	M. Ridell, GB053; bovine farcy, Dakar, Senegal
N723	*M. senegalense*	M. Ridell, GA816
N726	*M. senegalense*	M. Ridell, GA881
N728	*M. senegalense*	M. Ridell, GA883
M262^T	*M. farcinogenes*	NCTC 10955
M5	*M. farcinogenes*	M. Abdulle (Serum and Vaccine Institute, P.O. Box 919, Mogadishu, Somalia), MA5; bovine farcy, skin lesion, Somalia
M9	*M. farcinogenes*	M. Abdulle, MA9,
M30	*M. farcinogenes*	M. Abdulle, MA30
M39	*M. farcinogenes*	M. Abdulle, MA39

APPENDIX J. CONTINUED.

Laboratory numbers	Taxa	Strain histories
M191	*M. farcinogenes*	M. Abdulle, MA191
M192	*M. farcinogenes*	M. Abdulle, MA192
M217	*M. farcinogenes*	M. Abdulle, MA217
M612	*M. farcinogenes*	M. Abdulle, MA612
M768	*M. farcinogenes*	M. Abdulle, MA768
M555	*M. farcinogenes*	M. Abdulle, MA555, bovine farcy, mammary gland, Somalia
M687	*M. farcinogenes*	M. Abdulle, MA687, bovine farcy, liver, Somalia
M785	*M. farcinogenes*	M. Abdulle, MA785, bovine farcy, lung, Somalia
M52	*M. farcinogenes*	S.M. El-Sanousi, (Department of Microbiology, Faculty of Veterinary Science, P.O.Box 32, Khartoum North, Sudan), KHF9; bovine farcy, lymph nodes, Sudan
M54	*M. farcinogenes*	S.M. El-Sanousi, KHF1; bovine farcy, lymph nodes, Sudan
M55	*M. farcinogenes*	S.M. El-Sanousi, KF8; bovine farcy, lymph nodes, Sudan
M57	*M. farcinogenes*	S.M. El-Sanousi, KF6; bovine farcy, lymph nodes, Sudan
• **Genus *Nocardia***		
N97	*N. asteroides*	R.E. Gordon (Rutgers Univeristy, New Brunswick, USA), N659 (*N. caprae*); =NCTC 659
N105	*N. asteroides*	R.E. Gordon, 9969; =ATCC 9969
N317^T	*N. asteroides*	R.E. Gordon, IMRU 727; =ATCC 19247; Garden soil, Thailand
N668	*N. asteroides*	S.G. Bradley (Virginia Common-Wealth University, Georgia, USA), VAC 462
N688	*N. asteroides*	K.P. Schaal (Hygiene Institute, Cologne, Germany), N2
N692	*N. asteroides*	K.P. Schaal, N23; H. Mordarski, strain Copenhagen
N695	*N. asteroides*	K.P. Schaal, N67
N900	*N. asteroides*	J.L. Stanford (Middlesex Hospital, London, UK), N36, general nocardiosis in an immunosuppressed female
N901	*N. asteroides*	J.L. Stanford, N37; pulmonary nocardiosis
N902	*N. asteroides*	J.L. Stanford, N38; pulmonary nocardiosis
N907	*N. asteroides*	K.P. Schaal, N60; M. Tsukamura, M-129, 23046
N911	*N. asteroides*	K.P. Schaal, N199; R.E. Gordon, IMRU 3419 (*N. carnea*); replicate of ATCC 3847
N912	*N. asteroides*	K.P. Schaal, N228; V.A. Orchard, 255, soil isolate
N1132	*N. asteroides*	K.P. Schaal, N3
N1134	*N. asteroides*	K.P. Schaal, N11
N1135	*N. asteroides*	K.P. Schaal, N19
N1139	*N. asteroides*	K.P. Schaal, N89
N1140	*N. asteroides*	K.P. Schaal, N129; soil isolate

APPENDIX J. CONTINUED.

Laboratory numbers	Taxa	Strain histories
N14	*N. brasiliensis*	R.E. Gordon, IMRU 744; A. Gonzáles-Ochoa, 409
N118	*N. brasiliensis*	CBS 438.64; A.M. Klokke, 300
N318^T	*N. brasiliensis*	R.E. Gordon, IMRU 854; J.D. Schneidau Jr., 381; A. Batista, 631; =ATCC 19296
N428	*N. brasiliensis*	R.E. Gordon, IMRU 1336; *Nocardia spp.;* M.P. Lechevalier, L-36; soil isolate
N467	*N. brasiliensis*	IP 708
N471	*N. brasiliensis*	A. Gonzáles-Ochoa (Instituto de Salubridad y Enfermedades Tropicales, Mexico City, Mexico), 4115; mycetoma lower leg
N474	*N. brasiliensis*	A. Gonzáles-Ochoa, 4212; mycetoma ankle
N475	*N. brasiliensis*	A. Gonzáles-Ochoa, 4023; mycetoma forearm
N477	*N. brasiliensis*	A. Gonzáles-Ochoa, 4204; mycetoma heel; K.P. Schaal, N219
N1146	*N. brasiliensis*	K.P. Schaal, N214
N1148	*N. brasiliensis*	K.P. Schaal, N224
N13	*N. farcinica*	NCTC 8595
N233	*N. farcinica*	R.J. Olds (University of Cambridge, UK), CN 470
N669	*N. farcinica*	K.P. Schaal, N200
N670	*N. farcinica*	K.P. Schaal, N201
N671	*N. farcinica*	S.G. Bradley, VAC 330; M. Tsukamura, E-7549
N687	*N. farcinica*	K.P. Schaal, N1
N690	*N. farcinica*	K.P. Schaal, N5
N699	*N. farcinica*	M. Ridell (University of Göteborg, Sweden), N58; =ATCC 3318
N701	*N. farcinica*	M. Ridell, N67
N898^T	*N. farcinica*	M. Tsukamura, R-3318, 23102; =ATCC 3318
N1116^T	*N. seriolae*	JCM 4826
• **Genus *Nocardiopsis***		
A14^T	*N. dassonvillei*	NCTC 10488; R.E. Gordon, IMRU 509 (*Nocardia dassonvillei*); (=ATCC 23218; DSM 43111)
A27	*N. dassonvillei*	F. Mariat, IP 395; M. André, Fort Lamy, Chad; mycetoma of the arm (=DSM 40928)
A120	*N. dassonvillei*	H. Prauser (Zentralinstitut für Mikrobiologie und Experimentelle Therapie, Jena, Germany); R.E. Gordon, IMRU 1250 (=ATCC 23219; DSM 43235)
A147	*N. dassonvillei*	J. Lacey (Department of Plant Pathology, Rothamsted Experimental Station, Harpenden, Hertforshire, UK), A1441 (=DSM 44045)
A150	*N. dassonvillei*	J. Lacey, A1460, cotton dust, Cramlington, UK (=DSM 44046)

APPENDIX J. CONTINUED.

Laboratory numbers	Taxa	Strain histories
A202	*N. dassonvillei*	J. Lacey, A1505, cotton dust, Cramlington, UK (=DSM 44047)
A204	*N. dassonvillei*	J. Lacey, A1507, cotton dust, Cramlington, UK
A205	*N. dassonvillei*	J. Lacey, A1509, cotton dust, Cramlington, UK
A211	*N. dassonvillei*	J. Lacey, A1515, cotton dust, Cramlington, UK
• **Genus *Streptosporangium***		
TW006^T	*Str. albidum*	DSM 43870
TW001^T	*Str. amethystogenes*	DSM 43179
TW004^T	*Str. pesudovulgare*	DSM 43181
TW021^T	*Str. viridogriseum ssp. viridogriseum*	DSM 43850
TW007^T	*Str. vulgare*	DSM 43802
TW002^T	*Str. corrugatum*	DSM 43316
TW009^T	*Str. fragile*	ATCC 31519
TW022^T	*Str. nondiastaticum*	DSM 43848
TW005^T	*Str. roseum*	DSM 43021
TW006^T	*Str. viridialbum*	DSM 43801
HJ166	*Str. sp.*	H.-J. Kim; Cockle Park Experimental Farm, Northumberland, England, UK
HJ167	*Str. sp.*	H.-J. Kim; Cockle Park Experimental Farm, Northumberland, England, UK
HJ256	*Str. sp.*	H.-J. Kim; Cockle Park Experimental Farm, Northumberland, England, UK
HJ100	*Str. sp.*	H.-J. Kim; Oak copse, Corbridge, Northumberland, England, UK
HJ101	*Str. sp.*	H.-J. Kim; Oak copse, Corbridge, Northumberland, England, UK
HJ109	*Str. sp.*	H.-J. Kim; Oak copse, Corbridge, Northumberland, England, UK
HJ117	*Str. sp.*	H.-J. Kim; Oak copse, Corbridge, Northumberland, England, UK
HJ118	*Str. sp.*	H.-J. Kim; Oak copse, Corbridge, Northumberland, England, UK
HJ121	*Str. sp.*	H.-J. Kim; Oak copse, Corbridge, Northumberland, England, UK
• **Genus *Saccharomonospora***		
K168^T	*Sac. cyanea*	H. Runmao (New Antibiotic Research Department, Sichuan Industrial Institute of Antibiotics, Chengdu, Sichuan, People's Republic of China), NA-134 (SIIA 86134); soil, Guangyuan City, Sichuan, People's Republic of China
K201	*Sac. sp.*	J. Ruan (Institute of Microbiology, Academia Sinica, Zhong Guan Cun, Beijing 100080, People's Republic of China), 4029; soil
K169^T	*Sac. glauca*	DSM 43769

APPENDIX J. CONTINUED.

Laboratory numbers	Taxa	Strain histories
K170	*Sac. glauca*	E. Greiner-Mai (Institut für Mikrobiologie, Technische Hochschule Darmstadt, D-6100 Darmstadt, Germany), AA10
K171	*Sac. glauca*	E. Greiner-Mai, Än1
K172	*Sac. glauca*	E. Greiner-Mai, Äv1
K173	*Sac. glauca*	E. Greiner-Mai, KE4
K174	*Sac. glauca*	E. Greiner-Mai, Ko29
K175	*Sac. glauca*	E. Greiner-Mai, MK18a
K176	*Sac. glauca*	E. Greiner-Mai, R15
K194	*Sac. viridis*	J. Lacey, A66; hay, Rothamsted (CUB 62)
K195	*Sac. viridis*	J. Lacey, A1450; air sample, cotton mill, Cramlington, England, UK
K179	*Sac. viridis*	A. J. McCarthy (School of Biological Sciences, University of Bradford, UK), BD-125; hay, Rothamsted (A969; CUB 614)
K202	*Sac.* sp.	J. Ruan, 350 ; soil
K73^T	*Sac. viridis*	E. Küster (*Thermomonospora viridis*), P101 ; peat, Ireland
K197	*Sac. viridis*	J. Lacey, A1905, A1906; mushroom compost, Avon, England, UK
K198	*Sac. viridis*	J. Lacey, A1905, A1906; mushroom compost, Avon, England, UK
K177	*Sac. viridis*	A.J. McCarthy, BD-42, BD-89 ; grass compost
K178	*Sac. viridis*	A.J. McCarthy, BD-42, BD-89 ; grass compost
K181	*Sac. viridis*	E. Greiner-Mai, E13
K183	*Sac. viridis*	E. Greiner-Mai, Ko27
K184	*Sac. viridis*	E. Greiner-Mai, Ko33
K185	*Sac. viridis*	E. Greiner-Mai, L1v
K186	*Sac. viridis*	E. Greiner-Mai, MK5v
K187	*Sac. viridis*	E. Greiner-Mai, MK22
K188	*Sac. viridis*	E. Greiner-Mai, R18
K189	*Sac. viridis*	E. Greiner-Mai, R22
K190	*Sac. viridis*	E. Greiner-Mai, R24

^{T}, Type strains. Abbreviations: ATCC, American Type Culture Collection, 12301 Parklawn Drive, Rockville, Md., USA; DSM, Deutsche Sammlung von Microorganismen und Zellkulturen, Mascheroder Weg 1b, D-38124, Braunschweig, Germany; IFO, Institute for Fermentation, Osaka, Japan; IMRU, Institute of Microbiology, Rutgers State University, New Brunswick, N.J., USA.; JCM, Japan Collection of Microorganisms, Saitama, Japan; NCIMB, National Collection of Industrial and Marine Bacteria, St. Machar Drive, Aberdeen, Scotland, UK; NCTC, National Collection of Type Cultures, Central Public Health Laboratory, London, UK; and UQM, University of Queensland, Brisbane, Australia.