THE SOCIETY FOR
GENERAL MICROBIOLOGY

Symposium 32

Edited by M. J. CARLILE, J. F. COLLINS and
B. E. B. MOSELEY

Molecular and Cellular Aspects of Microbial Evolution

CAMBRIDGE UNIVERSITY PRESS

MOLECULAR AND CELLULAR ASPECTS OF MICROBIAL EVOLUTION

EDITED BY
M. J. CARLILE, J. F. COLLINS AND
B. E. B. MOSELEY

THIRTYSECOND SYMPOSIUM OF THE SOCIETY
FOR GENERAL MICROBIOLOGY
HELD AT
THE UNIVERSITY OF EDINBURGH
SEPTEMBER 1981

Published for the Society for General Microbiology
CAMBRIDGE UNIVERSITY PRESS
CAMBRIDGE
LONDON NEW YORK NEW ROCHELLE
MELBOURNE SYDNEY

CONTRIBUTORS

ABELSON, J., Department of Chemistry, University of California, San Diego, La Jolla, California 92093, USA

BAUMBERG, S., Department of Genetics, University of Leeds, Leeds LS2 9JT, UK

CAMMACK, R., Department of Plant Sciences, University of London King's College, 68 Half Moon Lane, London SE24 9JF, UK

CAVALIER-SMITH, T., Department of Biophysics, University of London King's College, 26–29 Drury Lane, London WC2B 5RL, UK

CORDINGLEY, J. S., MRC Biochemical Parasitology Unit, The Molteno Institute, University of Cambridge, Downing Street, Cambridge CB2 3EE, UK

CULLUM, J., Max-Planck-Institut für Züchtungsforschung, 5 Köln 30 (Vogelsang), German Federal Republic

DAWES, IAN W., Department of Microbiology, University of Edinburgh, Edinburgh, EH9 3JG, UK

DEVOS, R., Laboratory of Molecular Biology, State University of Ghent, Ledeganckstraat 35, B-9000, Ghent, Belgium

FANG, R.-X., Laboratory of Molecular Biology, State University of Ghent, Ledeganckstraat 35, B-9000, Ghent, Belgium

FIERS, W., Laboratory of Molecular Biology, State University of Ghent, Ledeganckstraat 35, B-9000, Ghent, Belgium

FINCHAM, J. R., Department of Genetics, University of Edinburgh, Edinburgh EH9 3JN, UK

GARLAND, P. B., Biochemistry Department, Medical Sciences Institute, University of Dundee, Dundee DD1 4HN, Scotland, UK

HUYLEBROECK, D., Laboratory of Molecular Biology, State University of Ghent, Ledeganckstraat 35, B-9000, Ghent, Belgium

JOHNSON, J. D., Department of Chemistry, University of California, San Diego, La Jolla, California 92093, USA

JOHNSON, P. F., Department of Chemistry, University of California, San Diego, La Jolla, California 92093, USA

KNAPP, G., Department of Chemistry, University of California, San Diego, La Jolla, California 92093, USA

KREBS, H., Metabolic Research Laboratory, Nuffield Department of Clinical Medicine, Radcliffe Infirmary, Oxford OX2 6HE, UK

MIN JOU, W., Laboratory of Molecular Biology, State University of Ghent, Ledeganckstraat 35, B-9000, Ghent, Belgium

OGDEN, R. C., Department of Chemistry, University of California, San Diego, La Jolla, California 92093, USA

PEEBLES, C. L., Department of Chemistry, University of California, San Diego, La Jolla, California 92093, USA

RAO, K. K., Department of Plant Sciences, University of London King's College, 68 Half Moon Lane, London SE24 9JF, UK

SAEDLER, H., Max-Planck-Institut für Züchtungsforschung, 5 Köln 30 (Vogelsang), German Federal Republic

STRACKEBRANDT, E., Technical University Munich, Arcisstr. 21, 8000 Munich 2, German Federal Republic

TURNER, M. J., MRC Biochemical Parasitology Unit, The Molteno Institute, University of Cambridge, Downing Street, Cambridge CB2 3EE, UK

VERHOEYEN, M., Laboratory of Molecular Biology, State University of Ghent, Ledeganckstraat 35, B-9000 Ghent, Belgium

WOESE, C. R., Department of Genetics and Development, College of Liberal Arts and Sciences, University of Illinois, 515 Morrill Hall, Urbana, IL 61801, USA

EDITORS' PREFACE

On the occasion of this Edinburgh Symposium, the Society for General Microbiology has returned to the theme of evolution, to which the Twenty-fourth Symposium in 1974, entitled 'Evolution in the Microbial World', was devoted. Evolution forms a natural meeting point for many areas of research which are of interest and importance to the Society's members, and we hope that this will recommend the Symposium to a wide audience.

Since 1974, progress in all fields has been rapid, and with the improvements in methodology pertinent to the study of microbial evolution, particularly of analysis at the molecular level, this book may be regarded as more a sequel than a companion to the previous volume.

Within such a unifying theme, however, there is a diversity of approach possible, and this is well illustrated here. For example, evolution is viewed both in an historical perspective and as a process which is still occurring in Nature. The contrast between these two approaches is very marked, the former being necessarily speculative and inferential, while the latter is essentially documentary; yet both are based upon the most recent experimental techniques.

The contributions have been chosen to cover a wide range of topics, including evolution of the major groups of microorganisms, aspects of their development, and analyses of the ways in which metabolic pathways and their necessary and sophisticated controls may have evolved. In these areas, we are made aware of the increasing role of studies on the sequences and properties of nucleic acids, which form the focus of another set of contributions. There is no doubt that nucleic acid sequence studies will be a rich source of evolutionary paradigms for a long time to come, and will play a unique role in linking or discriminating between groups of living organisms. One of the pleasures of being an Editor for this volume has been to sense the momentum in this field, as the manuscripts have arrived from our contributors, each one quickening the pace such that, in this instance, we may justifiably claim that the whole is greater than the sum of its parts.

This volume can also be regarded as a prelude of things to come, and nothing in it is necessarily final. As C. H. Waddington (in *Towards a Theoretical Biology*, ed. C. H. Waddington. Edinburgh University Press, 1968, p. 108) said, 'After all, evolution has had a

long time to cook up some really clever tricks'. We are confident that the Society will find it appropriate, from time to time in the future, to select the theme of evolution for its symposia, and we in our turn look forward to reading the next volume in this series.

Finally, we would like to thank all the authors and those members of the SGM Council and the Cambridge University Press who have been involved in the successful production of this volume.

M. J. CARLILE
J. F. COLLINS
B. E. B. MOSELEY

CONTENTS

THE EVOLUTION OF PROKARYOTES

ERKO STACKEBRANDT* AND CARL R. WOESE†

**Department of Microbiology, Technical University Munich, Arcisstr. 21,8000 Munich 2, German Federal Republic*
†Department of Genetics and Development, College of Liberal Arts and Sciences, University of Illinois, 515 Morrill Hall, Urbana, IL 61801, USA

INTRODUCTION

The microbiologist has sought for a century to establish the natural relationships among the myriad bacterial species. This has been largely a frustrating task because of the simplicity of their morphologies and other characteristics. In the higher forms where morphologies are indeed complex, morphological convergence is for the most part ruled out, and morphology is then a reliable phylogenetic indicator. However, distinctions involving spherical, rod, and spiral shapes, etc., are clearly not sufficient either reliably to group bacteria phylogenetically or necessarily to exclude species from groups so defined. The caveat concerning the use of these simple characters in attempting to determine the natural relationships among bacteria has been pronounced many times. Yet, to this day, morphological characters have been heavily relied upon for classification of bacteria simply because no better criteria existed until recently. It is obvious, therefore, that what bacterial classification we have (say up through the eighth edition of Bergey's Manual (1923–1974)) is probably not in very good accord with the natural relationships that exist among organisms.

Genetic sequence is an historical record (Bryson & Vogel, 1965; Zuckerkandl & Pauling, 1965*a, b*). Comparative analysis of genetic sequence can then be used to establish genealogical relationships among organisms. There are many approaches now available that directly or indirectly reflect genetic sequence to one degree or another, including the ultimate one of exact sequence determination. Although of limited use, one of the best techniques in principle is DNA–DNA hybridization. The method is relatively simple, rapid, and inexpensive. More importantly, the method gives an averaged measure for the entire genome, and so, unlike most other methods, it is necessarily representative of the whole organism. Unfortunately, this method permits detection of only the closest

genealogical relationships among bacteria, failing above the intrageneric level (Johnson, 1973; Steigerwalt, Fanning, Fife-Ashbury & Brenner, 1976) and this makes it of limited utility.

Most other methods that compare genetic sequence reflect one or a few genes only. Given that individual bacterial genes are often subject to lateral (interspecific) transfer, there is then a serious question as to whether a method based on a single gene reflects true bacterial phylogeny or merely the phylogeny of that gene *per se*. In fact, if lateral transfer of bacterial genes were extensive enough, there could be no such thing as a phylogeny representative of the whole bacterium. Fortunately this seems not to be the case. A prediction of the lateral transfer hypothesis is that independent genes (or gene clusters) will not exhibit the same pattern of interspecific transfer. In other words, a set of phylogenetic relationships determined with one gene would not be the same as a set determined with genes unrelated to that gene. As we will see below, phylogenies of an extensive group of bacteria determined by the use of two independent genes, give practically identical trees.

What is the optimal system for making phylogenetic measurements among bacteria? There are a number of requirements. First, the system must not be subject to appreciable lateral transfer. (This would rule out antibiotic resistance factors, nitrogen fixation, and so on.) Second, the system should be universally distributed. Third, the system must exhibit functional constancy; i.e., one does not want *selected* mutations (as opposed to neutral single or multiple mutations) to distort the measurement, to give the appearance of saltatory evolution. Fourth, the gene(s) involved has (have) to provide a sufficiently slow 'clock', i.e., change in sequence slowly enough that the largest of genealogical distances can be detected (which is definitely not the case, for example, with DNA–DNA hybridization methods). And finally, the system has to be an experimentally feasible one.

We decided more than a decade ago that the 16S ribosomal RNA was well suited to the purpose of measuring genealogical relationships among bacteria, and hopefully for constructing the universal phylogenetic tree as well. The molecule was universally distributed. It was easily isolated. It appeared to be highly constant in function (as ribosome reconstitution experiments demonstrated) (Nomura, Traub & Bechmann, 1968; Higo, Held, Kaham & Nomura, 1973). Parts of it, at least, seemed to change very slowly with time, as rRNA–DNA hybridization studies had shown (Pace

& Campbell, 1971; Moore & McCarthy, 1967) and, although the molecule was too large to sequence, it was possible, using the then current nucleic-acid-sequencing technology, to sequence large enough fragments of it to make feasible a comparative analysis of its primary structure. Moreover, the molecule was large (about 1 540 residues) which seems to give it a useful, but more subtle, advantage. Smaller molecules, e.g., cytochrome *c* and the 5S rRNA, exhibit saltatory evolutionary behaviour when the structure of one of their 'domains', loosely speaking, changes; we have noted several examples in the 5S rRNA in which the sequence of one of its four helical elements appears to change drastically for a given organism (Woese *et al.*, 1976*a*). These saltatory changes, undoubtedly involving some strongly selected mutations, distort the phylogenetic picture. For a large molecule like 16S rRNA, which has about fifty helical elements (Woese *et al.*, 1980*b*) a drastic 'redesigning' of any one of them would have far less effect on the apparent phylogenetic distance measure than would be the case for a small molecule like 5S rRNA, making the former the more accurate phylogenetic indicator.

The choice of 16S rRNA for measuring phylogenetic relationships among bacteria has proved to be a good one. Not only can the molecule span the greatest phylogenetic distances (Woese & Fox, 1977*a*), because parts of its sequence change only slowly with time, but it can be used to measure close phylogenetic relationships as well (Zablen *et al.*, 1975*a*; Stackebrandt *et al.*, 1980*b*) because other parts of its sequence change relatively rapidly with time. The main limitations of the technique are its expense and the fact that it is too slow to be used on thousands of bacterial species. Nevertheless, compared to the amount of time and effort that have gone into attempts to establish bacterial phylogenetic relationships in the past, the expense and time used for this one seem slight. Another potential drawback to the method in some cases, the necessity to incorporate moderately high levels of ^{32}P into the RNA of the growing cell, which is not always possible due to medium composition, pool levels, or radiation damage to cells, has now been alleviated by the development of *in vitro* labelling methods for the rRNA digest (E. Stackebrandt, W. Ludwig, K. H. Schleifer & H. J. Gross, 1981). The 16S rRNA cataloguing approach, as our method is known, should be used for establishing the major phylogenetic units down to the level of what is conventionally seen as a genus, e.g., the more 'recent' genera like *Escherichia,*

Arthrobacter, and so on. Within these recent genera, the bulk of the species can then more rapidly and less expensively be interrelated by the technique of DNA–DNA hybridization.

THE METHOD

The cataloguing of 16S rRNA is performed in one of two related ways. In the original method, a bacterial culture, typically 10–50 ml, is labelled in exponential phase growth with $^{32}PO_4$, and the 16S rRNA isolated by standard techniques, e.g., phenol extraction and polyacrylamide gel electrophoretic separation. Ideally 100 μg or more of purified, labelled 16S rRNA at a specific activity of roughly 1 μCi/μg is required. The RNA is then digested by T1 ribonuclease, to produce a set of oligonucleotides, usually up to 15–20 residues in length, each ending in a G residue and preceded by some number (from zero up) of non-G residues. These oligonucleotides are resolved from one another by the two dimensional paper electrophoretic method originally developed by Sanger and coworkers (Sanger, Brownlee & Barrell, 1965) and modified by us (Uchida *et al.*, 1974; Woese, Luehrsen, Pribula & Fox, 1976). This employs a cellulose acetate first dimension at pH 3.5 followed by transfer to DEAE cellulose, which is then run in 0.1 M buffer at a pH of about 2.3. (For a typical fingerprint, see Uchida *et al.*, 1974 and Balch *et al.* 1979.) The individual oligonucleotides are then sequenced by a combination of endonuclease digestion procedures, to produce finally a list, or catalogue, of sequences that is characteristic of the organism in question.

Alternatively, unlabelled 16S rRNA is digested with T1 ribonuclease, the terminal phosphate(s) removed, and a ^{32}P labelled phosphate enzymatically placed on the 5′ end of each oligonucleotide (Stackebrandt, Ludwig, Schleifer & Gross, 1980*b*). Separation of the oligonucleotides is then performed as described above or, more usually, the second dimension DEAE paper is replaced by a DEAE cellulose thin-layer plate, which is developed in a relatively high-salt-buffer system. Sequencing of the individual oligonucleotides in this case is somewhat different, since only the 5′ terminus is labelled, and involves a two-dimensional method of separating partial digests of the oligonucleotide (Silberklang, Gillum & RajBhandary, 1979).

Oligonucleotide catalogues of organisms are then generally analy-

sed by a binary method; each catalogue is compared individually to every other catalogue, and the oligonucleotides, of six residues or larger, common to any two catalogues are scored to produce a 'S_{AB} value' characteristic of that pair of organisms. (The function S_{AB} is defined as twice the total number of residues in sequences common to a pair of catalogues, divided by the total number of residues in all sequences in the two catalogues, consideration being confined, as stated, to hexamers and larger. These S_{AB} values range from 1.0 for identical RNAs to about 0.03 for randomly related sequences of 1 500 nucleotides length.) A table of S_{AB} values for any given set of organisms is then analysed by standard clustering procedures (average linkage among merged groups) to produce a dendrogram (Fox, Pechman & Woese, 1977).

The dendrogram so produced is a reasonable approximation to the true phylogenetic relationships provided that the 'mutational clocks' in all organisms are isochronic; in other words, all organisms introduce mutations into rRNA at the same rate. As we shall see, this is the case for the bacteria with a small number of interesting exceptions.

THE PRIMARY PHYLOGENETIC DIVISIONS

The ribosomal RNA cataloguing method is able to detect the most distant phylogenetic relationships. The lowest S_{AB} values, observed among the primary groups, are in the range of 0.10, which corresponds roughly to 50–60 residues in common oligonucleotides (hexamers and larger). The method can therefore be used to identify and define what should be called the 'primary kingdoms', or 'urkingdoms', those major phylogenetic units that directly stem from the common ancestor of all extant life (Woese & Fox, 1977*a*). The term 'primary kingdom' is used to distinguish these fundamental phylogenetic units from the classically defined 'kingdoms', such as animals, plants, etc. The classical eukaryotic kingdoms are not related to the primary kingdoms; the former are phylogenetic groupings within the eukaryotic domain of organization, whereas the primary kingdoms are defined for the underlying prokaryotic domain. (A more complete discussion of this question of phylogenetic units and domains, or levels, of biological organization can be found in Woese & Fox (1977*a*).)

Ribosomal RNA cataloguing shows that there exist at least three

primary kingdoms, or three primary lines of descent. This initial discovery came as a considerable shock to the scientific community, for biologists had for some time accepted that the dichotomy eukaryote–prokaryote somehow defined not only mutually exclusive categories of cell types, but in addition, mutually exclusive phylogenetic categories (Chatton, 1937; Allsopp, 1969; Margulis, 1970; Murray, 1974). In other words, all organisms were seen as belonging either to a prokaryotic or to an eukaryotic line of descent. As it turns out, there exist at least two distinct bacterial lines of descent, lines that are no more related to one another than either of them is to the 'eukaryotic line of descent' (see below).

The three primary kingdoms then are these (Woese & Fox, 1977*a*; Fox *et al.*, 1980): (I) A grouping that includes the vast majority of recognized bacteria. To date, this primary kingdom can also be defined as those organisms that do, or whose ancestors did, possess the muramic-acid type of cell wall. The true mycoplasmas, which come from an ancestry within the classical Gram-positive bacteria, are included herein (see below), and both the chloroplast and (plant) mitochondrion trace their ancestry back to this kingdom. This primary kingdom has been called the *true bacteria* or *eubacteria*. ('Eubacteria' is a term that has been used in many contexts, and microbiologists may be reluctant to see it in yet another.) (II) The second primary kingdom is known as the archaebacteria. These organisms are bacterial in size and simplicity, morphological and genetic, but differ from true bacteria in the details of most, if not all, of their organization at the molecular level. At present, the archaebacterial group is known to contain only three (rather bizarre) phenotypes, the methanogens, the extreme halophiles, and certain extreme thermoacidophiles. (III) The third primary kingdom may not exist in the strictest sense of the word. At present it is known to be represented only by the 18S rRNA of the eukaryotic cell; a prokaryotic example of the group has yet to be found. It is tempting to say that such should be called the eukaryotic line of descent, or the eukaryotic primary kingdom. However, the matter of eukaryotic origins is complicated and not yet well understood. The mitochondrion and chloroplast are of true bacterial ancestry (Bonen & Doolittle, 1975; Zablen, Kissil, Woese & Buetow, 1975*b*; Bonen, Cunningham, Gray & Doolittle, 1977). One of the eukaryotic ribosomal proteins seems to be of archaebacterial ancestry (Matheson, Möller, Amons & Yaguchi, 1980). The eukaryotic 18S rRNA, as we have just seen, seems of an ancestry neither

archaebacterial nor eubacterial. Therefore, the eukaryotic cell is a phylogenetic chimera; how radical a chimera, i.e., how many gene or gene cluster 'capture' events are involved we have no idea. At least until the genealogies for many of the eukaryotic gene families are traced into the prokaryotic realm, i.e., the various primary kingdoms they represent are identified, it is not useful to speak of an ancestral eukaryotic line of descent; there may be no single line of descent that accounts for enough of the eukaryotic gene families to be called *the* eukaryotic line of descent. What we do now know, however, is that there exist two primary kingdoms of bacteria, the true bacteria and the archaebacteria, and that these two lines of descent, along with others yet to be defined at the prokaryotic level, are variously represented in the genetic chimera that is the eukaryote. And for the present, we should leave it at that. The eukaryotes will not be discussed further in this chapter.

THE PHYLOGENETIC STRUCTURE OF THE TRUE BACTERIA

Conventionally, the true bacteria are divided into Gram-positive and Gram-negative groups, with some uncertainty and debate surrounding the cyanobacteria and mycoplasmas. As we will see, this grouping is only partially in accord with the phylogenetic structure of the true bacteria. Fig. 1 is an overview of the phylogeny of the true bacteria, as seen in terms of the rRNA cataloguing method.

The Gram-positive eubacteria

With the exception of the *Micrococcus radiodurans* group, whose members possess atypical walls and other features (Brooks *et al.*, 1980), the Gram-positive bacteria form a phylogenetically coherent unit, albeit a deep, i.e., ancient, one. By and large, the conventional separation of low G + C DNA-content organisms within the group from high G + C ones is seen to hold (Figs 2 and 3). In other words, the actinomycete–coryneform type of phenotype is phylogenetically distinct from the clostridium–bacillus–streptococcus type. However, in detail, there are some surprises and considerable rearrangement of the traditional groupings.

The high G + C Gram-positive bacteria are seen to structure

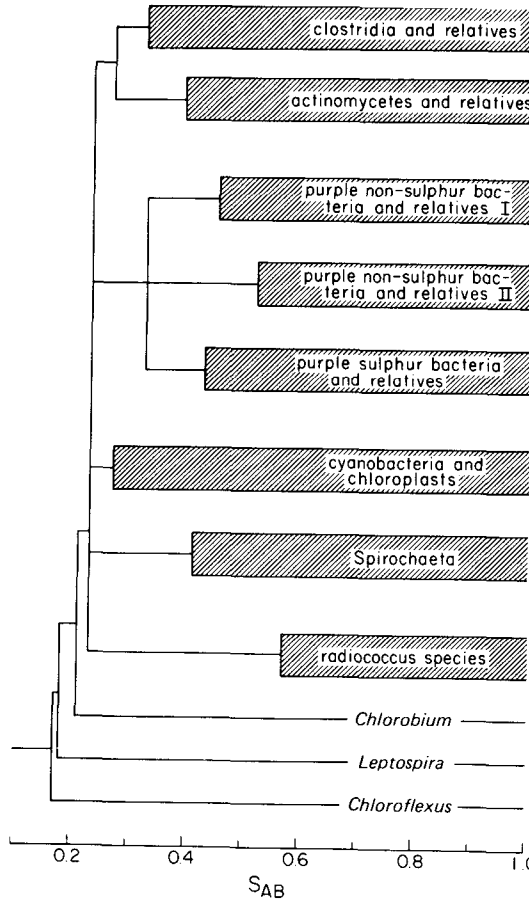

Fig. 1. Dendrogram of relationships among the true bacteria. Organisms forming the clusters of clostridia, actinomycetes, cyanobacteria, and purple non-sulphur bacteria are listed in Figs. 2 and 4–7 (Fox *et al.*, 1980; C. R. Woese, P. Blanz & C. M. Hahn, unpublished). Purple sulphur group: *Aerobacter aerogenes, Aeromonas hydrophila, Chromatium vinosum, Escherichia coli, Oceanospirillum maris, O. minutulum, Pasteurella multocida, Photobacterium phosphoreum, Proteus mirabilis, Pseudomonas aeruginosa, P. alcaligenes, P. fluorescens, P. pseudoalcaligenes, P. putida, P. stutzeri, P. syringae, Serpens flexibilis, Serratia marcescens, Thiocapsa pfennigii, Vibrio marinus, Yersinia pestis.* The *Spirochaeta* cluster contains *S. aurantia, S. halophila, S. litoralis, S. stenostrepta.* The cluster of the radio-resistant micrococci is defined by *M. radiodurans, M. radiophilus, M. roseus* UWO 294 (University of Western Ontario).

phylogenetically in the following way (Fig. 2). The members of *Arthrobacter* form a major subunit along with *Cellulomonas,* the plant pathogen coryneforms, *Microbacterium* and certain other genera. Peripherally these are related to members of the genus *Actinomyces.* It is surprising that the genuine species of *Micrococcus*

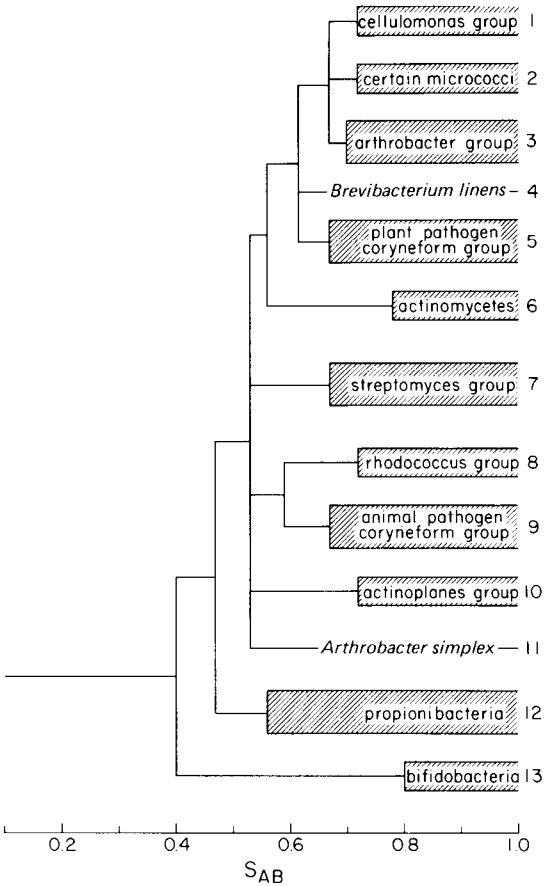

Fig. 2. Dendrogram of relationships among Gram-positive bacteria with a high DNA G + C content (≥55 Mol%). *Cellulomonas: Cellulomonas cartae, Cell. flavigena, Nocardia cellulans, Oerskovia turbata. Micrococcus: Micrococcus sedentarius, M. nishinomiyaensis. Arthrobacter: Arthrobacter globiformis, A. oxidans, A. atrocyaneus, M. luteus, M. lylae, M. roseus, M. varians.* Plant pathogen coryneforms: *Microbacterium lacticum, Corynebacterium betae, C. mediolanum.* Actinomycetes: *Actinomyces bovis, Act. viscosus. Streptomyces: Actinomadura dassonvillei, Chainia antibiotica, Elythrosporangium brasiliense, Kitasatoa kauaiensis, Microellobosporia cinerea, Streptomyces griseus, Streptosporangium roseum, Streptoverticillium baldacchi. Rhodococcus* group: *C. fascians, Nocardia corallina, N. calcarea.* Animal pathogen coryneforms: *C. diphtheriae, A. variabilis, C. glutamicum. Actinoplanes: Actinoplanes philippinensis, Ampulariella regularis, Dactylosporangium aurantiacum, Micromonospora chalcea.* Propionibacteria: *P. freudenreichii, P. acnes.* Bifidobacteria: *Bif. bifidum, Bif. breve.*

also fit into this grouping, and more surprising that these cannot be phylogenetically separated from species of *Arthrobacter*. It appears then that *Micrococcus* is not a phylogenetically valid genus. Rather the micrococci seem to have arisen as degenerate forms of the arthrobacteria, locked into the coccoid stage of the arthrobacterial