PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Biodiversity at the Molecular Level: The Domains, Kingdoms and Phyla of Life

T. Martin Embley, Robert P. Hirt and David M. Williams

Phil. Trans. R. Soc. Lond. B 1994 **345**, 21-33
doi: 10.1098/rstb.1994.0083

Biodiversity at the molecular level: the domains, kingdoms and phyla of life

T. MARTIN EMBLEY[1], ROBERT P. HIRT[1] AND DAVID M. WILLIAMS[2]

Departments of [1]Zoology and [2]Botany, The Natural History Museum, Cromwell Road, London SW7 5BD, U.K.

CONTENTS

	PAGE
1. Introduction: a microbiological perspective	21
2. The primary divisions of life	22
3. Diversity of uncultured Archaea	25
4. Domain Bacteria	26
5. Diversity of uncultured Bacteria	26
6. Domain Eucarya	28
7. Microbial diversity: a conclusion	30
References	31

SUMMARY

The results of comparative sequence analysis, mainly of small subunit (SSU) ribosomal (r)RNA sequences, have suggested that all of cellular life can be placed in one of three domains: the Archaea, Bacteria or Eucarya. There is some evidence that the Archaea may not be a monophyletic assemblage, but as yet this issue has not been resolved. Most of the lineages, and all of the deepest ones, in the tree based upon SSU rRNA sequences, are microbial. Traditional ideas of classification such as Whittaker's five kingdom scheme do not adequately describe life's diversity as revealed by sequence comparisons. There are many microbial groups that demonstrate much greater amounts of SSU rRNA sequence divergence than do members of the classical kingdoms, Animalia, Plantae and Fungi. The old microbial kingdoms Monera and Protista are clearly paraphyletic but as yet there is no consensus as to how they should be reorganized in taxonomic terms. New data from environmental analysis suggests that much of the microbial world is unknown. Every environment which has been analysed by molecular methods has revealed many previously unrecorded lineages. Some of these show great divergence from the sequences of cultured microorganisms suggesting that fundamentally new microbial groups remain to be isolated. The relationships of some of these new lineages may be expected to affect how the tree of life is organized into higher taxa, and to also influence which features will be recognized as synapomorphies. There is currently no objective measure whereby microbial diversity can be quantified and compared to the figures which are widely quoted for arthropods and other Metazoa.

> '"I tell you what", said William, confidingly, "let's say eggs for both of them. Then we shan't get so muddled."'
>
> (from *William the Pirate*, Richmal Crompton, 1932)

1. INTRODUCTION: A MICROBIOLOGICAL PERSPECTIVE

Most discussions of biodiversity focus on Metaphyta and Metazoa and seldom do they consider microorganisms. To microbiologists this appears unjustified; practically all key environmental processes are driven by microbial activity and numerically at least, microbial diversity can be expected to be the most abundant. For instance, microorganisms will include all prokaryotes and practically all taxa termed, at one time or another, protists. But these names are problematic as both 'prokaryote' and 'protist' have undergone radical re-definition, and as indicators of 'natural' taxa they appear to have limited immediate value. Value in the sense that the common term 'animal' will be more or less understood by most, the term 'protist' is not, even by those whose focus on the problem directly. These issues are developed below. For the purposes of this paper we will use the term prokaryote to indicate organisms which lack a nucleus and the term protist to indicate those organisms that are not metaphytes, Metazoa, 'true' fungi or 'prokaryotes'. This cumbersome definitional problem is symptomatic: microbial diversity has been so little studied that precision over higher taxon

Phil. Trans. R. Soc. Lond. B (1994) **345**, 21–33
Printed in Great Britain

definition is, and may remain, problematic for a while to come.

The scale of the microbial world is immense. For example, Fenchel (1992) recently estimated that a one centimetre core of coastal marine sediment may contain 4×10^{10} bacterial cells, 10^4 heterotrophic flagellates and amoebae, 10^8 chlorophyll a-containing microorganisms and about 2000 ciliates. As previously mentioned, microorganisms are fundamental to biological nutrient cycling; sulphur oxidation and reduction, ammonification, nitrogen fixation, methanogenesis and methane-oxidation are the provenance of microorganisms. Photosynthesis originated in prokaryotes and tropical rain forests produce oxygen because plants contain the remnants of endosymbiotic photosynthetic bacteria.

The fossil record indicates that the timescale for microbial evolution and diversification is much greater than that for Metazoa and Metaphyta. Stromatolite communities containing oxygenic phototrophic cyanobacteria were already abundant and widespread 3.5 billion years ago (Schopf 1992). Microfossils which are recognizable as multicellular red algae are evident in rocks at least as old as 1.25 billion years (Butterfield *et al.* 1990). In contrast, the earliest recorded (Ediacaran) metazoan fossils occur in deposits which are barely 700 million years old (Gould 1989; Runnegar 1992) and the first hominid, *Australopithicus afarensis* ('Lucy'), is a mere 3.4 million years old (Kimbel *et al.* 1994). The results of molecular sequence studies published during the past 15 or so years, have now confirmed that all of the deepest divisions of life are microbial. It is the aim of the present paper to review the results of these studies, discuss their implications, and attempt to place them in context for a general understanding of all biodiversity.

2. THE PRIMARY DIVISIONS OF LIFE

When the discovery of the Archaebacteria was announced (Balch *et al.* 1977; Woese *et al.* 1978), the classical divisions of the living world were brought into a different and sharper perspective. Woese and his colleagues revolutionized the classification of life by estimating genetic diversity from comparison of small subunit (SSU) ribosomal (r) RNA sequences, to establish three groups of organisms: Eubacteria, Archaebacteria and Eukaryotes (figure 1). Prior to Woese's work, the living world was regarded as neatly subdivided into two parts based on the presence or absence of a nucleus, the Eukaryotes and Prokaryotes, respectively. Eubacteria and Archaebacteria were previously considered as subdivisions of the non-eukaryote organisms, the 'prokaryotes' referred to above. Other alternatives for the subdivision of life

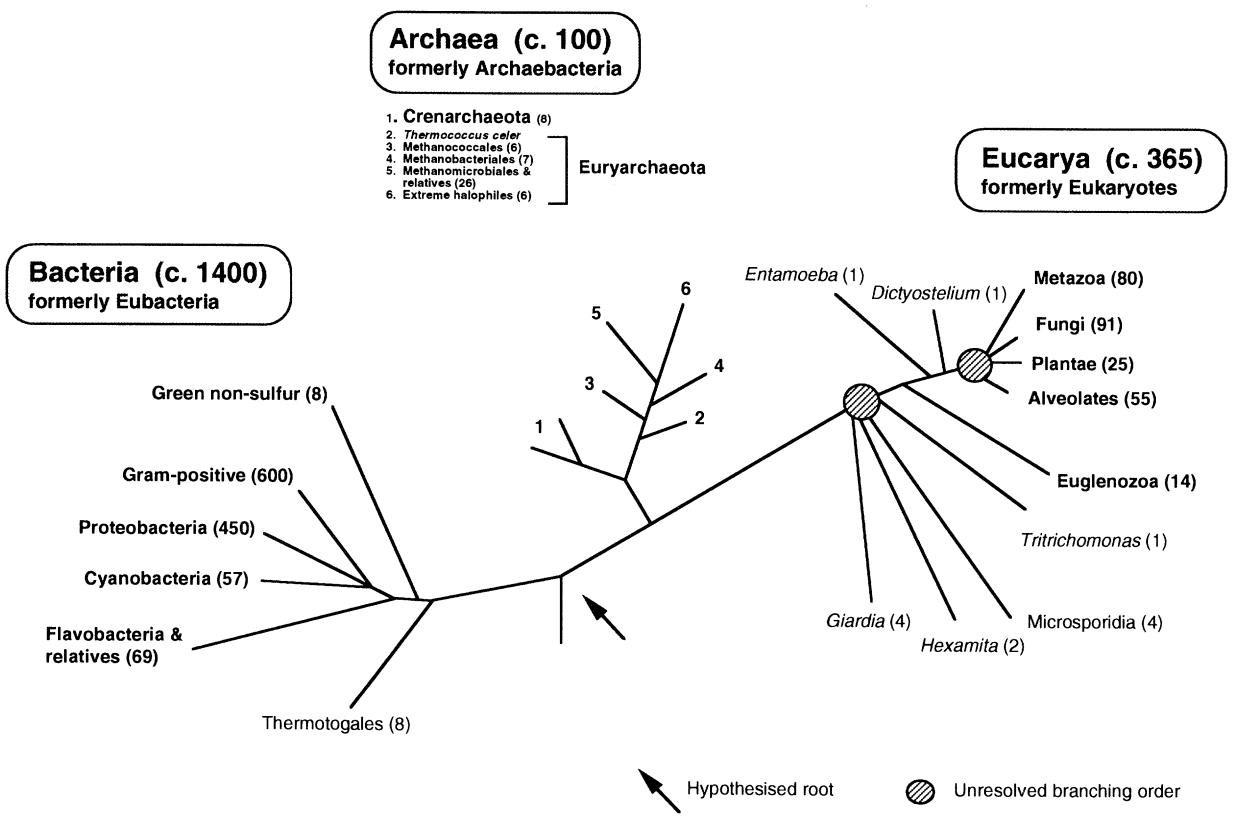

Figure 1. The domains of life based upon small subunit (SSU) ribosomal (r)RNA sequence comparisons. The figure is adapted from Woese *et al.* (1990) and shows only a representative set of lineages from each domain. Figures in parentheses indicate the approximate number of SSU rRNA sequences for taxa, the figures are extracted from Larsen *et al.* (1993) and Neefs *et al.* (1993). The root is currently thought to occur on the bacterial branch based upon the work of Iwabe *et al.* (1989) and Gogarten *et al.* (1989) but see text for discussion of this issue.

existed. Whittaker (1959) proposed a five kingdom system where prokaryotes (following the early nomenclature of Haeckel) were referred to as the Monera and microbial eukaryotes as Protista. The 'classical' kingdoms Plantae, Fungi and Animalia made up the remaining three divisions and life's diversity appeared to be very neatly encapsulated in such a system. Subsequent study, with the bulk of confirmatory evidence coming from the SSU rRNA studies (Woese 1987; Sogin 1991), convincingly demonstrated that prokaryotes and protists were not monophyletic groups and testified conclusively to the preponderance of life's diversity residing among microorganisms. It may be salutary to consider these results in terms of what is still not known. For instance, the major question of the root of the tree of life remains. Put simply, of the three taxa Eukaryote, Archaebacteria and Eubacteria, which two are more closely related to each other than they are to the third? The answer has a direct bearing on the question of the origin of life as well as the resolution of its most fundamental divisions.

The most successful way to root a tree is by reference to taxa which are outside of the group under consideration (outgroups). Clearly this is an impossibility when dealing with all of life. Schwartz & Dayhoff (1978) suggested a means of circumventing the problem by using ancestrally duplicated (paralogous) genes, those genes which had duplicated prior to the divergence of the major clades of life. The gene tree derived from one paralogue can be used to root the gene tree from the other, and *vice versa*. Iwabe *et al.* (1989) and Gogarten *et al.* (1989) used this principle for comparing gene trees derived from the paralogues of elongation factors and ATPases. Both analyses suggested the root be placed along the bacterial, rather than the archaebacterial or eukaryote, branches. Woese *et al.* (1990) proposed an improved classification to incorporate these findings. All three kingdoms were re-named and elevated to a new level, the domain. Thus Eubacteria was renamed as Bacteria, Archaebacteria as Archaea and Eukaryotes as Eucarya (figure 1). Such beginnings are not without controversy and the position of the root of the universal tree, as well as the related issue of the monophyly of Archaea have both been questioned from the perspective of further protein sequence data and new methods for analysis.

The position of the root linking cultured Archaea to the other Domains resolves two major subdivisions on the basis of 16*S* rRNA sequences (figure 2). One branch comprises a rather homogeneous group of sulphur-dependent, extreme thermophiles and has been named the kingdom Crenarchaeota (Woese *et al.* 1990). The second branch was named the kingdom Euryarchaeota (Woese *et al.* 1990) and is more phenotypically heterogeneous containing methanogens, extreme halophiles and miscellaneous thermophiles some of which lack cell walls (Woese 1987). Members of both branches of the archaeal tree contain rRNAs which possess some features which are not present in Bacteria or Eucarya (Winker & Woese 1991). These include a small number of homologous sequence positions where the archaeal base is different from the other two, at least two well defined structural differences (bases 500–545 and 991–1045), and a small number of insertions or deletions (Woese 1987; Winker & Woese 1991). Some of the lipids in the plasma membranes of Archaea appear to be unique, they share three simultaneous features, ether linkage, isopranic aliphatic chains and a different stereochemistry of glycerol linkage (Gambarcorta *et al.* 1994). Some Archaea do not possess a cell wall, but when they do, it is distinct from the bacterial version; they do not possess murein which with few exceptions is characteristic of Bacteria (Kandler 1994).

James Lake and his co-workers (e.g. Lake 1987*a*, 1989) also recognized the group of sulphur-dependent, extreme thermophilic Archaea (the Crenarchaeota) but called them Eocytes. Studies from Lake's laboratory suggested the possibility that rRNA sequence data may be incorrectly placing the two archaeal lineages together, due to extreme rate differences in sequence evolution of adjacent taxa. This is the so called 'long branch effect', first identified as a theoretical notion by Felsenstein (1978). Using a new and ingenious approach called evolutionary parsimony (Lake 1987b), Lake suggested that Felsenstein's theoretical idea might have empirical currency when dealing with the basal branches of the universal tree. Using evolutionary parsimony the Eocytes grouped with Eukaryotes (which together Lake called the Karyotes; Lake 1987*b*), while some representatives of the Euryarchaeota (methanogens and halophiles; Lake 1987*b*) grouped with Eubacteria (which together Lake called Parkaryotes). This aligning of the archaeal kingdoms has been referred to as the Eocyte tree, in contrast to the archaebacterial tree of Woese. In the former, Archaea are not monophyletic, in the latter they are. Lake's conclusions are controversial and have inspired much debate over both the appropriate way to analyse DNA sequence data as well as the implication behind his results (e.g Lake 1989; Olsen & Woese 1989). However, Lake has recently identified a molecular character from elongation factor genes which has bearing on the debate (Rivera & Lake 1992).

Elongation factor EF-1α (EF-Tu in Bacteria) gene sequences have a highly conserved sequence at the terminus of a β-strand near to the guanosine diphosphate (GDP) binding site. The eocytes sampled by Rivera & Lake (1992) all share an 11-amino acid sequence at this position which is also found in all eukaryotes so far analysed (human, tomato, yeast and recently *Giardia lamblia*; Hashimoto *et al.* 1994). In contrast, the Euryarchaeota *Thermococcus celer* and *Halobacterium marismortui* have a 4-amino acid motif similar to the one found in Bacteria. To establish which of these conditions is the derived state and hence which grouping is best supported, its paralogous gene EF-2 (EF-G in Bacteria) was analysed. Sequences examined from taxa across the entire spectrum of domains showed the presence of the 4-amino acid motif, indicating that the 11-amino acid motif in the EF-α gene is derived and is a putative molecular synapomorphy for eocytes plus eukaryotes.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

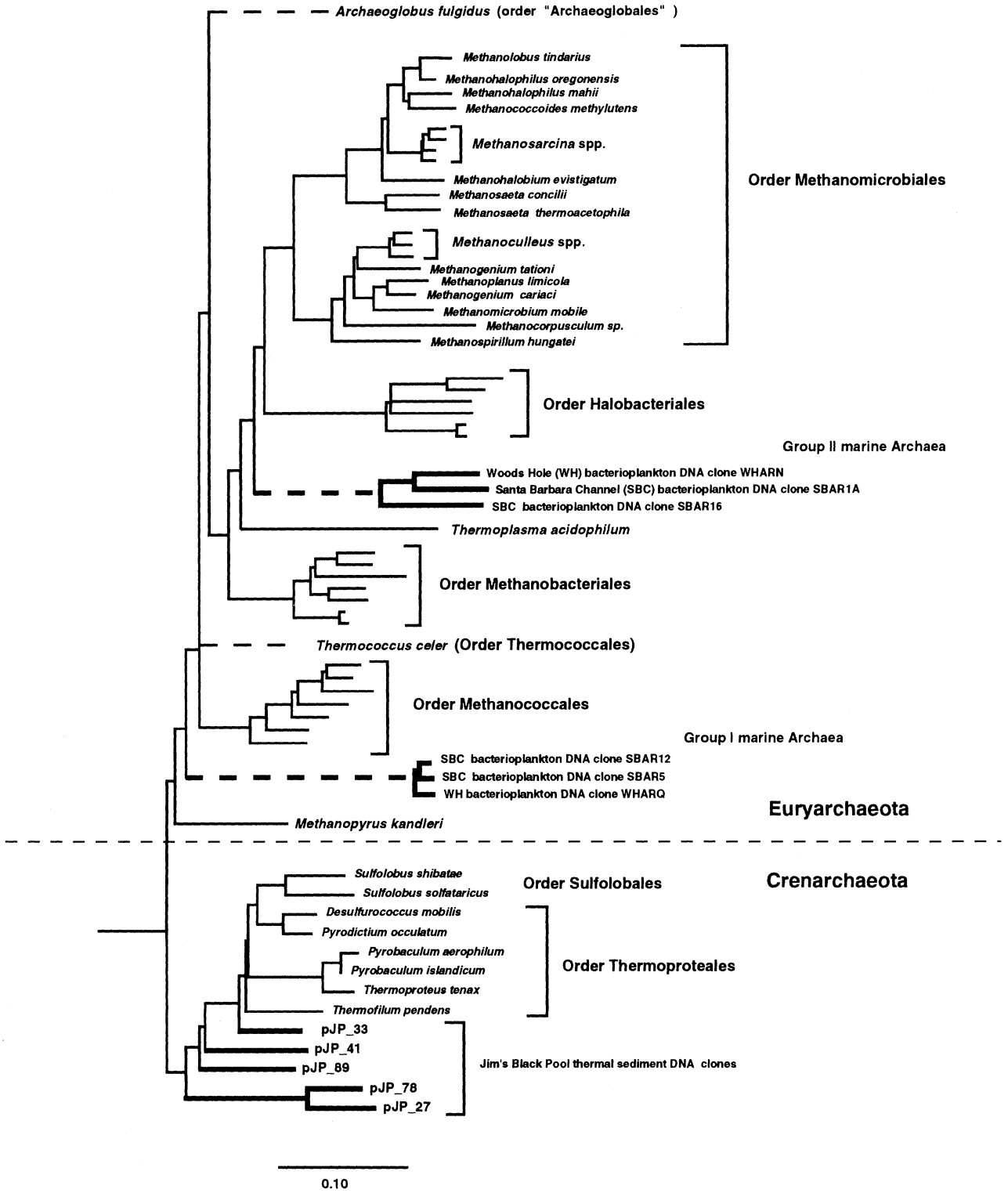

Figure 2. Relationships among Archaea based upon SSU rRNA sequences. The tree was constructed from 828 bases of SSU sequence using the Olsen Correction (G. J. Olsen, unpublished data; Larsen *et al.* 1993) and least squares analysis (DeSoete 1983). Sequences were obtained from the Ribosomal Database Project (Larsen *et al.* 1993), Barnes *et al.* (1994) and DeLong (1992). Dotted lines indicate lineages which tend to move around in different analyses. The names Group I and II marine Archaea are from DeLong (1992). Scale bar indicates 0.10 estimated substitutions per sequence position.

These data represent some of the strongest evidence so far put forward for the eocyte tree.

The resolution of archaeal monophyly by comparison of protein sequences is confounded by difficulties in sequence alignment and in interpreting possible patterns of duplication and/or horizontal transfer of genes. For instance, the elongation factor EF-1α (EF-Tu) is only half the size of EF-2(G) and they can only be aligned with each other in four short regions located at the N-terminal part of the molecule. A new

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

alignment (Creti *et al.* 1991) is different to that used by Iwabe *et al.* (1989). Again, Lake (1991) examined the effect of different alignments from these protein sequences and suggested that alignment artefact could be responsible for producing spurious trees (in his view, those that support the monophyly of Archaea). He demonstrated that under certain alignment conditions the eocyte tree receives greatest support. The protein sequences themselves may lack the appropriate information. In a detailed analysis of the EF genes, Forterre *et al.* (1993) examined an alignment across all domains. They enumerated 45 cladistically informative sites (where taxa from two of the three domains share a common amino acid) and found only one site that unambiguously grouped Eucarya with Archaea. At this level of relationship the genes coding for elongation factors appear to have limited utility. Nevertheless, recent analyses of EF-Tu/EF-1α and EF-G/EF-2 sequences (Creti *et al.* 1991; Cammarano *et al.* 1992) support the monophyly of Archaea, Bacteria and Eucarya regardless of the method of analysis. However, evolutionary parsimony, while recovering all three domains as monophyletic, has the euryarchaeote *Halobacterium* as the deepest branch in the tree based upon EF-G/EF-2, rather than the crenarchaeote *Sulfolobus* (Cammarano *et al.* 1992).

A second gene family, the ATPases, are seemingly also ambiguous on this question. Vacuolar (V) ATPase is present in Eucarya, the bacterial (F) ATPase is present in Bacteria, and the archaeal (A) ATPase is present in Archaea. All three ATPases have catalytic and non-catalytic subunits. Gogarten *et al.* (1989) suggested that the subunits were the result of an ancient gene duplication arising from a single common ancestral gene. They used this fact to root the universal tree as well as to confirm the monophyly of the three clades. Once again, the situation may be more complex. *Thermus* is a bacterium but it has an A/V-ATPase and both subunits are archaeal (Yokoyama *et al.* 1990; Tsutsumi *et al.* 1991; Gogarten *et al.* 1992), *Enterococcus* (also a bacterium) has an A/V-ATPase (Sumi *et al.* 1992) and the euryarchaeote *Methanosarcina* has an F-ATPase (Sumi *et al.* 1992). Forterre *et al.* (1993) interpreted these findings to suggest that both ATPases (A/V and F) were present in the common ancestor prior to the separation of all three clades. The first gene duplication gave rise to the two subunits, and the second yielded the A/V and F-ATPase paralogues. If this is true then the A/V- F-ATPase pair cannot be used to root the tree. Hilario & Gogarten (1993) analysed the available V, F and A-ATPase gene sequences and argued that if the duplication events occurred as Forterre *et al.* (1993) suspected, then the A/V-ATPase should reflect the correct relationships among domains. Yet their results place *Thermus* and *Enterococcus* among the Archaebacteria, a highly suspicious arrangement (Weisburg *et al.* 1989; Embley *et al.* 1993). Finally, analysis of a partial sequence from the non-catalytic subunit of F-ATPase groups *Methanosarcina* with the cyanobacteria *Anabaena* to the exclusion of other Archaea. All three of these contradictory analyses suggested to Hilario & Gogarten (1993) that explanation should be sought not in duplication events, but in horizontal gene transfer, an explanation which allows conformation to the tree derived from rRNA data.

It appears clear (if nothing else is) that the questions of archaeal monophyly and the position of the root of the tree have not yet been fully resolved. One theme that runs through such studies is the issue of sampling: how many taxa are required to reach a satisfactory solution and how much data from different genes are required? In our opinion such questions are impossible to answer with any degree of accuracy, the experiments must be done and the trees constructed and evaluated. The tree based upon SSU rRNA sequences has a role to play in helping to decide which taxa should be studied for additional gene sequences. The different results obtained using a collection of key molecules (e.g. SSU rRNA, ATPases, elongation factors), does indicate that our current knowledge of the evolutionary scenario is far from complete.

Cultured microbial diversity probably represents only a small fraction of extant microorganisms. Selective isolation, which has been a mainstay of microbial ecology, is precisely that; it selects for those organisms which are best fitted for growth on the selective medium (Wagner *et al.* (1993) have nicely demonstrated this using probes). Molecular tools are now available whereby ribosomal RNA sequences can be extracted and analysed directly from environmental nucleic acid extracts thus circumventing the need to culture prior to phylogenetic identification (Pace *et al.* 1985; Giovannoni *et al.* 1990; Ward *et al.* 1990; reviewed in Embley & Stackebrandt 1994). These tools are providing the first glimpse of a natural microbial world of potentially vast, but currently unknown dimension.

3. UNCULTURED ARCHAEAL DIVERSITY

Although methanogens have been found in virtually all anaerobic environments where fermentation takes place, most other Archaea have been sought in samples from environmental extremes. However, there is no *a priori* reason why Archaea should not be found in more mundane habitats. Recently published data have now demonstrated that Archaea are more ecologically widespread than previously thought. Molecular tools were used to detect uncultured Archaea in coastal and deep marine waters (DeLong 1992; Fuhrman *et al.* 1992, 1993). Members of the same deep branching archaeal lineage with many 'tip branches' comprising several highly related sequences, were recovered in aerobic samples from both the Pacific and Atlantic Oceans. The phylogenetic placement of this transoceanic, but previously unknown, major lineage is uncertain (we are discussing the group 1 marine Archaea in figure 2). In a recently published maximum likelihood tree they were recovered next to extremely thermophilic Crenarchaeota at the base of the archaeal tree (Olsen *et al.* 1994). The extreme thermophiles possess rRNA genes with high G + C base content (63 to 67 mol%;

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Woese *et al.* 1991) whereas the environmental sequences contain genes with a G + C content of between 51 and 54 mol%. This suggests that the environmental sequences do not originate from thermophiles, but it also complicates their comparison to those from the Crenarchaeota. Current analytical methods cannot adequately compensate for strongly opposing base compositional biases (Weisburg *et al.* 1989; Embley *et al.* 1993). Transversion analysis and signature analysis also provided support for a deep placement within the Archaea but again could not provide a precise placement (DeLong 1992).

The most recent molecular analysis of environmental Archaea focused on sediment (74°C) samples from a hot spring (Jim's Black Pool) in Yellowstone National Park (Barns *et al.* 1994). A large number of lineages was recovered, most of which clustered with sequences from the small number of cultured Crenarchaeota (figure 2). Two clones (pJP 27 and pJP 78) grouped, with only low to moderate support, with the Crenarchaeota, on the basis of comparisons of 397 bases, but shared slightly more signatures for Euryarchaeota than for Crenarchaeota. Although the root still separates the two archaeal kingdoms (figure 2), the new sequences extend the diversity of Crenarchaeota to such an extent that the boundary is beginning to blur.

4. DOMAIN BACTERIA

Most of the published 16*S* rRNA sequences are from members of the domain Bacteria; but even here coverage is limited and patchy. There are many sequences for Gram-positive bacteria and proteobacteria, groups which contain organisms of medical or commercial interest. There are fewer sequences from taxa at the base of the tree, the two deepest branching lineages *Aquifex* and the Thermotogales are currently represented by one and eight sequences, respectively (Larsen *et al.* 1993; Olsen *et al.* 1994). The 16*S* rRNA sequence of *Aquifex pyrophilus*, the deepest branch in most analyses, lacks some of the signatures which Winker & Woese (1991) have used to define the domain Bacteria. It also possesses four signatures which previously defined Archaea but which now appear to be plesiomorphic (Burggraf *et al.* 1992). As with Archaea, most of the deepest branching lineages in the bacterial domain are thermophiles, suggesting that this is a primitive trait.

The bacterial tree has been divided into a number of phyla (Woese 1987) on the basis of rRNA sequences (most of these groups are indicated on figure 3). It has been suggested that most of these could eventually be assigned the rank of kingdom (Woese *et al.* 1990). Certainly in almost all cases the degree of rRNA sequence divergence resembles or exceeds that observed between plants, animals and 'true' fungi (an impression can be gained from comparing figures 1, 3 and 4). Some of the phyla correspond to classically recognized taxa such as spirochaetes and cyanobacteria (Woese 1987; Olsen & Woese 1993). In these cases unifying phenotypic motifs can be identified. However, other phyla contain what are, from a phenotypic perspective, rather unlikely partners. Here it should be noted that there is little detailed comparative information for most prokaryote or eukaryote microorganisms and that explicit investigations for congruence between genotype and phenotype have seldom been performed. One example of a group with an unexpected composition is that which comprises *Cytophaga*, *Bacteroides* and *Flexibacter* (Woese 1987). In this case a relationship between these three taxa is supported by sequences of an unrelated gene, the β-subunit of ATPase (Amann *et al.* 1988). Some of the best examples of phenotypic diversity over relatively short amounts of rRNA sequence divergence are within the proteobacteria. This group contains most of the bacteria traditionally recognized as Gram-negative and all of the old purple photosynthetic bacteria (Woese 1987). Here intermingle a variety of very different phenotypes, including magnetotactic, nitrogen fixing, anaerobic, aerobic and photosynthetic. Many of the sequences from environmental samples fall within the radiation of proteobacteria (Giovannoni *et al.* 1990; Schmidt *et al.* 1991; Fuhrman *et al.* 1993; DeLong *et al.* 1993), it follows that it is impossible to infer a phenotype for these uncultured organisms without isolating them.

5. DIVERSITY OF UNCULTURED BACTERIA

Most prokaryotic clones in libraries constructed from environmental rRNA sequences are from Bacteria rather than Archaea (Schmidt *et al.* 1991; Fuhrman *et al.* 1992, 1993; DeLong *et al.* 1993; Weller *et al.* 1991, 1992; Liesack & Stackebrandt 1992; Stackebrandt *et al.* 1993). As yet there is too little information to decide if this represents the natural situation or a methodological bias. Almost all of the sequences discovered show sufficient divergence from those of cultured taxa to suggest they are distinct entities. All of the studies published so far have also identified some bacterial lineages which, by their depth of branching or absence of signature sequences, could be considered new phyla (Giovannoni *et al.* 1990; Britschgi & Giovannoni 1991; Schmidt *et al.* 1991; Fuhrman *et al.* 1993; Liesack & Stackebrandt 1992; Stackebrandt *et al.* 1993). Given the phenotypic versatility exhibited by cultured Bacteria, nothing can be said about the physiologies possessed by members of these new lineages. Some of the deep branching lineages from marine environments (as with the marine Archaea) are trans-oceanic and they have been detected in almost all studies. One such example is the so-called SAR 11 cluster of α-proteobacteria, members of which have been detected in coastal and deep water samples in the Pacific and Atlantic Oceans (Giovannoni *et al.* 1990; Britschgi & Giovannoni 1991; Schmidt *et al.* 1991; Fuhrman *et al.* 1993). In the only published probing experiment, rRNA from members of this group comprised approximately 12% of probe-able rRNA in a sample from the Sargasso Sea and a small but detectable amount in some other samples (Giovannoni *et al.* 1990). At the moment nothing is known about the biology of this group and we have no

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

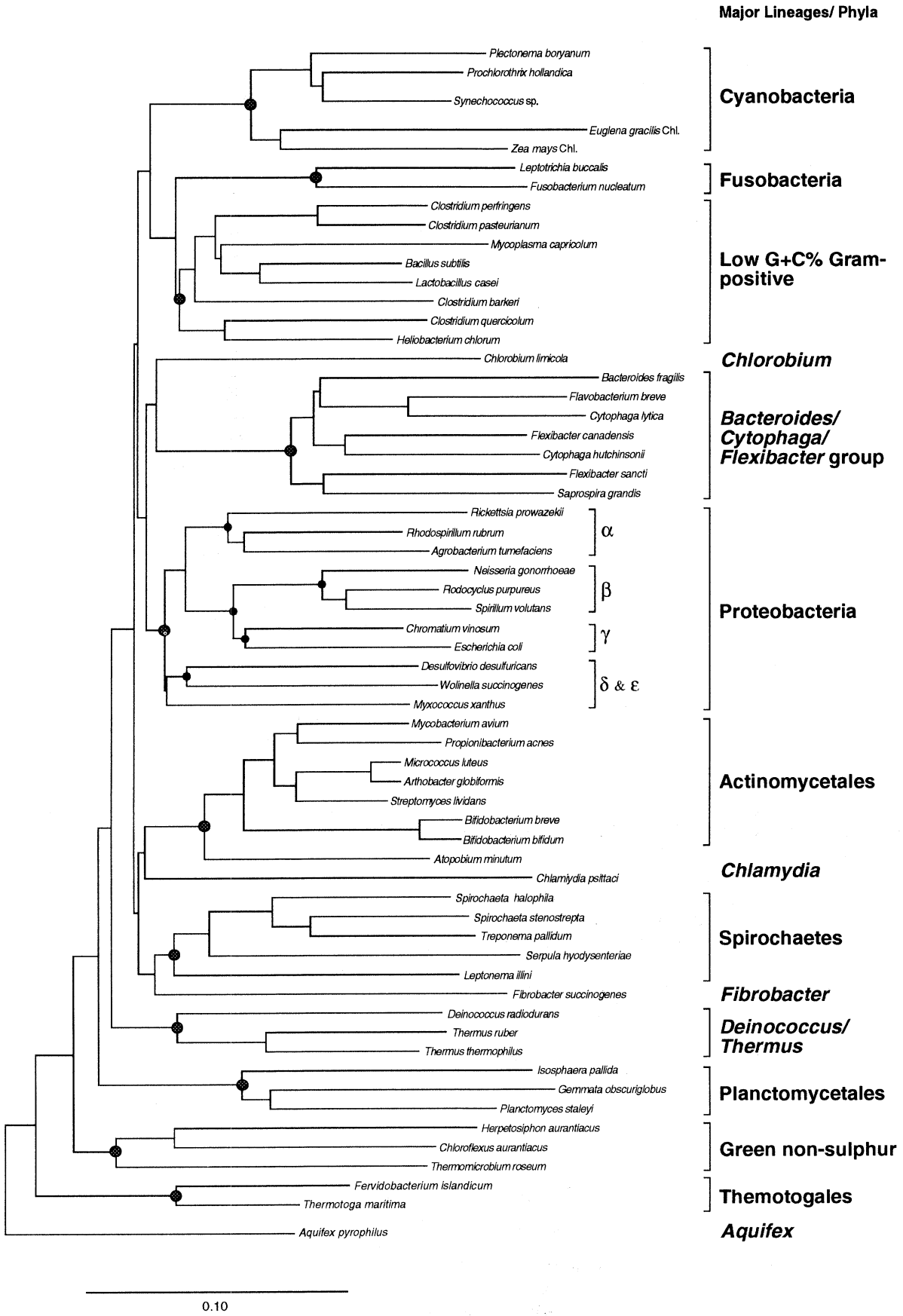

Figure 3. Relationships among a sample of Bacteria based upon SSU rRNA sequences and showing representatives of the major clades and phyla so far recognized. The tree is based upon 1179 bases of aligned sequence and was constructed using Neighbor Joining (Saitou & Nei 1987) and the Jukes & Cantor (1969) correction for multiple hits. Circles at nodes indicate clades which are stable under different methods of analysis. All sequences were obtained from the Ribosomal Database Project (Larsen *et al.* 1993). Scale bar indicates 0.10 estimated substitutions per sequence position.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

knowledge of the means whereby its success (i.e. apparently ubiquitous, abundant in some samples) has been achieved. As the sequences are from widely distributed samples, their biogeography is somewhat paradoxical. Should we expect widespread distributions of closely related microorganisms? or levels of endemism comparable to those of metaphytes and metazoa? It is too early to say much about this on present data and sampling, but in the marine environment where mixing can probably occur fairly easily some microbial lineages appear to be global. The SAR 11 lineage comprises a group of closely related sequences which produce a cluster of shallow branches (Giovannoni *et al.* 1990; Schmidt *et al.* 1991; DeLong 1992; Fuhrman *et al.* 1993). It has been suggested (Giovannoni *et al.* 1990) that these 'bushes' are not artefacts; when checked the individual 16*S* rRNA sequences show internal consistency for secondary structure. However, at the moment it is not clear if the separate lineages represent biological selection for new genotypes, individual clonal lineages, or some other phenomenon (Giovannoni *et al.* 1990; Maynard Smith *et al.* 1993).

The potential impact of environmental rRNA sequence analyses on the perception of the diversity of a single bacterial group, is nicely illustrated by the morphologically conspicuous order Planctomycetales. This clade currently contains four cultured genera: *Isosphaera*, *Gemmata*, *Pirellula* and *Planctomyces*. Planctomycetes have some unusual features for bacteria; at least one member contains a membrane bound nucleoid (Fuerst & Webb 1991), their 5*S* rRNA shows several unique idiosyncrasies (Bomar *et al.* 1988) and their rRNA operon organization deviates from the 'bacterial norm' in that 16*S* and 23*S* genes are separate (Liesack & Stackebrandt 1989; Liesack *et al.* 1992). Planctomycetes have proved difficult to culture and there are few strains available from established culture collections. The genera *Isosphera* and *Gemmata* are each represented by a single culturable strain. In contrast, environmental rRNA sequences and microscopic evidence indicate that planctomycetes are widespread and common in water and associated habitats. New sequences have been recovered from soil (Liesack & Stackebrandt 1992; Stackebrandt *et al.* 1993), compost (E. Stackebrandt, personal communication) and marine snow (Delong *et al.* 1993). In fact there are now more published 16*S* rRNA sequences from uncultured planctomycetes, than there are from cultured strains.

6. DOMAIN EUCARYA

A quick glance at a tree based upon eukaryote nuclear rRNA sequences (figure 4) reveals that plants, animals and fungi are only some of the most recently diverged ('crown') groups in an already diversified tree (Woese 1987; Sogin 1991). All of the deepest branching lineages, as well as some of the most recent, are protist. Most of the sequencing effort among eukaryotes has concentrated on the taxa of the crown. Coverage at the base of the tree is limited. There are approximately 11 sequences spread among the three basal lineages: the diplomonads, microsporidia and trichomonads. To put this into some kind of perspective, there are over 1500 described species for these three groups alone and a tentative and conservative estimate of described protist species exceeds 40 000 (Vickerman 1992). Even among crown groups such as the ciliates, which have perhaps received more attention than others, there are only 26 SSU rRNA sequences (including 14 for the single genus *Tetrahymena*) for over 8000 described species (Vickerman 1992). Heterotrophic flagellates, which play crucial roles as consumers in terrestrial and aquatic, aerobic and anaerobic environments (Fenchel 1988), are almost entirely neglected.

The earliest branches in the eukaryotic domain are anaerobic, sometimes aerotolerant, protists which lack mitochondria and peroxisomes. It is presumed that they diversified prior to the endosymbiotic events which gave rise to these organelles (Gray 1992). There is a small amount of published evidence that suggests that some of the basal eukaryote lineages retain prokaryote-sized 70*S* ribosomes and all of them sequenced so far contain small (typically less than 1580 bases) small subunit RNAs. Most of the early branches comprise parasites, but free-living relatives, such as *Hexamita*, also branch deep in the tree (Leipe *et al.* 1993; van Keulen *et al.* 1993). Anaerobic habitats contain many bactivorous flagellates and amoebae; anaerobiosis among eukaryotes is not rare or exotic, but these communities are largely unexplored phylogenetically. The anaerobic eukaryotes which have been sequenced fall into two groups: primitively anaerobic taxa, such as those already discussed, and taxa which are secondarily adapted to life without oxygen. Examples of the latter include chytrids (Marvin-Sikema *et al.* 1992) and ciliates (Fenchel & Finlay 1991). In ciliates, aerobes and anaerobes are intermingled (Embley & Finlay 1994) suggesting that in this group at least, the transition may not be too arduous.

Resolution of relationships at the base of the eukaryote tree is uncertain; the relative branching order between diplomonads, microsporidians and trichomonads changes with different analyses and choice of outgroup (Sogin *et al.* 1989; Leipe *et al.* 1993; van Keulen *et al.* 1993). The limited number of sequences available is part of the problem but some of the rRNA sequences also exhibit such strongly opposing base compositional biases that they inevitably confound tree construction. All of the published microsporidian sequences are AT rich, whereas some diplomonads, e.g. *Giardia lamblia*, are GC rich (Sogin *et al.* 1989). The most frequently used outgroup taxa are thermophilic Archaea which contain rRNAs with very high amounts of G and C. It is somewhat reassuring that recently sequenced diplomonads, such as *Giardia ardeae* and *Hexamita inflata*, have more balanced base compositions and also branch near the base of the tree (Leipe *et al.* 1993; van Keulen *et al.* 1993). Significantly, analysis of elongation factor EF-α of *Giardia lamblia* supports a deep divergence for this organism (Hashimoto *et al.* 1994).

In between the basal eukaryotes and the crown

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

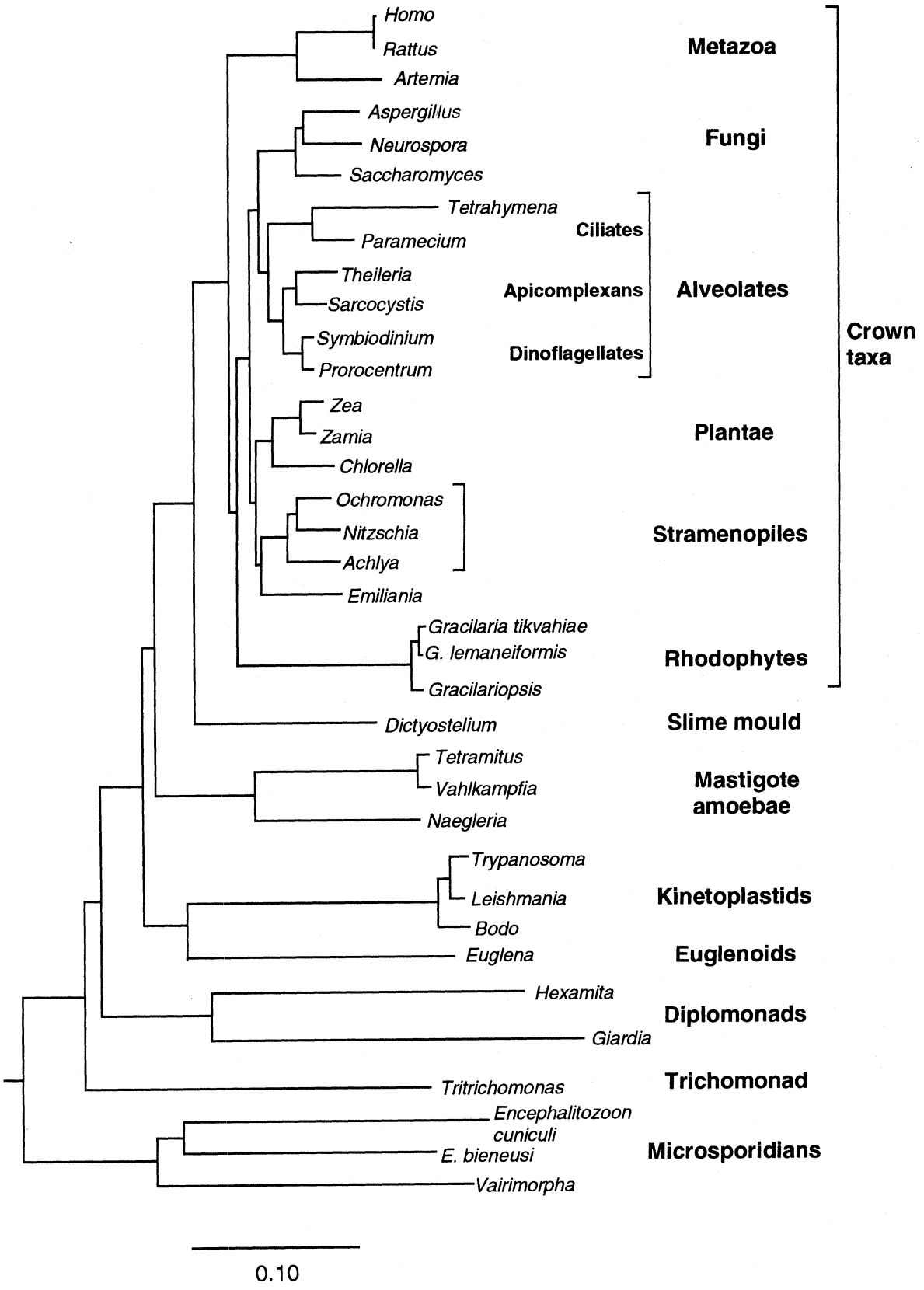

Figure 4. Relationships among a representative sample of Eucarya based upon SSU rRNA sequence comparisons. The tree is based upon 1014 positions and the tree was constructed using Neighbor Joining (Saitou & Nei 1987) and the Jukes & Cantor (1969) correction. The branching order of some lineages near the base of the tree and between clades at the crown of the tree is uncertain, as discussed in the text. All sequences were obtained from the Ribosomal Database Project (Larsen *et al.* 1993). Scale bar indicates 0.10 estimated substitutions per sequence position.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

groups are a number of poorly sampled lineages some of which appear to form distinct clades (figure 4). These include, among others, *Euglena* and kinetoplastids, *Naegleria*, *Valkampfia* and *Paratetramitus*, cellular slime moulds, and *Entamoeba histolytica* (Sogin 1991). The relative order of branching among the taxa which comprise the crown of the tree is difficult to assess reliably, because small amounts of sequence differences separate nodes. Robust groups, i.e. those that have resisted change through sampling, include ciliates, dinoflagellates and apicomplexans, which together form the monophyletic group Alveolata (Cavalier-Smith 1993; Patterson & Sogin 1993). The kingdoms Animalia, Plantae (including green algae) and Fungi (excluding the phycomycetes) are robust as well. Based on SSU rRNA sequence comparisons, Animalia and the choanoflagellates appear to form a clade with true Fungi as the sister taxon (Wainright *et al.* 1993). The relationships between other crown taxa are less well established and may change as new sequences become available. Particularly interesting is the question of how many monophyletic groups comprise photosynthetic eukaryotes. Currently the green algae (Chlorophytes) and all other green plants comprise one group, Rhodophytes another, and brown/golden algae (with non-photosynthetic phycomycetes) a third the Stramenopiles, with the composition of the latter still controversial. The Rhodophyta, in some published rRNA trees, appear at the base of the crown (e.g. Sogin 1991) but evidence is emerging to suggest that they may form a more recent divergence with the green algae and plants (M. Ragan, personal communication). A separate, but related issue, is the interpretation of endosymbiotic events that may have given rise to the photosynthetic phenotype in all the host lineages. This is complicated by so far unknown patterns of plastid gain and loss which, through intermingling of lineages, may become more apparent as a fuller tree develops. Evidence pertinent to this question may arise from unlikely quarters. For instance, it was recently shown that the non-photosynthetic apicomplexan *Plasmodium* contains the remnants of a plastid genome (Howe 1992; Gardner *et al.* 1993).

The kingdom Protista is paraphyletic but as yet there is no consensus for its replacement with a new higher level classification. This is partly due to a perception that many groups remain unsampled or poorly sampled and that as a result the tree topology is incomplete or unstable. The experience of microbiologists studying environmental rRNA sequences from Archaea and Bacteria lends some credence to this perception. Nevertheless, some monophyletic groups appear to be robust to change and one can argue that these should be named to indicate and communicate shared ancestry. The ciliates, dinoflagellates and apicomplexans are among the best candidates for naming (Cavalier-Smith 1993; Patterson & Sogin 1993). The question of whether only strictly monophyletic groups should form the basis of higher protist taxa has been questioned by Cavalier-Smith (1993). He proposed a six kingdom classification for an empire Eukaryota which comprises a paraphyletic superkingdom Archezoa (essentially the basal taxa discussed previously) and a paraphyletic superkingdom Metakaryota containing kingdoms Protozoa, Animalia, Fungi, Plantae (green algae, plants and red algae) and Chromista (mainly the remaining photosynthetic groups). Cavalier-Smith (1993) based his classification partly on a tree from rRNA sequences, in which only the Fungi are actually monophyletic. Others have advocated that only monophyletic taxa should be used as the basis for named groups (e.g. Patterson & Sogin 1993). It appears to us, while considering the issues surrounding a 'universal' diversity measure, that rather than name groups which are clearly paraphyletic (and therefore almost certain to change composition in the future), the challenge is to establish monophyletic groups and the synapomorphies by which we can recognize them. Recognition in this sense appeals to both molecular and morphological features.

7. MICROBIAL DIVERSITY: A CONCLUSION

Charting the diversity within the Metazoa and Metaphyta has had almost a century and a half of advance over the 'invisible' natural microbial world. Some may argue that science is either technology driven or concept driven. In truth, it is a symbiotic relationship between both. Although technology has allowed access to the natural world, old concepts: monophyly or paraphyly, phylogeny, systematics and the like, are necessary to interpret the results and to understand them. The first window into the hidden world of microbial diversity came with the advent of microscopes, developing from simple optics to sophisticated electron emission beam models. It is a nice irony that one of the first microbes to be described, *Giardia intestinalis* in 1681 from Leeuwenhoek's diarrhetic stools (Dobell 1920), is now recognized as among the earliest eukaryote offshoots. The development of molecular systematics and new tools such as the polymerase chain reaction (PCR), has opened access to microbial diversity in a manner which was previously unimagined. To quote May (1990, p. 301) '... studies of natural populations of microorganisms ... are, in their own way even more astonishing than Erwin's and others' revelations about tropical canopy faunas'. The tools for estimating microbial diversity in terms of abundance are only just beginning to emerge. Any figures which are currently quoted are probably best considered as representing efforts to estimate numbers, rather than realistic estimates of numbers. How unrealistic (we think) some of these figures are, can be appreciated from the Systematics Agenda 2000 report (Systematics Agenda 2000 1994). In this document arthropod diversity alone is estimated as in the region of 10–100 million species whereas 'species' diversity of all the microbial world is only about 10% of that figure. Yet all arthropods harbour microorganisms, and experience suggests that in each case at least some of these will be previously unrecorded. One factor which helps to obscure our understanding of diversity is the seemingly intractable problem of a universal species concept. Thus far, the intractability

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

may have something to do with the focus of the question: are concepts originally proposed for metaphytes and metazoans applicable to all of life ? We would suggest that they are obviously not, but no-one has yet suggested an acceptable alternative.

An appreciation and understanding of the natural world depends not only on the description of the biodiversity of plants and animals, but on knowledge of the diversity of all Domains of life. Darwin closed his *Origin of Species* with the following words: '... from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved'. In the last quarter of the 20th century we are only starting to appreciate just how far from simple was the beginning.

The authors would like to acknowledge the generosity of Carl Woese, Mitchel L. Sogin and their colleagues for making freely available their compilations of aligned small sub-unit rRNA sequences through the Ribosomal Database Project. The authors wish to thank Dr Darrell Siebert and Dr Mark Wilkinson for useful discussions and comments on the manuscript. The work of R.P.H. is supported by Natural Environment Research Council grant GR3/8146 to T.M.E.

REFERENCES

Amann, R., Ludwig, W. & Schleifer, K.-H. 1988 β-subunit of ATP-synthase: a useful marker for studying the phylogenetic relationships of eubacteria. *J. gen. Microbiol.* **134**, 2815–2821.

Balch, W.E., Magrum, L.J., Fox, G.E., Wolfe, R.S. & Woese, C.R. 1977 An ancient divergence among the bacteria. *J. molec. Evol.* **9**, 305–311.

Barns, S.M., Fundyga, R.E., Jeffries, W. & Pace, N.R. 1994 Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc. natn. Acad. Sci. U.S.A.* **91**, 1609–1613.

Bomar, D., Giovannoni, S.G. & Stackebrandt, E. 1988 A unique eubacterial 5*S* rRNA in members of the order *Planctomycetales*. *J. molec. Evol.* **27**, 121–125.

Britschgi, T. & Giovannoni, S.J. 1991 Phylogenetic analysis of a natural marine bacterioplankton population by rRNA gene cloning and sequencing. *Appl. Environ. Microbiol.* **57**, 1707–1713.

Burggraf, S., Olsen, G.J., Stetter, K.O. & Woese, C.R. 1992 A phylogenetic analysis of *Aquifex pyrophilus*. *Syst. appl. Microbiol.* **15**, 352–356.

Butterfield, N.J., Knoll, A.H., & Swett, K. 1990 A bangiophyte red alga from the Proterozoic of Arctic Canada. *Science, Wash.* **250**, 104–107.

Cammarano, P., Palm, P., Creti, R., Ceccareli, E., Sanangelantoni, A.M. & Tiboni, O. 1992 Early evolutionary relationships among known life forms inferred from elongation factor Ef-2 (Ef-G) sequences. Phylogenetic coherence and structure of the archaeal domain. *J. molec. Evol.* **34**, 396–405.

Cavalier-Smith, T. 1993 Kingdom protozoa and its 18 phyla. *Microbiol. Rev.* **57**, 953–994.

Creti, R., Citarella, R., Citarella, F., Tiboni, O., Sanangelantoni, A., Palm, P. & Cammaranno, P. 1991 Nucleotide sequence of a DNA region comprising the gene for elongation factor 1 α (EF-1 α) from the ultrathermophilic archeaeota *Pyrococcus woesei*: phylogenetic implications. *J. molec. Evol.* **33**, 332–342.

DeLong, E.F. 1992 Archaea in coastal marine environments. *Proc. natn. Acad. Sci. U.S.A.* **89**, 5685–5689.

DeLong, E.F., Franks, D.G. & Alldredge, A.L. 1993 Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. *Limnol. Oceanogr.* **38**, 924–934.

De Soete, G. 1983 A least squares algorithm for fitting additive trees to proximity data. *Psychometrica* **48**, 621–626.

Dobell, C. 1920 The discovery of intestinal protozoa in man. *Proc. R. Soc. Med.* **13**, 1–15.

Embley, T.M. & Finlay, B.J. 1994 The use of rRNA sequences to unravel the relationships between anaerobic ciliates and their methanogen endosymbionts. *Microbiology* **140**, 225–235.

Embley, T.M. & Stackebrandt E. 1994 The use of ribosomal RNA sequences in microbial ecology. In *Molecular approaches in environmental microbiology* (ed. R. W. Pickup, J. Saunders & G. A. Codd). London: Ellis Horwood. (In the press.)

Embley, T.M., Thomas, R.H. & Williams, R.A.D. 1993 Reduced thermophilic bias in the 16*S* rRNA sequence from *Thermus ruber* provides further support for a relationship between *Thermus* and *Deinococcus*. *Syst. appl. Microbiol.* **16**, 25–29.

Felsenstein, J. 1978 Cases in which parsimony or compatibility methods will be positively misleading. *System. Zool.* **27**, 401–410.

Fenchel, T. 1988 Marine plankton food chains. *A. Rev. Ecol. System.* **19**, 19–38.

Fenchel, T. 1992 What can ecologists learn from microbes: life beneath a square centimetre of sediment surface. *Funct. Ecol.* **6**, 499–507.

Fenchel, T. & Finlay, B.J. 1991 The biology of free-living anaerobic ciliates. *Eur. J. Protistol.* **26**, 201–215.

Forterre, P., Benachenhou-Lahfa, Confalonieri, F., Duguet, M., Elie, C. & Labedan, B. 1993 The nature of the last universal ancestor and the root of the tree of life, still open questions. *BioSystems* **28**, 15–23.

Fuerst, J.A. & Webb, R.I. 1991 Membrane bounded nucleoid in the eubacterium *Gemmata obscuriglobus*. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8184–8188.

Fuhrman, J.A., McCallum, K. & Davis, A.A. 1992 Novel major archaebacterial group from marine plankton. *Nature, Lond.* **356**, 148–149.

Fuhrman, J.A., McCallum, K. & Davis, A.A. 1993 Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific oceans. *Appl. Environ. Microbiol.* **59**, 1294–1302.

Gambacorta, A., Trincone, A., Nicolaus, B., Lama, L. & De Rosa, M. 1994 Unique features of lipids of *Archaea*. *System. appl. Microbiol.* **16**, 518–527.

Gardner, M.J., Feagin, J.E., Moore, D.J., Rangachari, K., Williamson, D.H. & Wilson, R.J.M. 1993 Sequence and organisation of large subunit rRNA genes from the extrachromosomal 35 kb circular DNA of the malaria parasite *Plasmodium falciparum*. *Nucl. Acid. Res.* **21**, 1067–1071.

Giovannoni, S.J., Britschgi, T.B., Moyer, C.L. & Field, K.G. 1990 Genetic diversity in Sargasso Sea bacterioplankton. *Nature, Lond.* **345**, 60–63.

Gogarten, J.P., Kibak, H., Dittrich, P., Taiz, I., Bowman, E.J., Bowman, B.J., Manolson, M.F., Pool, R.J., Date, T., Oshima, T., Konishi, J., Denda, K. & Yoshida, M. 1989 Evolution of the vacuolar H^+ ATPase: implications for the origins of the eukaryotes. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6661–6664.

Gogarten, J.P., Starke, T., Kibak, H., Fichman, J. & Taiz, L. 1992 Evolution and isoforms of V-ATPase subunits. *J. exp. Biol.* **172**, 137–147.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Gould, S.J. 1989 *Wonderful life*. London: Penguin Books.
Gray, M.W. 1992 The endosymbiont hypothesis revisited. *Int. Rev. Cytol.* **141**, 233–357.
Hashimoto, T., Nakamura., Y., Nakamura., F., Shirakura, T., Adachi, J., Goto, N., Okamoto, K.-I. & Hasegawa, M. 1994 Protein phylogeny gives robust estimation for early divergences of a mitochondria-lacking protozoa, *Giardia lamblia*. *Molec. biol. Evol.* **11**, 65–71.
Hilario, E. & Gogarten, P. 1993 Horizontal transfer of ATPase genes-the tree of life becomes a net of life. *Biosystems* **31**, 111–119.
Howe, C.J. 1992 Plastid origin of an extrachromosomal DNA molecule from *Plasmodium*, the causative agent of malaria. *J. theor. Biol.* **158**, 199–205.
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. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9355–9359.
Jukes T.H. & Cantor C.R. 1969 Evolution of protein molecules. In *Mammalian protein metabolism*. (ed. H. N. Munro), pp. 21–132. New York: Academic Press.
Kandler, O. 1994 Cell wall biochemistry and three-domain concept of life. *System. appl. Bacteriol.* **16**, 501–509.
Kimbel, W.H., Johanson, D.C. & Rak, Y. 1994 The first skull and other new discoveries of *Austrlalopithecus afarensis* at Hadar, Ethiopia. *Nature, Lond.* **368**, 449–451.
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. *Nucl. Acid. Res.* **21**, 3021–3023.
Lake J.A. 1987*a* Prokaryotes and archaebacteria are not monophyletic: rate invariant analysis of rRNA genes indicates that eukaryotes and eocytes form a monophyletic taxon. *Cold Spring Harb. Symp. quant. Biol.* **52**, 839–846.
Lake J.A. 1987*b* A rate independent technique for analysis of nucleic acid sequences: evolutionary parsimony. *Molec. biol. Evol.* **4**, 167–191.
Lake J.A. 1988 Origin of the eukaryotic nucleus determined by rate invariant analysis of rRNA sequences. *Nature, Lond.* **331**, 184–186.
Lake J.A. 1989 Origin of eukaryotic nucleus: eukaryotes and eocytes are genotypically related. *Can. J. Microbiol.* **35**, 109–118.
Lake J.A. 1991 The order of sequence alignment can bias the selection of tree topology. *Molec. biol. Evol.* **8**, 378–385.
Leipe, D.L., Gunderson, J.H., Nerad, T.A. & Sogin M.L. 1993 Small subunit ribosomal RNA of *Hexamita inflata* and the quest for the first branch in the eukaryotic tree. *Molec. biochem. Parasitol.* **59**, 41–48.
Liesack, W. & Stackebrandt, E. 1989 Evidence for unlinked rrn operons in the planctomycete *Pirellula marina*. *J. Bacteriol.* **171**, 5025–5030.
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. *J. Bacteriol.* **174**, 5072–5078.
Liesack, W. , Söller, R., Stewart, T., Haas, H., Giovannoni, S. & Stackebrandt, E. 1992 The influence of tachytelically (rapidly) evolving sequences on the topology of phylogenetic trees – intrafamily relationships and the phylogenetic position of *Planctomycetaceae* as revealed by comparative analysis of 16*S* ribosomal RNA sequences. *Syst. Appl. Microbiol.* **15**, 357–362.
Marvin-Sikema, F.D., Lahpor, G.A., Kraak, M.N., Gottschal, J.C. & Prins, R.A. 1992 Characterisation of an anaerobic fungus from Llama faeces. *J. gen. Microbiol.* **138**, 2235–2241.
May, R.M. 1990 How many species? *Phil. Trans. R. Soc. Lond.* B **330**, 171–182.
Maynard Smith, J., Smith, N.H., O'Rourke, M. & Spratt, B. 1993 How clonal are bacteria? *Proc. natn. Acad. Sci. U.S.A.* **90**, 4384–4388.
Neefs, J.-M., Van de Peer, Y., De Rijk, P., Chapelle, S. & De Wachter, R. 1993 Compilation of small ribosomal subunit structures. *Nucl. Acid. Res.* **21**, 3025–3049.
Olsen, G.J. & Woese C.R. 1989 A brief note concerning archaebacteria phylogeny. *Can. J. Microbiol.* **35**, 119–123.
Olsen, G.J. & Woese C.R. 1993 Ribosomal RNA: a key to phylogeny. *FASEB J.* **7**, 113–123.
Olsen, G.J., Woese, C.R. & Overbeek, R. 1994 The winds of evolutionary change: breathing new life into microbiology. *J. Bacteriol.* **176**, 1–6.
Pace, N.R., Stahl, D.A., Lane, D.J. & Olsen, G.J. 1985 Analyzing natural microbial populations by rRNA sequences. *Am. Soc. Microbiol. News* **51**, 4–12.
Patterson, D.J. & Sogin, M.L. 1993 Eukaryote origins and protistan diversity. In *The origin and evolution of the cell* (ed. H. Hartman & K. Matsuno), pp. 13–46. Singapore: World Scientific Publishing.
Rivera, M.C. & Lake, J.A. 1992 Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science, Wash.* **257**, 74–76.
Runnegar, B. 1992 Evolution of the earliest animals. In *Major events of the history of life* (ed. J. W. Shopf), pp. 65–95. Boston: Johns and Bartlett.
Saitou, N. & Nei, M. 1987 The neighbor joining method: a new method for constructing phylogenetic trees. *Molec. biol. Evol.* **4**, 406–425.
Schopf, J.W. 1992 The oldest fossils and what they mean. In *Major events of the history of life* (ed. J. W. Shopf), pp. 29–64. Boston: Johns and Bartlett.
Schmidt, T.M., DeLong, E.F. & Pace, N.R. 1991 Analysis of a marine picoplankton community by 16*S* rRNA gene cloning and sequencing. *J. Bacteriol.* **173**, 4371–4378.
Schwartz, R.M. & Dayhoff, M.D. 1978 Origin of prokaryotes, mitochondria and chloroplasts. *Science, Wash.* **199**, 395–403.
Sogin, M. 1991 Early evolution and the origin of eukaryotes. *Curr. Opin. Genet. Dev.* **1**, 457–463.
Sogin, M.L., Gunderson, J.H. & Elwood, H.J., Alonso, R.A. & Peattie, D.A. 1989 Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from Giardia lamblia. *Science, Wash.* **243**, 75–77.
Stackebrandt, E., Liesack, W. & Goebel, B.M. 1993 Bacterial diversity in a soil sample from a subtropical Australian environment as determined by 16*S* rDNA analysis. *FASEB J.* **7**, 232–236.
Sumi, M., Sato, M.H., Denda, K., Date, T. & Yoshida, M. 1992 A DNA fragment homologous to F1-ATPase *beta* subunit was amplified from genomic DNA of *Methanosarcina barkeri*. Indication of an archeabacterial F-type ATPase. *FEBS Lett.* **314**, 207–210.
Systematics Agenda 2000 1994 Charting the biosphere. Technical report. Produced by the Systematics Agenda 2000.
Tsutsumi, S., Denda, K., Yokoyama, K., Oshima, T., Date, T. & Yoshida, M. 1991 Molecular cloning of genes encoding major two subunits of a eubacterial V-Type ATPase from *Thermus thermophilus*. *Biochim. biophys. Acta* **1098**, 13–20.
van Keulen, H., Gutell, R.R., Gates, M.A., Campbell, S.R., Erlandsen, S.L., Jarroll, E.L., Kulda, J. & Meyer, E.A. 1993 Unique phylogenetic position of diplomonadida based on complete small subunit ribosomal RNA sequence of *Giardia ardeae*, *G. muris*, *G. duodenalis* and *Hexamita* sp. *FASEB J.* **7**, 223–231.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Vickerman, K. 1992 The diversity and ecological significance of protozoa. *Biodiv. Conserv.* **1**, 334–341.

Wainright, P.O., Hinkle, G., Sogin, M.L. & Stickel, S.K. 1993 Monophyletic origins of the metazoa: an evolutionary link with fungi. *Science, Wash.* **260**, 340–342.

Wagner, M., Amann, R.I., Lemmer, H. & Schleifer, K.-H. 1993 Probing activated sludge with oligonucleotides specific for proteobacteria: inadequacy of culture dependent methods for describing microbial community structure. *Appl. Environ. Microbiol.* **59**, 1520–1525.

Ward, D.M., Weller, R. & Bateson, M.M. 1990 16*S* rRNA sequences reveal numerous uncultured inhabitants in a natural community. *Nature, Lond.* **345**, 63–65.

Weller, R., Weller, J.W. & Ward, D.M. 1991 16*S* rRNA sequences of uncultivated hot spring cyanobacterial mat inhabitants retrieved as randomly primed cDNA. *Appl. Environ. Microbiol.* **57**, 1146–1151.

Weller, R., Bateson, M.M., Heimbuch, B.K., Kopczynski, E.D. & Ward, D.M. 1992 Uncultivated cyanobacteria, *Chloroflexus*-like inhabitants, and spirochaete-like inhabitants of a hot spring cyanobacterial mat. *Appl. Environ. Microbiol.* **58**, 3964–3969.

Weisburg, W.G., Giovannoni, S.J. & Woese, C.R. 1989 The *Deinococcus-Thermus* phylum and the effect of rRNA composition on phylogenetic tree construction. *System. appl. Microbiol.* **11**, 128–134.

Winker, S. & Woese, C.R. 1991 A definition of the domains Archaea, Bacteria and Eucarya in terms of small subunit ribosomal RNA characteristics. *System. appl. Microbiol.* **14**, 305–310.

Whittaker, R.H. 1959 On the broad classification of organisms. *Q. Rev. Biol.* **34**, 210–226.

Woese, C.R. 1987 Bacterial evolution. *Microbiol. Rev.* **51**, 221–271.

Woese, C.R. 1992 Prokaryote systematics: the evolution of a science. In *The prokaryotes* (ed. A. Balows, H. G. Truper, M. Dworkin, W. Harder and K.-H. Schleifer), pp. 3–18. New York: Springer-Verlag.

Woese, C.R., Magrum, L.J. & Fox, G.E. 1978 Archaebacteria. *J. molec. Evol.* **11**, 245–252.

Woese, C.R., Kandler, O. & Wheelis, M.L. 1990 Towards a natural system for organisms : proposal for the domains Archaea, Bacteria and Eucarya. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4576–4579.

Yokoyama, K., Oshima, T. & Yoshida, M. 1990 *Thermus thermophilus* membrane associated ATPase: indication of a eubacterial V-type ATPase. *J. biol. Chem.* **265**, 21946–21950.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES