THE PHYLOGENY OF THERMOPHILES AND HYPERTHERMOPHILES AND THE THREE DOMAINS OF LIFE

The Phylogeny of Thermophiles

DEBASHISH BHATTACHARYA
University of Iowa
Department of Biological Sciences
Biology Building, Iowa City, Iowa 52242-1324
United States

THOMAS FRIEDL
Department of Biology, General Botany
University of Kaiserslautern
P.O. Box 3049, D-67653 Kaiserslautern, Germany

HEIKO SCHMIDT
Deutsches Krebsforschungszentrum
Theoretische Bioinformatik
Im Neuenheimer Feld 280 , D-69120 Heidelberg, Germany

1. Introduction

The nature of the first cells and the environment in which they lived are two of the most interesting problems in evolutionary biology. All living things are descendents of these primordial cells and are divided into three fundamental lineages or domains, Archaea (formerly known as Archaebacteria), Bacteria (formerly known as Eubacteria), and the Eucarya (formerly known as Eukaryotes, Woese et al. 1990). The Archaea and Bacteria are prokaryotic domains whereas the Eucarya includes all other living things that have a nucleus (i.e., the genetic material is separated from the cytoplasm by a nuclear envelope). The observation of the three primary domains, first made on the basis of small subunit (i.e., 16S, 18S) ribosomal DNA (rDNA) sequence comparisons (Woese 1987), has created a framework with which the nature of the last common ancestor (LCA) can be addressed. In this review we present phylogenies of the prokaryotic domains to understand the origin and distribution of the thermophiles (organisms able to grow in temperatures > 45°C) and the hyperthermophiles (organisms able to grow in temperatures > 80°C). Hyperthermophiles are limited to the Archaea and Bacteria. In addition, we inspect the distribution of extremophiles within the cyanobacteria. The cyanobacteria are unique in being able to tolerate rapidly fluctuating environmental conditions. This capacity has presumably allowed some cyanobacteria to survive in niches unexplored by other prokaryotes and eukaryotes which have a more restricted environmental tolerance. We do not deal with eukaryotic extremophiles in this

chapter because no multicellular plant or animal yet found can survive in temperatures of about 50°C and no protists can tolerate long-term exposure to temperatures in excess of 60°C (Madigan and Marrs 1997).

The phylogenetic data regarding thermophiles and hyperthermophiles are used to gain insights into the evolutionary relationship of extant cells to the LCA and the environment in which the LCA may have lived. These attempts at "retelling" evolution are purely hypothetical exercises since there is no way to "glimpse" into the distant past and the fossil record offers no clear answers. The analysis of living taxa to infer the characteristics of long-dead ancestors is however a remarkably powerful way to recreate evolutionary history.

2. The Origin of the Three Domains of Life

Fossilized cyanobacterial mats (stromatolites) have been found in strata dating to 3.5 billion years before the present (BYP, Schopf 1993) suggesting that photosynthetic prokaryotes were already abundant on the Earth. This hypothesis is supported by the presence of iron oxide in banded formations that date from the same time (Holland and Beukes 1990). The oxidization is believed to have arisen from the biological production of oxygen via photosynthesis in the marine ecosystem. Since modern cyanobacteria presently form only one branch of the bacterial lineage it would follow that the stromatolites are the remnants of a well-evolved bacterial world (Schopf 1994). The common ancestor of these bacteria must therefore already have existed for millions of years. Alternatively, the cyanobacteria in the stromatolites may be a long-dead branch of photosynthetic life which has left no survivors among the modern-day bacteria. This would be one explanation for the protein sequence trees of Doolittle et al. (1996) who have suggested that the LCA existed only about 2 BYP. The Doolittle et al. hypothesis is however controversial for a number of reasons including the questionable use of protein sequence divergence rates to place dates on deep branches in the universal phylogeny (see Golding 1996, Martin 1996 for comments). In addition, the strong similarity between living cyanobacteria and the fossilized stromatolite cells, and the existence of living stromatolites with dominant cyanobacterial communities (e.g., Shark Bay, Australia; Schopf and Walter 1982, Whitton and Potts 1982) argue for an unbroken evolutionary line connecting fossilized cyanobacteria with those existing today. The current consensus opinion remains that life has probably existed for at least 3.5 billion years and that the earliest recognizable life-form resembles modern-day cyanobacteria. The eukaryotes, which first appear about a billion years later in the fossil record (Knoll 1994), apparently evolved from a prokaryotic ancestor though likely not from a highly evolved cyanobacterium but rather from a heterotrophic common ancestor, be it a Bacteria or an Archaea.

The recognition that rDNA coding regions are universally distributed, highly conserved, and encode significant evolutionary signal has led to their usage to create a molecular phylogeny of living things. This work, set in motion by Woese and his colleagues have, for example, shown that the prokaryotes are comprised of two distinct lineages (i.e., Archaea, Bacteria) that are as distinct from each other as either is from eukaryotes (Woese 1987). The evolutionary distinctness of the three domains has since

then been confirmed with comparisons of a number of protein sequences (e.g., ATPase subunits, elongation factors [Iwabe et al. 1989], isoleucyl-tRNA synthetases [Brown and Doolittle 1995]). All of these protein coding regions were presumably also found in the LCA and, due to their fundamental importance in the cell, have maintained their original function. In spite of the exemplary characteristics of rDNA genes, one issue was however left unresolved with these phylogenetic analyses, the position of the root of the universal tree. This problem arises because there are no outgroups of rDNA genes since all taxa contain the same genes and there is no unbiased way to decide which domain came first in evolution. The rooting problem has since been solved using protein coding regions that underwent duplications in the LCA prior to the radiation of the three domains. One set of gene duplication products can then be used as a "natural" outgroup of the other (e.g., EF-Tu/G, ATPase V/F [see Gupta and Golding 1996 for review]). With this approach, the root of the universal tree has been placed in the branch leading to the Bacteria, thereby supporting a monophyletic origin of the Archaea and Eucarya (see Iwabe et al.1989, Gogarten et al. 1989). This view has since then been widely accepted and was used as the basis for the formal recognition of the three domains of life (Woese et al. 1990) and for the rooting of subsequent rDNA trees (see Fig. 1A). Like the fossil record, the molecular systematic scheme then also supports a prokaryotic LCA. In many molecular trees there is additional evidence that the Archaea diverged from the LCA and subsequently gave rise to the eukaryotes.

The hypothesis that the Eucarya diverge from within the Archaea is however not universally accepted. The most convincing evidence against this scheme comes from comparisons of the 70 kDa heat schock proteins (hsp70/dnaK) that suggest a sister group relationship between eukaryotes and gram-negative bacteria. The hsp70, as well as the glutamine synthetase and the glutamate dehydrogenase, trees show clearly that gram-negative bacteria gave rise to prokaryotes whereas gram-positive bacteria are directly related to the Archaea. In global phylogenies using comparisons of a number of conserved protein sequences, Gupta and Golding (1996) have also shown that about one-half of the trees support the Eucarya/Gram-Negative clade whereas about the same number support a Eucarya/Archaea clade. None of the phylogenies are consistent with an Eucarya/Gram-Positive clade. Since two universal trees of life obviously cannot be correct, Gupta and Singh (1994) and Gupta and Golding (1996) have suggested a chimeric model for the origin of the eukaryotic cell. In this model, a gram-negative bacterium engulfs a "eocyte"-type (see Lake and Rivera 1994 for details) thermoacidophilic archaebacterium (see Fig. 1B). Fusion of the chromosomes of these taxa leads to the retention/loss of certain genes from each partner. The usage of the surviving genes in sequence analyses therefore ultimately lead to different positions of the eukaryotes (i.e., either with the gram-negative bacteria or with the Archaea) in the trees. The high similarity in both sequence and secondary structure of the proteins and protein complexes involved in DNA transcription (Langer et al. 1995, Rowlands et al. 1994, Kosa et al. 1997) found between Eucarya and Archaea could be explained if the DNA transcription machinery traced its ancestry to the archaeal partner in the original cell fusion. The fusion hypothesis would also explain why eukaryotes do not appear until relatively late in the fossil record. The eukaryotes would be recognizable as such only after the fusion event that had created them and the ensuing intracellular differentiation leading to the evolution of a nucleus and a endomembrane system (Golding 1996). In conclusion, the molecular data do not yet clarify the role of direct

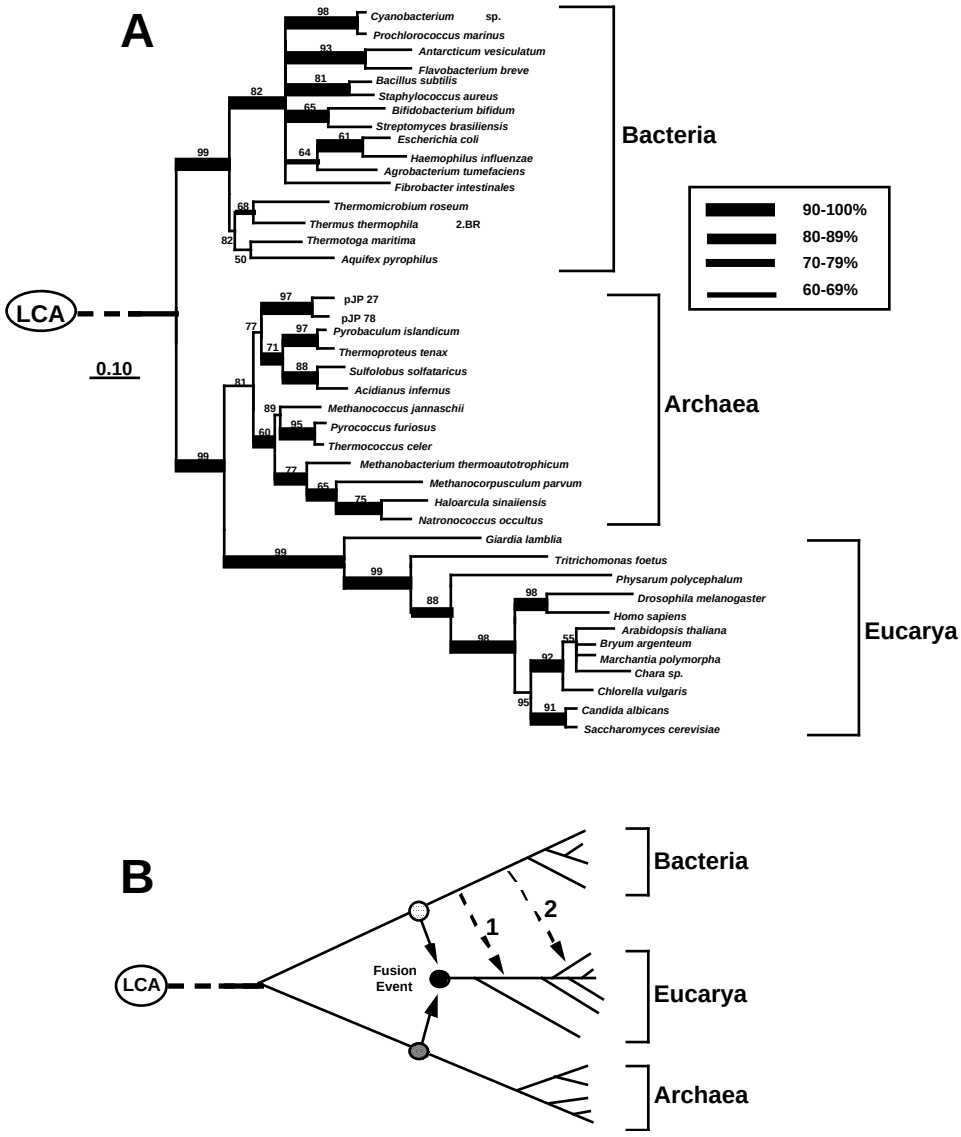

Fig. 1. A. Universal tree of life inferred from small subunit rDNA sequence comparisons. This tree was inferred using quartet puzzling and rooted within the branch leading to the Bacteria. Puzzle values are shown above the branches. Results of a distance bootstrap analysis (1000 replications) are shown as branch lengths of differing thicknesses (see legend). B. Origin of the eukaryotes via a fusion involving an Archaea and a gram-negative bacterial cell (based on Gupta and Golding 1996). The numbers 1 and 2 are the putative origins of the mitochondrion and plastid, respectively, in eukaryotes via endosymbiosis. This figure supports a highly chimeric origin of the eukaryotic cell. LCA refers to the last common ancestor of the three domains.

filiation from an archaeal ancestor and/or the fusion of archaeal and gram-negative (or some other) bacterial cells in the origin of the Eucarya. It seems relatively certain however that the first cells were prokaryotic in nature (i.e., lacked a nucleus). In the following sections, we present phylogenies of the Archaea and Bacteria, and discuss the kind of environment that may have existed when the prokaryotic LCA existed. These sections will concentrate on the evolution of thermophilic and hyperthermophilic prokaryotes.

3.0 Phylogeny of the Archaea

3.1 INTRODUCTION

Archaea seem to occupy almost all harsh habitats. They are found in extremely hot niches which until recently were considered to be sterile. Some hyperthermophiles grow between 60°C and 115°C and have an optimal growth temperature near the boiling point of water. Other Archaea are found in marine environments with temperatures around the freezing point. Other extreme environments like those strictly anaerobic, extremely halophilic (3-20% salt) sometimes related to the presence of heavy metals (e.g., at the bottom of the Dead Sea), acidophilic at pH<2 or alkaliphilic at pH>9, or "barophilic" (Pennisi 1997) at high pressures of 1100 atmospheres are also home to Archaea. Often two or more of those environmental factors occur together where Archaea live (Lowe et al. 1993, Stetter 1996).

The Archaea resemble the Bacteria in cellular structure and organisms in both these domains contain single circular genomes with very few introns. Archaeal genes are arranged in polycistronic operons and transcription and translation is coupled like in the Bacteria. The Archaea have however distinguishing characteristics such as cell walls made of protein, glycoprotein, carbohydrate, or pseudomurein, but not peptidoglycan, and isopranyl glycerol ether lipids rather than fatty acid glycerol ester lipids. The ether-linked 40-carbon long lipids form lipid monolayers rather than bilayers (Driessen et al. 1996). The Archaea also share a number of biochemical and genetic features with the Eucarya which are not found in Bacteria (Keeling and Doolitle 1995, Zillig 1991). The best studied of these characters is DNA transcription, the components of which share high sequence and overall tertiary structure similarity among the two domains (see Rowlands et al. 1994, Langer et al. 1995, Kosa et al. 1997). In addition, although archaeal ribosomes are of the 70S bacterial type, translation is initiated by an unformylated methionine like in the Eucarya. So, Archaea share characteristics with both Eucarya and Bacteria which together seem to indicate an intermediate evolutionary position in the universal tree (for reviews, see Zillig 1991, Woese 1987, Lowe et al. 1993, Jones et al. 1987.)

3.2 PHYLOGENETIC METHODS

We used aligned small subunit rRNA sequences from the rRNA WWW server in Antwerp (Van de Peer et al. 1997) for the phylogenetic analyses of the Archaea. As representatives of the euryarchaeotes we included sequences from the extreme halophiles (Halobacteriales, *Haloferax denitrificans*, *Haloarcula sinaiiensis*, *Natronococcus occultus*), the Methanomicrobiales (*Methanosaeta thermoacetophila*, *Methanolobus vulcani*, *Methanocorpusculum parvum*), the Methanobacteriales (*Methanothermus fervidus*, *Methanobacterium bryantii*, *Methanobacterium thermoautotrophicum*), Methanococcales (*Methanococcus jannaschii*, *Methanococcus igneus*, *Methanococcus*

voltae), and the Thermococcales (*Pyrococcus furiosus*, *Thermococcus celer*, *Thermococcus litoralis*). As representatives of the crenarchaeotes we used *Acidianus infernus*, *Metallosphaera sedula* 2, *Stygiolobus azoricus*, *Sulfolobus metallicus*, *Sulfolobus solfataricus* (Sulfolobales), *Desulfurococcus mobilis* (Desulforococcales), *Pyrobaculum islandicum*, *Thermofilum pendens*, *Thermoproteus tenax*) (Thermoproteales), and *Pyrodictium occultum* (Pyrodictiales). The Korarchaeota sequences, pJP78 and pJP27, were manually aligned with the existing Antwerp rRNA sequence alignment. The rRNA sequences of the primitive, mitochondrial-lacking Eucarya, *Tritrichomonas foetus* and *Giardia lamblia*, were used to outgroup root the Archaea phylogeny.

Trees were constructed with the pfastDNAml 2.1.4 (a new parallel version of fastDNAml published by Olsen et al. in 1994) for maximum likelihood analyses (ML), puzzle 3.1 for ML analyses using quartet puzzling (Strimmer and von Haeseler 1996), and CLUSTAL W 1.7 (Thompson et al. 1994) for the neighbor-joining analyses (NJ). Bootstrap analyses were done using 100 pseudosamples with the pfastDNAml method and using 1000 pseudosamples with the neighbor-joining method. Gaps within the alignment columns were excluded from all analyses.

3.3 PHYLOGENY

Although Archaea are a hot topic in research, especially industrial research focused on exploiting "super" extremophile enzymes (Madigan and Marrs 1997, Pennisi 1997), we are far from understanding the breadth of archaeal taxonomy. rDNA sequence analyses suggest that the Archaea is divided into two major kingdoms, the Euryarchaeota and the Crenarchaeota (see Fig. 2). The Euryarchaeota include several methanogenic, halophilic, and some sulfur-metabolizing hyperthermophiles. Since all known Crenarchaeota are sulfur-dependent hyperthermophiles, the members of this kingdom are thought to not be widespread but restricted to hyperthermophilic habitats (Woese et al. 1990). There have been hyperthermophilic Archaea also found in cold water conditions, suggesting that hyperthermophilic organisms might be able to exist in the cold, although they cannot grow under such conditions (Stetter 1996).

In the past, Archaea from environmental samples were cultured prior to PCR and sequence analysis of the rDNA coding regions from single isolates. It has been shown however that about 99% of the organisms from an environmental sample are lost during culturing (Fuhrman et al. 1992). Since the development of methods for isolating rDNA directly from environmental samples by PCR amplification (without culturing), a large number of new sequences have been identified and used for phylogenetic analyses. This method led to the identification of mesophilic Archaea (e.g., soil Archaea [Bintrim et al. 1997]). These taxa group with the Crenarchaeotes in the phylogenetic analyses thereby showing that Crenarchaeota are not exclusively hyperthermophilic. In addition, analyses using environmental DNA clones have shown that Archaea are highly abundant in the marine picoplancton, constituting up to 34% of the prokaryotic biomass in coastal arctic surface waters (DeLong et al. 1994). Interestingly, a pair of putative Archaea were also found by cloning environmental rDNA (Barns et al. 1996) that branch before the dichotomy of euryarchaeotes and crenarchaeotes. This group is provisionally called Korarchaeota, and is apparently another kingdom or Archaea (Fig. 2). The different tree analysis methods do not however all support unequivocally the monophyly of Korarchaeota and the Archaea (see below).

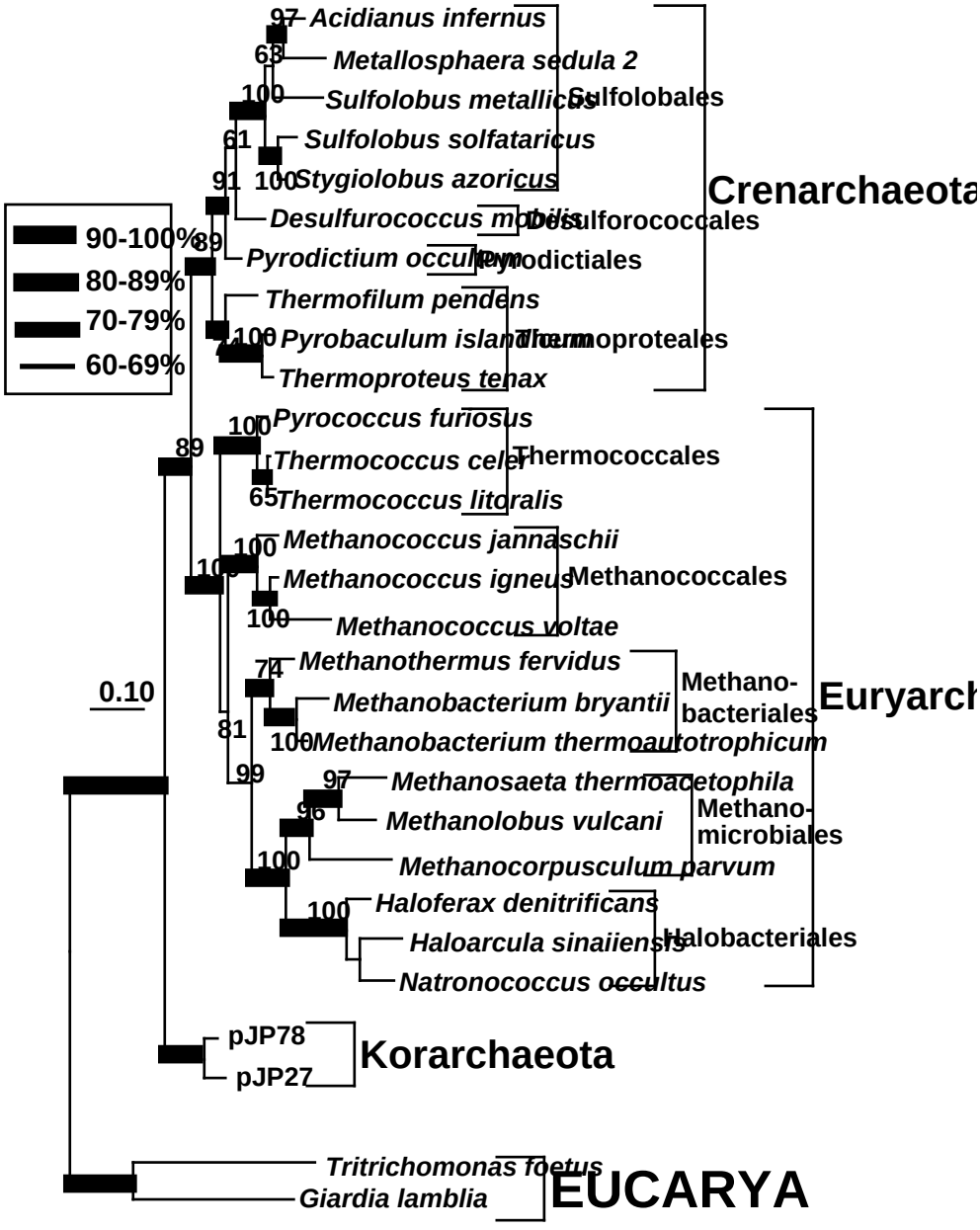

Fig. 2. Phylogeny of Archaea inferred from small subunit rDNA sequences. This tree has been inferred with the pfastDNAml program and rooted with the eukaryotic sequences. Bootstrap values (100 replications) from the ML analysis are shown above branches whereas results of distance analyses using 1000 bootstrap pseudosamples are shown as branch lengths of differing thicknesses (see legend).

In the archaeal tree (Fig 2), most of the groups are clearly separated from each other. Within the Euryarchaeota there is a clear split between the Thermococcales on the one hand and the group consisting of the methanogens and the extreme halophiles on the other. Both clusters have good bootstrap support, although the methanogen/halophile cluster have lower bootstrap support (81%) and no puzzle values. The branching order inside this cluster could not be clearly resolved, since the neighbor-joining method gave no support for the ML-supported branching order of the Methanococcales before the Methanobacteriales, and then the splitting of the Methanomicrobiales from the Halobacteriales. Whereas the Methanomicrobiales/halophile group, and its split, is strongly supported by all three methods, the neighbor-joining bootstrap analyses support weakly (52%, not shown) a clustering of the Methanococcales with the Methanobacteriales. The Crenarchaeota is split into two groups, the Thermoproteales and the Pyrodictiales/Sulfolobales/Desulfurococcales which are not resolved any further by all methods. The crenarchaeote and the euryarchaeote lineages together are clearly separated from the Korarcheota with both bootstrap analyses, whereas puzzle could not resolve the branching order of the three archeal kingdoms.

4. Phylogeny of the Extremophilic Cyanobacteria

4.1 INTRODUCTION

The broad environmental tolerance range of most extremophile cyanobacteria is in contrast to most other prokaryotic and eukaryotic organisms from extreme environments which occur only within a small window of almost constant conditions, exemplified by hyperthermophilic or acidophilic Archaea. From an evolutionary point of view it seems that the prokaryotic cyanobacteria (also called the "blue-green algae" by some authors, thereby using "algae" not as a taxonomic term, but as a convention), among the oldest organisms on earth, have been forced by the "modern" eukaryotes to withdraw themselves into habitats of extreme and fluctuating conditions (Whitton 1992, Büdel et al. 1993). An excellent overview of the diversity, ecology, and taxonomy of cyanobacteria is given by Whitton (1992).

4.2 PHYLOGENETIC METHODS

We have compared small subunit rDNA sequences from the cyanobacteria, with emphasis on extremophiles/thermophiles, to reconstruct the phylogeny of this lineage (see also Wilmotte 1994, Nelissen et al. 1995, Turner 1997). Most phylogenetic relationships were, however, resolved with only poor support from bootstrap analyses. This may be due to the fact that the majority of available cyanobacterial sequences are incomplete. In this study, representatives of almost all groups that were resolved by rDNA phylogenies presented by Wilmotte (1994) and Turner (1997) were selected, with preference for those taxa that exist in extreme habitats. The monophyletic groups resolved in our trees are consistent with those shown in the more comprehensive analysis of Turner (1997).

The alignment of cyanobacterial small subunit rDNA sequences we used for

phylogenetic analyses were taken from the rRNA WWW servers in Antwerp (Van de Peer et al. 1997) from the Ribosomal Data Project (Maidak et al. 1997). The alignment was slightly modified after the addition of some strains of *Synechococcus* using conserved small subunit rRNA secondary structure models. Trees were reconstructed with a weighted (rescaled consistency index over an interval of 1-1000) maximum parsimony method and with the neighbor-joining method with matrices inferred with the LogDet (Lockhart et al. 1994) or "HKY85" (Hasegawa et al. 1985) corrections using a test version of PAUP (version 4.0d59,written by D.L. Swofford and with permission of the author). Phylogenies inferred with maximum likelihood methods (as used in Figs. 1, 2) were almost identical with the phylogeny shown in Fig. 3. A total of 988 unequivocally aligned sequence positions were included in the phylogenetic analyses from which the constant sites were removed resulting in 530 variable sites (403 parsimony-informative characters). The cyanobacterial phylogeny was rooted with the lineage leading to the green sulfur bacteria (*Chlorolobium limicola* and *C. vibrioforme*), and other bacteria (*Agrobacterium tumifaciens*, *Alcaligines faecalis*, *Escherichia coli*, and *Pseudomonas andropogonis*; outgroups not shown in tree). The stability of internal nodes was tested with the bootstrap method (500 replications) using both neighbor-joining and parsimony methods. The rDNA sequences used in these analyses are available from the Genbank/EBI/DDBJ or RDP databases.

4.3 PHYLOGENY

Oscillatoria limnetica may serve as a model for cyanobacteria that are adapted to rapidly fluctuating environments and thus have multipotential metabolic patterns. This species is, for example, capable of shifts from oxygenic to anoxygenic photosynthesis, having multiple dark energy-generating systems, and fixing nitrogen anaerobically (Shilo 1980). In the small subunit rDNA tree, *O. limnetica* clusters with other filamentous non-heterocystous cyanobacteria from hypersaline habitats, *Arthrospira platensis*, *Lyngbya aestuarii* and *Microcoleus chthonoplastes* which form the OSCILLATORIA group (Turner 1997, Fig. 3). These cyanobacteria are primary producers in saline marshes and mangrove communities in which evaporation-dilution cycles result in broad salinity fluctuations (Javor 1989). Although heterocysts are lacking in these cyanobacteria, it is their nitrogen-fixing capacity (Stal et al. 1985, Villbrandt et al. 1990) that allows them to colonize sandy estuaries that are extremely poor in nutrients and thus they play an important role in the formation and stabilization of intertidal sediments. The various ways in whoch nitrogen fixation can be performed and is regulated in the abscence of heterocysts in cyanobacteria has been reviewed by Mitsui et al. (1986), Villbrandt et al. (1990) and Haselkorn & Buikema (1992). That *Microcoleus chthonoplastes* is dominant in intertidal microbial communities ("microbial mats", Van Gemerden 1993) is consistent with its capacity to tolerate enormous changes in salinity and forming various ecotypes within the species (Karsten 1996) which, however, are genetically rather uniform (Garcia-Pichel et al. 1996). *M. chthonoplastes* strains may be phylogenetically removed from other species of that genus as is indicated by a chemosystematic study of carotenoids and mycosporine-like amino acid compounds in various strains of *Microcoleus* (Karsten and Garcia-Pichel 1996). The paraphyletic nature of the genus *Microcoleus* is shown by the rDNA analyses (Fig. 3), *M. chthonoplastes* appears to be distant from *Microcoleus* sp. strain 10mfx. Another organism from hypersaline evironments is the unicellular *Synechocystis* sp. strain PCC 6906 which, however, is

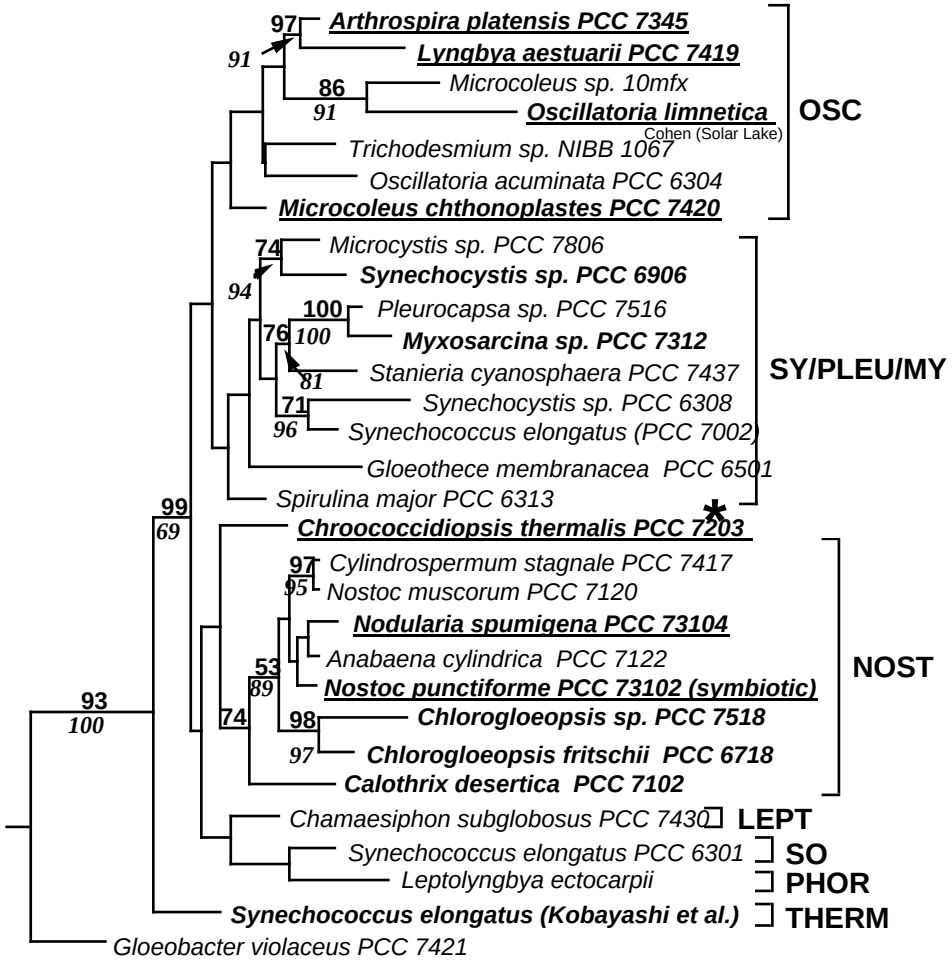

Fig. 3. Phylogeny of extremophile/thermophile cyanobacteria inferred from small subunit rDNA sequences. The phylogeny shown is from a weighted (rescaled consistency index over an interval of 1-1000) maximum parsimony analysis, the phylogram has a consistency index of 0.6633. The values above the internal nodes are from a neighbor-joining bootstrap analysis (500 replications) whereas values shown in italics below the internal nodes were inferred from a weighted maximum parsimony bootstrap analysis (100 replications). Bootstrap values are shown only for those groupings that are resolved by both the neighbor-joining and parsimony methods in more than 60% of replications. Taxa from extreme habitats are underlined, thermophiles are shown in bold face. This phylogeny has been rooted within the branch leading to the green sulfur bacteria (*Chlorolobium limicola* and *C. vibrioforme*), and other bacteria (*Agrobacterium tumifaciens*, *Alcaligines faecalis*, *Escherichia coli*, and *Pseudomonas andropogonis*; not shown). Abbreviations for groupings (bold face and upper case letters) follow Turner (1997): OSC, OSCILLATORIA group (filamentous non-heterocystous cyanobacteria); SY/PLEU/MY, SYNECHOCYSTIS/ PLEUROCAPSA/MICROCYSTIS group (unicellular and filamentous cyanobacteria); NOST, NOSTOC group (filamentous heterocystous cyanobacteria); SO, SYNECHOCOCCUS group; LEPT, LEPTOLYNGBYA group; PHOR, PHORMIDIUM group; THERM, THERMOPHILIC unicellular group. An asterisk indicates the uncertain position of *Chroococcidiopsis.*

within another cluster (SY/PLEU/MY, Fig. 3). Its close relationship with *Microcystis* sp. strain PCC 7806 from a non-extreme environment is well supported in the rDNA-tree.

Cyanobacteria are also pioneer organisms in nitrogen-depleted sandy desert soils where, in addition, they are exposed to high temperatures and dryness. Some filamentous cyanobacteria, such as *Calothrix desertica* and species of *Microcoleus*, may play an enormously important role in desert ecosystems for stabilizing soil and carbon input (Lange et al. 1992, Garcia-Pichel & Belnap 1996). *Nodularia spumigena* strain PCC 73104, another filamentous cyanobacterium that has been isolated from alkaline soils, is phylogenetically removed from *Calothrix desertica* and the two species of *Microcoleus* (Fig. 3). Species of *Nostoc* (e.g., *N. commune*, *N. flagelliforme*) frequently occur on arid soil surface. They can withstand dramatic wetting and drying cycles in combination with intense UV-light irridation (Scherer et al. 1984, Scherer & Zhong 1991). As many other poikilohydric organisms, cyanobacteria are able to "switch" their metabolic activities on and off in response to rapid changes in environmental conditions. *Nostoc* overcomes the hostile conditions on the soil surface during drought by producing a "water stress protein" (Scherer & Potts 1989) and particular UV-protecting pigments (Scherer et al. 1988, Garcia-Pichel & Castenholz 1991); the latter are found in virtually all cyanobacteria (Garcia-Pichel & Castenholz 1991). The genus *Nostoc* is a paraphyletic taxon in the rDNA analyses (Wilmotte 1994, Turner 1997). *N. punctiforme* strain PCC 73102 (associated with the roots of *Macrozamia*) is phylogenetically removed from the free-living *Nostoc muscorum* strain PCC 7120 (Fig. 3). The phylogenetic position of other terrestrial and symbiotic *Nostoc* species that are frequent photobionts of lichens, however, has not yet been investigated.

Cyanobacteria frequently occur beneath the surface of light-colored porous rock in a discrete band that follows the surface contours of the rock. Upon microscopic examination, the cyanobacteria are seen as free-living colonies attached to rock crystals or in loose lichen associations (Friedmann & Ocampo-Friedmann 1984, Bell et al. 1986, Büdel & Wessels 1991). Extremes of insolation and temperature, scouring winds and moisture stress which make the rock surface uninhabitable are thought to be the likely driving force behind colonization of the endolithic habitat. Cyanobacteria may actively bore carbonate rock or snail shells (eu-endoliths, e.g. *Myxosarcina* sp. strain PCC 7312) or inhabitat the cavates of porous rock (crypto-endoliths, e.g. *Pleurocapsa* strain PCC 7516). In hot deserts, the combination of temperature increase in a hydrated state and the concomitant loss of water imposes a severe environmental stress to cryptoendolithic cyanobacteria that eucaryotic organisms do not seem to be able to tolerate (Friedmann & Ocampo-Friedmann 1984). While temperature conditions inside the cyanobacterial zone may not be different from rock surface temperature, the more favorable water conditions and lower light intensities inside the rock may be essential factors promoting cryptoendolithic growth (Weber et al. 1996). Hot deserts are typically colonized by one or more species of a taxonomically confusing group of cyanobacteria, usually referred to as *Chroococcidiopsis*. In addition, the genus *Chroococcidiopsis* has been reported from such diverse habitats as thermal springs, marine littoral, freshwater, and from lichens (Büdel & Henssen 1983, Büdel 1992, Billi & Grilli Caiola 1996). *Chroococcidiopsis* has been regarded as the most primitive living cyanobacterium (Friedmann et al. 1994). In the rDNA analyses the phylogenetic position of

Chroococcidiopsis is unresolved. Maximum parsimony analysis places the single small subunit rDNA sequence available for *Chroococcidiopsis* (*C. thermalis* strain PCC7203, isolated from soil) close to the NOSTOC group (Fig. 3, Turner 1997) whereas *C. thermalis* clusters with *Microcoleus chthonoplastes* (OSCILLATORIA group) in neighbor-joining analyses (not shown). It is anticipated that sequence analyses of additional species of *Chroococcidiopsis* are needed to assess its phylogenetic position among the other unicellular terrestrial cyanobacteria.

Thermophilic cyanobacteria from hot springs (reviewed in Castenholz 1973) are distributed over two independent lineages of the small subunit rDNA phylogeny (Fig. 3). A thermophilic strain of the polyphyletic *Synechococcus* (Turner 1997) groups together with *Gloeobacter violaceus* strain PCC7421 at the base of the phylogram in Fig. 3 and appears to be phylogenetically distant from *S. elongatus* PCC 6301 which is from freshwater and the strain from a mud sample. Further thermophilic cyanobacteria shown in Fig. 3 are the filamentous heterocystous species of *Chlorogloeopsis* (that includes the former genus *Mastigocladus*); these cluster within the NOSTOC group.

5. A Hot Origin of Life?

The discovery of organisms that can live and grow at temperatures in excess of 100°C and branch deeply in rDNA trees has fueled the idea that these taxa are the relatives of the LCA which was, as well, thermophilic or hyperthermophilic. Eukaryotes cannot live in temperatures greater than about 60°C and in any case appear to have originated relatively "late" in evolution. The real thermophiles are within the prokaryotic domains, in particular the Archaea. The idea of a "hot" origin of life is however controversial. It has been pointed out, for example, that a hyperthermophile world for the LCA is not consistent with the RNA-world that is believed to have preceded it (Forterre 1996). RNA is highly unstable at the boiling point of water (however, see Riera et al. 1997). In addition, there is data from the distribution and origin of reverse gyrase, a topoisomerase that stabilizes topologically-closed DNA and is found in all prokaryotes living at temperatures above 80°C, that suggests that hyperthermophily is a derived and not an ancestral trait (see Forterre 1995, 1996 for details). This result is supported by other studies which show that protection of molecules against heat damage in hyperthermophiles can be achieved by relatively simple base modifications (tRNAs, Kowalak et al. 1994) or increases in the number of ion pairs in protein tertiary structures (*Taq* polymerase, Korolev et al. 1995). This suggests that thermophily may have evolved from a mesophilic ancestor via thermoadaptation.

In all our phylogenies however, the thermophilic organisms show up as early divergences. This was also shown by other analyses (Stetter 1996). The (known) korarchaeotes also appear to be thermophilic with respect to the hot environment from which they were extracted (Barns et al. 1996) and the culturing and characterization of pJP27 as a thermophilic Archaea by Burggraf et al. (1997). These observations imply that the common ancestor of the Archaea might have been thermophilic. Although many small subunit rRNA sequences of putative mesophilic crenarchaeotes have been extracted directly from the environment, most of those species have not yet been cultured and further characterized. Therefore, we do not yet know which of these are really mesophilic

or perhaps are thermophiles existing in a mesophilic environment. The lower branches of the bacterial domain are also occupied by hyperthermophilic bacteria such as *Aquifex* and *Thermotoga* (see Fig. 1A). This branching order, also shown by Barns et al (1996) and Stetter (1996), is consistent with a hyperthermophilic ancestry of the Bacteria. Given the thermophilic life-style of the (known) korarchaeotes, it is plausible therefore that the common ancestor of the Archaea and the Bacteria (therefore also the LCA) may have been a thermophile or hyperthermophile. Within the cyanobacteria, we also find thermophiles as early diverging (*Synechococcus elongatus*) and as more derived (*Chlorogloeopsis* sp.) taxa suggesting that thermophily was a primitive trait in these derived Bacteria.

In spite of these results the origin of life need not have happened in a thermophilic environment. Another explanation for our results is that all (currently known) organisms may have had a common ancestor that was adapted to the hyperthermophile environment (Forterre 1996). In this scenario, life evolved in a mesophile world that later became hot and led to the evolution of the hyperthermophilic LCA from which all extant cells are derived. To gain a better understanding therefore of the origin of thermophiles in the phylogeny of life and in the Archaea in particular, we need analyses of sequences drawn from a larger diversity of environments, and the biological characterization of these organisms to gain deeper insights into the phylogeny of the genes that allow life at extreme temperatures. The results published in the last few years promise that we will have the possibility to reconstruct a much more precise tree of life in the near future.

6. Acknowledgements

This research was supported in part from a grant from the University of Iowa to DB. We thank the German National Research Center for Information Technology (GMD, Sankt Augustin), NEC Europe Ltd. (C&C Research Laboratories, Sankt Augustin) and the German Cancer Resarch Center (DKFZ, Heidelberg) for access to computing facilities. TF thanks Burkhard Buedel (Kaiserslautern) for valuable discussions.

7. References

Barns, S.M., Delwiche, C.F., Palmer, J.D., and Pace, N.R. (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences, *Proc. Natl. Acad. Sci. USA* **93**, 9188-9193.

Bell, R.A., Athey, P.V., and Sommerfeld, M.R. (1986) Cryptoendolithic algal communities of the Colorado Plateau, *J. Phycol.* **22**, 429-435.

Billi, D. and Grilli Caiola, M. (1996) Effects of nitrogen limitation and starvation on *Chroococcidiopsis* sp. (Chroococcales), *New Phytol.* **133**, 563-571.

Bintrim, S.B., Donohue, T.J., Handelsman, J., Roberts, G.P., and Goodman, R.M. (1997) Molecular phylogeny of Archaea from soil, *Proc. Natl. Acad. Sci. USA* **94**, 277-282.

Brown, J.R. and Doolittle. W.F. (1995) Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications, *Proc. Natl. Acad. Sci. USA* **92**, 2441-2445.

Büdel, B. (1992) Taxonomy of lichenized procaryotic blue-green algae, in W. Reisser (ed.), *Algae and Symbioses*, Biopress Ltd., Bristol. pp. 301-324.

Büdel, B. and Henssen, A. (1983) *Chroococcidiopsis* (Cyanophyceae), a phycobiont in the lichen family Lichinaceae, *Phycologia* **22**, 367-375.

Büdel, B. and Wessels, D.C.J. (1991) Rock inhabiting blue-green algae/cyanobacteria from hot arid regions, *Algological Studies* **64**, 385-398.

Büdel, B., Wessels, D.C.J., and Mollenhauer, D. (1993) Massenentwicklung von *Nostoc* cf. *microscopicum* (Carmichael) Harvey ex Bornet & Flahault in alkalischem, salzhaltigem Milieu in einer Höhle in den Drakensbergen Südafrikas (Golden Highlands National Park), *Arch. Protistenkd.* **143**, 229-235.

Burggraf, S., Huber, H., and Stetter, K.O. (1997) Reclassification of the crenarchael orders and families in accordance with 16S rRNA sequence data, *Int. J. Syst. Bacteriol.* **47**, 657-660.

Castenholz, R.W. (1973) Ecology of blue-green algae in hot springs, in N.G. Carr and B.A. Whitton (eds.), *The biology of blue-green algae*, Blackwell, Oxford, pp. 379-414.

DeLong, E.F., Wu, K.Y., Prézelin, B.B., and Jovine, R.V.M. (1994) High abundance of Archaea in Antarctic marine picoplancton, *Nature* **371**, 695-697

Doolittle, R.F., Feng, D.-F., Tsang, S., Cho, G., and Little, A. (1996) Determining divergence times of the major kingdoms of living organisms with a protein clock, *Science* **271**, 470-477.

Driessen, A.J.M., van de Vossenberg, J.L.C.M, and Konings, W.N. (1996) Membrane composition and ion-permeability in extremophiles, *FEMS Microbiol. Rev.* **18**, 139-148

Forterre, P. (1996) A hot topic: the origin of hyperthermophiles, *Cell* **85**, 789-792.

Forterre, P. (1995) Thermoreduction, a hypothesis for the origin of prokaryotes, *C R Acad Sci III* **318**, 415-422.

Friedmann, E.I. and Ocampo-Friedmann, R. (1984) Endolithic microorganisms in extreme dry environments: analysis of a lithobiontic microbial habitat, in: M.J. Klug, C.A. Reddy (eds.) *Current perspectives in microbial ecology*, American Society for Microbiology, Washington D.C., pp.177-185.

Friedmann, E.I., Ocampo-Friedmann, R., and Hua, M. (1994) *Chroococcidiopsis*, the most primitive living cyanobacterium?, *Origins Life Evol. Biosphere* **24**, 269-270.

Fuhrman, J.A., McCallum, K., and Davis, A.A. (1992) Novel major archaebacterial group from marine plankton, *Nature* **356**, 148-149.

Garcia-Pichel, F. and Belnap, J. (1996) Microenvironments and microscale productivity of cyanobacterial desert crusts, *J. Phycol.* **32**, 774-782.

Garcia-Pichel, F. and Castenholz, R.W. (1991) Characterization and biological implications of scytonemin, a cyanobacterial sheath pigment, *J. Phycol.* **27**, 395-409.

Garcia-Pichel, F., Prufert-Bebout, L., and Muyzer, G. (1996) Phenotypic and phylogenetic analyses show *Microcoleus chthonoplastes* to be a cosmopolitan cyanobacterium, *Applied and environmental Microbiology* **62**, 3284-3291.

Gogarten, J.P., Kibaj, H., Dittrich, O., Taiz, I., Bowman, E.J., Bowman, B.J., Manolson, M.F., Poole, R.J., Date, T., Oshima, T., et al. (1989) Evolution of the vacuolar H^+-ATPases: implications for the origin of eukaryotes, *Proc. Natl. Acad. Sci. USA* **86**, 6661-6665.

Golding, B. (1996) Evolution: when was life's first branch point?, *Curr. Biol.* **6**, 679-682.

Gupta, R.S. and Golding, G.B. (1996) The origin of the eukaryotic cell. *Trends Biochem. Sci*. **21**, 166-171.

Gupta, R.S. and Singh, B. (1994) Phylogenetic analysis of 70 kD heat shock protein sequences suggests a chimeric origin for the eukaryotic cell nucleus, *Curr. Biol.* **4**, 1104-1114.

Hasegawa, M., Kishino, H., and Yano, T. (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA, *J. Mol. Evol.* **22**, 32-38.

Haselkorn, R. and Buikema, W.J. (1992) Nitrogen fixation in cyanobacteria, in G. Stacey, R.H. Burris, and H.J. Evans (eds.), *Biological nitrogen fixation*, Chapman and Hall, New York, London, pp. 166-190.

Holland, H. and Beukes,, N. (1990) A paleoworthy profile from Griqualand West, South Africa: evidence for a dramatic rise in atmospheric oxygen between 2,2 and 1.9 BYBP, *Am. J. Sci.* **290A**, 1-34.

Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., and Miyata, T. (1989) Evolutionary relationships of archaebacteria, eubacteria and eukaryotes inferred from phylogenetic trees of duplicated genes, *Proc. Natl. Acad. Sci. USA* **86**, 9355-9359.

Javor, B. (1989) *Hypersaline environments. Microbiology and biogeochemistry*, Springer-Verlag, Berlin, New York, London, Paris, Tokyo, pp. 134-146.

Jones, W.J., Nagle, D.P., and Whitman, W.B. (1987) Methanogens and the Diversity of Archaebacteria, *Microbiol. Rev.* **51**, 135-177

Karsten, U. (1996) Growth and organic osmolytes of geographically different isolates of *Microcoleus chthonoplastes* (Cyanobacteria) from benthic microbial mats: response to salinity change, *J. Phycol.* **32**, 501-506.

Karsten, U. and Garcia-Pichel, F. (1996) Carotenoids and mycosporine-like amino acid compounds in members of the genus *Microcoleus* (Cyanobacteria): a chemosystematic study, *System. Appl. Micobiol.* **19**, 285-294.

Keeling, P.J. and Doolittle, W.F. (1995) Archaea: narrowing the gap between prokaryotes and eukaryotes, *Proc. Natl. Acad. Sci. USA* **92**, 5761-5764.

Knoll, A.H. (1994) Proterozoic and Early Cambrian protists: evidence for accelerating tempo, *Proc. Natl. Acad. Sci. USA* **91**, 6743-6750.

Korolev, S., Nayal, M., Barnes, W.M., Di Cera, E., and Waksman, G. (1995) Crystal structure of the large fragment of *Thermus aquaticus* DNA polymerase I at 2.5-Å resolution: structural basis for thermostability, *Proc. Natl. Acad. Sci. USA* **92**, 9264-9268.

Kosa, P.F., Ghosh, G., DeDecker, B.S., and Sigler, P.B. (1997) The 2.1-Å crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box, *Proc. Natl. Acad. Sci. USA* **94**, 6042-6047.

Kowalak, J.A., Dalluge, J.J., McCloskey, J.A., and Stetter, K.O. (1994) The role of posttranscriptional modification in stabilization of transfer RNA from hyperthermophiles, *Biochemistry* **33**,7869-7876.

Lake, J.A., and Rivera, M.C. (1994) Was the nucleus the first endosymbiont? *Proc. Natl. Acad. Sci. USA* **91**, 2880-2881.

Lange, O.L., Kidron, G.J., Büdel, B., Meyer, A., Kilian, E., and Abeliovich, A. (1992) Taxonomic composition and photosynthetic characteristics of the "biological soil crusts" covering sand dunes in the western Negev desert, *Functional Ecology* **6**, 519-527**.**

Langer, D., Hain, J., Thuriaux, P., and Zillig, W. (1995) Transcription in Archaea: similarity to that in Eucarya, *Proc. Natl. Acad. Sci. USA* **92**, 5768-5772.

Lockhart, P.J., Steel, M.A., Hendy, M.D., and Penny, D. (1994) Recovering evolutionary trees under a more realistic model of sequenc evolution, *Mol. Biol. Evol.* **11**, 605-612.

Lowe, S.E., Jain, M.K., and Zeikus, G. (1993) Biology, ecology, and biotechnological applications of anaerobic bacteria adapted to environmental stresses in temperature, pH, salinity, or substrates, *Microbiol. Rev.* **57**, 451-509.

Madigan, M.T. and Marrs, B.L. (1997) Extremophiles, *Sci. Amer.* **276**, 82-87.

Maidak, B.L., Olsen, G.J., Larsen, N., Overbeek, R., McCaughey, M.J., and Woese CR (1997) The RDP (Ribosomal Database Project), *Nucleic Acids Res.* **25**, 109-111.

Martin, W.F. (1996) Is something wrong with the tree of life? *BioEssays* **18**, 523-527.

Mitsui, A., Kumazawa, S., Takahashi, A., Ikemoto, H., Cao, S., and Arai, T. (1986) Strategy by which nitrogen-fixing unicellular cyanobacteria grow photoautotrophically, *Nature* **323**, 720-722.

Nelissen, B., Van de Peer, Y., Wilmotte, A., and De Wachter, R. (1995) An early origin of plastids within the cyanobacterial divergence is suggested by evolutionary trees based on complete 16S rRNA sequences, *J. Mol. Evol.* **12**, 1166-1173.

Olsen, G.J., Matsuda, H., Hagstrom, R., and Overbeek, R. (1994) fastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood, CABIOS **10**:41-48.

Pennisi, E. (1997) In Industry, Extremophiles Begin to Make Their Mark, *Science* **276**, 705-706

Riera, J., Robb, F.T., Weiss, R., and Fontecave, M. (1997) Ribonucleotide reductase in the archaeon Pyrocossus furiosus: a critical enzyme in the evolution of DNA genomes, *Proc. Natl. Acad. Sci. USA* **94**, 475-478.

Rowlands, T., Baumann, P., and Jackson, S.P. (1994) The TATA-binding protein: a general transcription factor in eukaryotes and archaebacteria, *Science* **264**, 1326-1329.

Scherer, S., Chen, T.W., and Böger, P. (1988) A new UV-A/B protecting pigment in the terrestrial cyanobacterium *Nostoc commune*, *Plant Physiol.* **88**, 1055-1057.

Scherer, S., Ernst, A., Chen, T.W., and Böger, P. (1984) Rewetting of drought resistant blue-green algae: time course of water uptake and reappearance of respiration, photosynthesis and nitrogen fixation, *Oecologia* **62**, 418-423.

Scherer, S. and Potts, M. (1989) Novel waterstress protein from a desiccation-tolerant cyanobacterium. Purification and partial characterization, *J. Biol. Chem.* **264**, 12546-12553.

Scherer, S. and Zhong, Z.P. (1991) Desiccation independence of terrestrial *Nostoc commune* ecotypes (Cyanobacteria), *Microb. Ecol.* **22**, 271-283.

Schopf, J.W. (1993) Microfossils of the early Archaean apex chart: new evidence of the antiquity of life, *Science* **260**, 640-646.

Schopf, J.W. (1994) Disparate rates, differing fates: tempo and mode of evolution changed from the Precambrian to the Phanerozoic, *Proc. Natl. Acad. Sci. USA* **91**, 6735-6742.

Schopf, J.W. and Walter, M.R. (1982): Origin and early evolution of cyanobacteria: the geological evidence, in N.G. Carr and B.A. Whitton (eds.) *The. - biology of cyanobacteria*. Blackwell, Oxford, pp. 543-564.

Shilo, M. (1980) Strategies of adaptation to extreme conditions in aquatic microorganisms, *Naturwissensch.* **67**, 384-389.
Stal, L.J., van Gemerden, H., and Krumbein, W.E. (1985) Structure and development of a benthic marine microbial mat, *FEMS Microbiol. Ecol.* **31**, 111-125.
Stetter, K.O. (1996) Hyperthermophilic procaryotes, *FEMS Microbiol. Rev.* **18**, 149-158.
Strimmer, K. and von Haeseler, A. (1996) Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies, Mol. Biol. Evol. **13**:964-969.
Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.* **22**, 4673-4680.
Turner, S. (1997) Molecular systematics of oxygenic photosynthetic bacteria, in D. Bhattacharya (ed.), *Origins of Algae and Their Plastids*, Springer-Verlag, Wien, in press.
Van de Peer, Y., De Rijk, P., and De Wachter, R. (1997) SSU rRNA database. Department of Biochemistry, University of Antwerp http://rrna.uia.ac.be/rrna/ssuform.html#Eukarya.
Van Gemerden, H. (1993) Microbial mats: a joint venture, *Mar. Geol.* **113**, 3-25.
Villbrandt, M., Stal, L.J., and Krumbein, W.E. (1990) Interactions of nitrogen fixation and oxygenic photosynthesis in marine cyanobacterial mats, *FEMS Microbiol. Ecol.* **74**, 59-72.
Weber, B., Wessels, D.C.J., and Büdel, B. (1996) Biology and ecology of cryptoendolithic cyanobacteria of a sandstone outcrop in the Northern Province, South Africa, *Algological Studies* **83**, 565-579.
Whitton, B.A. (1992) Diversity, ecology, and taxonomy of the Cyanobacteria, in N.H. Mann and N.G. Carr (eds.), *Photosynthetic Prokaryotes*, Plenum Press, New York, pp. 1-51.
Wilmotte, A. (1994) Molecular evolution and taxonomy of the Cyanobacteria, in A. Bryant (ed.) *The Molecular Biology of Cyanobacteria*, Kluwer Academic Publishers, Dordrecht, pp. 1-25.
Whitton, B.A. and Potts, M. (1982): Marine littoral, in N.G. Carr, and B.A. Whitton (eds.) *The biology of cyanobacteria*. Blackwell, Oxford, pp. 515-542.
Woese, C.R (1987) Bacterial evolution, *Microbiol. Rev.* **51**, 221-271.
Woese, C.R., Kandler, O., and Wheelis, M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya, *Proc. Natl. Acad. Sci. USA* **87**, 4576-4579.
Zillig, W. (1991) Comparative biochemistry of Archaea and Bacteria, *Curr. Opin. Genet. Dev.* **1**, 544-551.