*Geomicrobiology Journal*, 20:1–14, 2003

0149-0451/03 $12.00 + .00
DOI: 10.1080/01490450390144330

Taylor & Francis
Taylor & Francis Group

# Estimated Minimal Divergence Times of the Major Bacterial and Archaeal Phyla

PETER P. SHERIDAN

Department of Biological Sciences
Idaho State University
Pocatello, Idaho, USA

KATHERINE H. FREEMAN

Department of Geosciences
The Pennsylvania State University
University Park, Pennsylvania, USA

JEAN E. BRENCHLEY

Department of Biochemistry and Molecular Biology
The Pennsylvania State University
University Park, Pennsylvania, USA

*Previous calibrations of prokaryotic phylogenetic events were based on the vertebrate fossil record because a detailed microbial fossil record does not exist. Recently, compounds (2-methylhopanoids) that are found in cyanobacterial membranes were identified among compounds extracted from late Archean sedimentary rocks. These lipids establish a minimum time, 2.65 Ga (Billion Years Ago), in the geologic record for the existence of cyanobacteria. We have used this new information to calibrate a Neighbor Joining distance tree generated from SSU rRNA sequences of major prokaryotic lineages and to estimate the dates for significant events in the history of life on Earth. The results suggest that the Last Common Ancestor occurred at about 4.29 Ga and that the individual Bacterial and Archaeal Domains began radiating about 3.46 Ga.*

**Keywords** 16S rRNA, bacteriohopanepolyols, divergence, phylogeny, prokaryotes

## Introduction

### *Fossils and the Phylogeny of Life in the Archean Era*

The timing of events from the origin of life to the Last Common Ancestor (LCA) and to the divergences of the major taxa in the phylogeny of life is of great interest to Earth and

Received 5 June 2001; accepted 22 October 2001.

We thank our laboratory coworkers for suggestions and helpful discussions. This work was supported by Department of Energy Grant to J.E.B. P.P.S. has also been supported by an Alfred P. Sloan Foundation Fellowship in Molecular Evolution from the National Science Foundation and partial funding from the Penn State Astrobiology Center and NSF/IGERT for Biogeochemical Research Initiative for Education (BRIE). Support was also provided by NSF-EAR to K.H.F.

Address correspondence to Peter P. Sheridan, Department of Biological Sciences, Box 8007, Idaho State University, Pocatello, ID 83209, USA. E-mail: sherpete@isu.edu

Life scientists alike. Most efforts at dating the significant events in the early history of life focused on dating the origin and divergences of multicellular eukaryotes, which have a much more extensive fossil record than the prokaryotes (Doolittle et al. 1996; Wray et al. 1996; Feng et al. 1997; Ayala et al. 1998; Wang et al. 1999). The lack of a well-defined microbial fossil record has meant that previous calibrations of phylogenetic distance trees containing microbial sequence data are based on the vertebrate fossil record (Doolittle et al. 1996; Feng et al. 1997), or used a data set skewed towards a limited number of bacterial taxa and calibration dates which are questionable (Ochman and Wilson 1987). The selective pressures on single-celled prokaryotes are vastly different from those on reptiles and mammals, possibly resulting in different rates of molecular evolution. Thus, an independent date of phylogenetic divergence is an important test of the validity of using vertebrate data to calibrate prokaryotic events. The compounds associated with cyanobacterial membranes in rocks 2.65 Ga provide this independent marker for calibrating these events.

Bacteriohopanepolyols with a methyl carbon at position 2 are found in a wide distribution of cultured and environmental samples of cyanobacteria (Summons et al. 1999) and although reported in other organisms, they are only found in significant quantities in cyanobacteria (Summons et al. 1999). The geological forms of these compounds, 2-methylhopanes, are formed following diagenesis and extended burial of bacterial remains in sediments, and are considered definitive markers for the presence of cyanobacteria on the ancient Earth (Summons et al. 1999). The occurrence of these compounds in Late Archean shales (Brocks et al. 1999; Summons et al. 1999) within the Hamersley Basin of Western Australia, place the oldest molecular evidence of cyanobacteria between 2.63 and 2.68 Ga (Arndt et al. 1991; Trendall et al. 1998). This is consistent with widespread microfossil and isotopic evidence for photosynthetic life at this time (Schidlowski et al. 1983; Schopf and Packer 1987) and confirms an early presence of oxygenic photosynthesis (e.g., Knoll 1999; Watanabe et al. 2000; Nisbet and Sleep 2001; Schidlowski 2001). Younger morphological fossils are invaluable in estimating divergence events for multicellular eukaryotes; likewise, this earliest occurrence of cyanobacterial biomarkers provides a unique and useful tool for examining the divergence dates of prokaryotic phyla.

The vast phylogenetic diversity of prokaryotic organisms was revealed by Woese's analysis of SSU rRNA (16S) gene sequences (Woese et al. 1990). The SSU rRNA gene sequence tree indicated that the Gram-positive bacteria, Proteobacteria, and cyanobacteria diverged at about the same time and that several other taxa were well established before this divergence. The cyanobacteria-Proteobacteria-Gram-positive bacteria divergence was also shown to have occurred well after the divergence of the Bacterial and Archaeal Domains. The proposed dates of the Proteobacteria-Gram-positive bacteria split (and therefore also the cyanobacterial divergence) of ∼1.5 Ga (Doolittle et al. 1996) and ∼2.2 Ga (Feng et al. 1997) are comparable to the age of rocks (∼1.0–2.0 Ga) in which Schopf (Schopf 1994) and others have found fossils that closely resemble modern cyanobacterial taxa. However, Schopf (1993) has proposed that fossils found in rocks dated to ∼3.5 Ga are also cyanobacteria. The existence of cyanobacteria as early as 3.5 Ga requires that life originated and split into the two prokaryotic Domains which then underwent extensive radiations within the relatively short time after the formation of the Earth and the cessation of the bombardment phase. Acceptance of the Apex Chert fossils as cyanobacteria would indicate that the rates of evolution in early prokaryotic taxa and later branching organisms are so dissimilar that making molecular comparisons between most genera is meaningless.

The assignment of divergence dates requires well-dated earliest appearances of fossils that are specific to certain metabolisms or taxa. The analysis of cyanobacteria-specific molecular fossils such as 2-methylhopanoids allows calibration of distance trees and permits estimates of the major prokaryotic divergences. Discrepancies between fossil record

divergence dates and divergence dates based on the analysis of protein molecular data using vertebrate fossil calibrations (Doolittle et al. 1996; Feng et al. 1997) have been acknowledged for early prokaryotic evolutionary patterns. An estimation of a few divergence dates for some of the well-known proteobacterial and Gram-positive species (Ochman and Wilson 1987) was hampered by the extremely limited SSU rRNA data set used, as well as by the use of a mixture of eukaryotic and eubacterial fossil data. The report (Brocks et al. 1999; Summons et al. 1999) of 2-methylhopanes in Late Archean shales dated to ages between 2.63 and 2.68 Ga (Arndt et al. 1991; Trendall et al. 1998) inspired a limited attempt (Knoll 1999) to place dates on a Woese's SSU rRNA phylogenetic tree (Woese et al. 1990). The use of molecular fossils such as 2-methylhopanes and a data set that more accurately reflects the abundant phylogenetic diversity of prokaryotic organisms may help to resolve the discrepancies in divergence dates previously reported.

### *Selection of SSU rRNA for Comparison*

The distance measurements previously used to determine the timing of events in Earth's early biotic history were based on protein sequence data (Doolittle et al. 1996; Feng et al. 1997). We chose to analyze SSU rRNA sequences for our distance tree for the following reasons. First, SSU rRNA provides much more extensive coverage of prokaryotic diversity than the available protein sequences. For example, a recent release (version 8.0) from the Ribosomal Database Project (RDP), at the Center for Microbial Ecology at Michigan State University (Maidak et al. 2000) contained 16,277 prokaryotic SSU rRNA sequences. Although there is some duplication among the 16S rRNA sequences, there is no other molecule in any database with so many accumulated phylogenetically distinct sequences. Likewise, the growing number of prokaryotic genomes being sequenced have concentrated on a relatively few genera. Thus, the use of protein sequence data can overlook major prokaryotic phylogenetic groups. Second, proteins are subjected to numerous selective pressures that SSU rRNA sequences are not. The rRNA sequence is mainly under selective pressures relating to structural interactions, as opposed to proteins that are also under more direct functional, kinetic, physiological, and mechanistic constraints that can affect the rates of change of the proteins in the individual taxa. Third, while the alignments of the SSU rRNA data are based on secondary structural analysis, in most cases the alignments of distantly related proteins are not. This may lead to misalignment of the protein sequences, which will adversely affect sequence comparisons. Fourth, the phylogenies of proteins can be confusing in cases where gene transfers (Nelson et al. 1999) and gene duplications (Delwiche and Palmer 1996) have occurred. This could lead to the analysis of paralogous genes (resulting from a duplication event in a shared ancestor) which are incorrectly believed to be homologues (resulting from speciation events).

In contrast, reports of the possible lateral transfer of SSU rRNA gene sequences between prokaryotic genera are exceedingly rare (Wang et al. 1997; Dennis et al. 1998; Ueda et al. 1999; Yap et al. 1999) based on extensive study of sequences in a wide diversity of organisms. We suggest that SSU rRNA data rather than protein sequence data enable a more accurate determination of the divergence times of the major prokaryotic lineages.

## Materials and Methods

### *Generation and Calibration of the Distance Tree*

All estimates of divergence times are subject to the difficulty that molecular data exhibit "clock-like" behavior over a range of temporal scales (Ayala et al. 1996). The number of

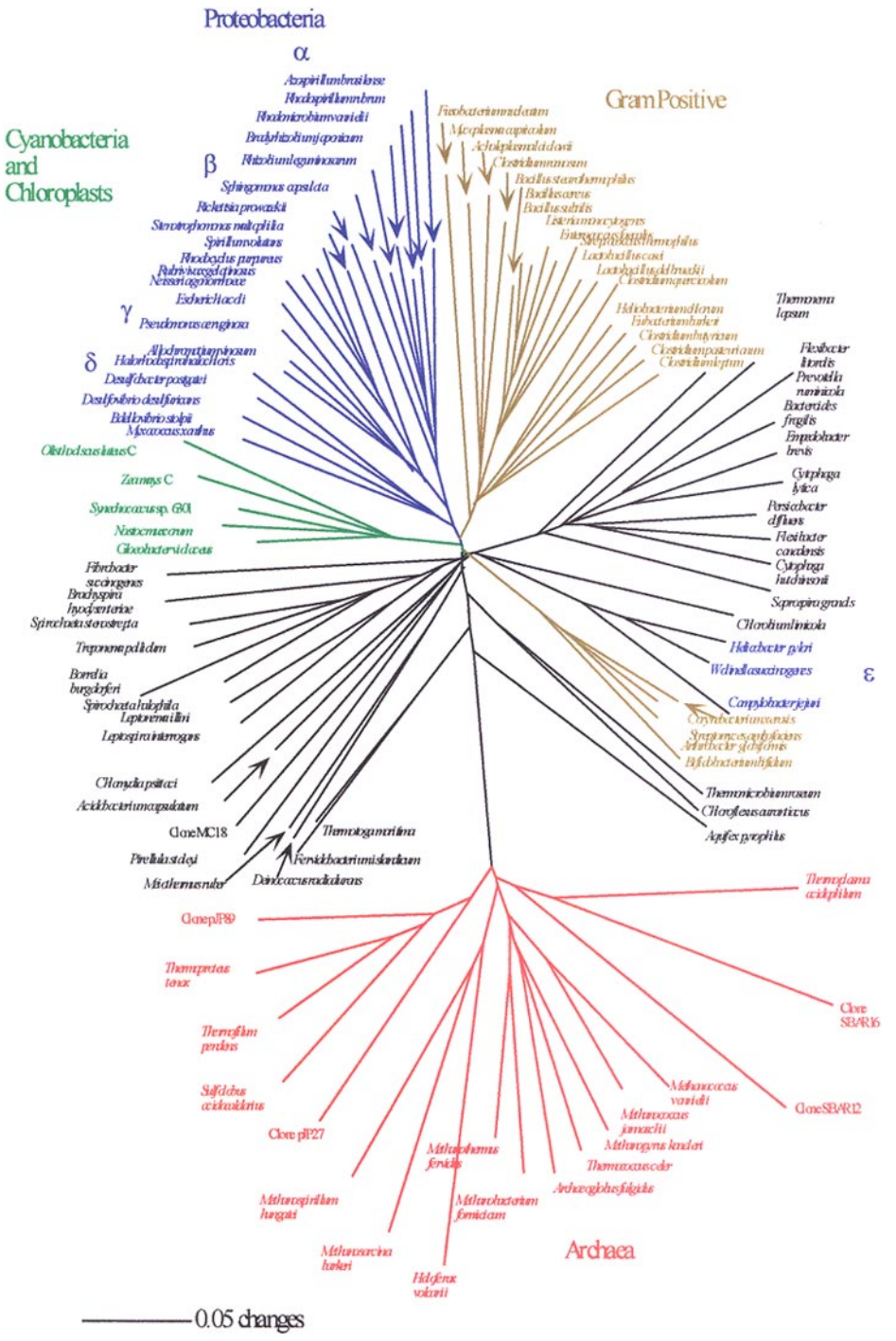


**FIGURE 1** Unrooted phylogram of 98 representative prokaryotic organisms. The SSU rRNA sequences of 98 representative prokaryotic organisms were downloaded from the Ribosomal Database Project (RDP), version 7.0 at the Center for Microbial Ecology at Michigan State University (Maidak et al. 2000). The sequence alignment used 1840 characters, and was based on the RDP alignment, which incorporates secondary structural *(Continued)*

changes in a sequence versus time deviates from linearity due to multiple mutations at a single site. Some of these mutations result in the reversion of a character state to the original state (homoplasies), thus leading to an underestimate of differences and distance between distantly related sequences. The need to correct for the effect of homoplasies is evident in the multitude of algorithms designed for phylogenetic analysis, but attempts to correct for such rate effects can introduce biases into the analyses. Recent work (Takahashi and Nei 2000), however, has shown that when the data set is large (48 sequences) relative to the number of nucleotides (1,000 for low divergence [$d\max = 0.1$] or 300 for high divergence [$d\max = 1.0$]), less computationally intense analyses produce the correct phylogeny as those incorporating various correction factors. To increase the likelihood that the estimated distances we obtained were accurate, we used a large data set (an alignment of 1840 characters for each SSU rRNA gene sequence from 98 representative prokaryotic taxa). The major prokaryotic branch point divergences (Figures 1 and 2) were established using the Neighbor Joining algorithm and an uncorrected "p" distance measure generated by the PAUP (version 4.0b4a) program (Swofford 2000). To determine the effects of using various corrections on these results, we compared our results with Neighbor Joining distance trees generated using the Jukes-Cantor, Kimura 2-parameter, Kimura 3-parameter, Tamura-Nei, Tajima-Nei, and HKY85 models with equal rates for variable sites. These trees were roughly equivalent to the tree obtained using the uncorrected "p" distance (data not shown). Unequal rates for variable sites were also examined for all six models using gamma parameters with shapes of 0.5, 1.0, 2.0, and 5.0. Distance trees for gamma parameters of 1.0, 2.0, and 5.0 were basically equivalent to the corresponding tree obtained with equal rates for variable sites (data not shown).

The times of major divergences on our prokaryotic distance tree (Figure 2) were determined using the average cyanobacterial distance and a minimum time estimate for the origin of the cyanobacterial lineage obtained using 2-methylhopanoid molecular fossils. We used the date of 2.65 Ga determined for sedimentary rocks bearing the 2-methylhopanes (Summons et al. 1999) to calibrate the cyanobacterial distance. The average cyanobacterial distance was calculated in two ways: as the average distance between the two chloroplast sequences and the common ancestor of all cyanobacteria, and as the average distance between the three free-living cyanobacterial sequences and the common ancestor of all cyanobacteria. The results of these two calculations were averaged and this distance was used in all subsequent calculations as the representative distance of the cyanobacterial lineage. The use of this average should correct for sample bias from the selection of cyanobacterial/chloroplast taxa which had either slower or faster generation times than the average generation time, which would distort branch lengths on the tree.

Placing more cyanobacterial and/or chloroplast sequences on the tree would distort the tree in favor of the cyanobacterial lineage, artificially inflating its importance with respect to the other taxa. The distances between the Last Common Ancestor and the Bacterial and Archaeal roots were set to be equal, under the assumption that the rates of evolution were similar in both domains. The distance between the Last Common Ancestor and the common ancestor of all cyanobacteria was determined, and independently added to the representative distance of the cyanobacterial lineage. The age of the Last Common Ancestor was calculated

**FIGURE 1** (Continued) information. The data were analyzed using the PAUP package, version 4.0b3a (Swofford 2000). A Neighbor Joining search was performed, using an uncorrected "p" distance measure, and the resulting tree was visualized as an unrooted phylogram. Archaea are shown in red, Proteobacteria in purple, cyanobacteria and chloroplasts in green, and Gram-positive organisms in brown.

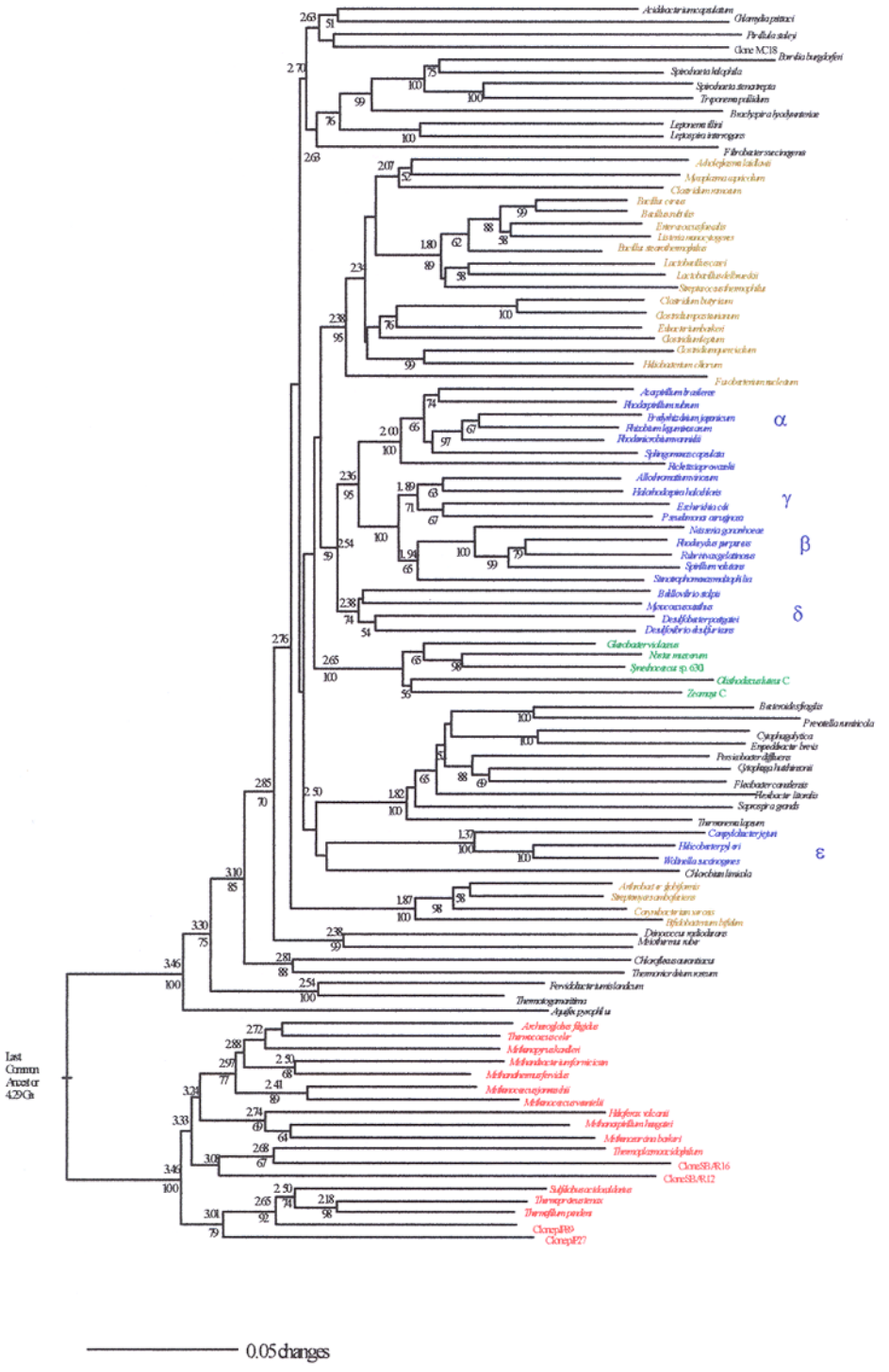


**FIGURE 2** Major divergence dates in prokaryotic history. This is the rectangular phylogram resulting from a Neighbor Joining search using an uncorrected "p" distance measure of SSU rRNA sequence data from 98 representative prokaryotic organisms for 1840 characters. We chose to root our tree with the LCA of the Bacteria and Archaea *(Continued)*

**TABLE 1** Divergence dates of major nodes in the history of the Archaeal Domain

| Node | Distance | When node arose (Ga) |
|---|---|---|
| LCA | 0 | 4.29 |
| Euryarchaeota/Crenarchaeota split | .037 | 3.46 |
| Euryarchaeota group | | |
| Archaeoglobus/Thermoplasma split | .043 | 3.33 |
| Archaeoglobus/Haloferax split | .047 | 3.24 |
| Archaeoglobus/Methanococcus split | .059 | 2.97 |
| Archaeoglobus/Methanobacterium split | .063 | 2.88 |
| Thermoplasma/SBAR12 split | .054 | 3.08 |
| Thermoplasma/SBAR16 split | .072 | 2.68 |
| Methanosarcina/Haloferax split | .069 | 2.74 |
| Methanococcus split | .084 | 2.41 |
| Methanopyrus/Archaeoglobus split | .070 | 2.72 |
| Methanothermus/Methanobacterium split | .080 | 2.50 |
| Crenarchaeota group | | |
| Clone JP27/Thermofilum split | .057 | 3.01 |
| Clone JP89/Thermofilum split | .073 | 2.65 |
| Sulfolobus/Thermofilum split | .080 | 2.50 |
| Thermoproteus/Thermofilum split | .094 | 2.18 |

based on this distance. The time from the divergence of the LCA to major prokaryotic divergence events was calculated based on the distance between the LCA and the node in question, using a simple ratio of the representative distance of the cyanobacterial lineage and the 2.65 billion year cyanobacterial lineage calibration. This time was then subtracted from the age of the LCA to give the date of the divergence of the node. Unfortunately, at this time the PAUP program does not allow for the estimation of the error in the calculation of internode distances. However, one should keep in mind that the divergence dates proposed here are minimum estimates, and are not meant to signify specific biological events.

## Results

### *Divergence Time Estimates for the LCA and the Major Prokaryotic Lineages*

Our estimation of minimum divergence times for major events in early prokaryotic history leads to several interesting observations and speculations. For example, the Last Common Ancestor of the prokaryotic domains occurred 4.29 Ga, while the Last Common Ancestors of the individual Bacterial and Archaeal domains occurred at least 3.46 Ga (Table 1). The 3.46 Ga date would also be when the Crenarchaeota and the Euryarchaeota split, with the

**FIGURE 2** (Continued) occurring midway between the two domains, under the assumption that the rates of evolution were similar in both domains. The LCA was placed on the tree subsequent to the analysis. Estimated divergence times (in billions of years ago Ga) are shown above the branches, while bootstrap values are shown below the branches. A 50% majority rule consensus limit was in effect, and the number of bootstrap replicates was 10,000. Archaea are shown in red, Proteobacteria in purple, cyanobacteria and chloroplasts in green, and Gram-positive organisms in brown.

**TABLE 2** Divergence dates of major nodes in the history of the Bacterial Domain

| Node | Distance | When node arose (Ga) |
|---|---|---|
| LCA | 0 | 4.29 |
| Aquifex | .037 | 3.46 |
| Thermotoga group | .044 | 3.30 |
| Thermotoga/Fervidobacterium split | .078 | 2.54 |
| Chloroflexus group | .053 | 3.10 |
| Chloroflexus/Thermomicrobium split | .066 | 2.81 |
| Deinococcus group | .064 | 2.85 |
| Deinococcus/Meiothermus split | .085 | 2.38 |
| Bifidobacterium/Arthrobacter/Streptomyces group | .068 | 2.76 |
| Bifidobacterium/Arthrobacter/Streptomyces radiation | .108 | 1.87 |
| Chlamydia/Pirellula group | .071 | 2.70 |
| Chlamydia/Pirellula split | .074 | 2.63 |
| Fibrobacter/Treponema group | .074 | 2.63 |
| Chlorobium/Cytophaga/Flexibacter group | .080 | 2.50 |
| Campylobacter/Helicobacter ($\varepsilon$) split | .130 | 1.37 |
| Cytophaga/Flexibacter split | .110 | 1.82 |
| Fusobacterium group | .085 | 2.38 |
| Mycoplasma group | .099 | 2.07 |
| Bacillus/Lactobacillus group | .111 | 1.80 |
| Clostridium group | .087 | 2.34 |
| Cyanobacteria/Chloroplast group | .073 | 2.65 |
| Proteobacteria | .078 | 2.54 |
| Desulfovibrio ($\delta$) group | .085 | 2.38 |
| Escherichia/Allochromatium ($\gamma$) group | .107 | 1.89 |
| Neisseria/Stenotrophomonas ($\beta$) group | .105 | 1.94 |
| Rickettsia/Rhizobium ($\alpha$) group | .102 | 2.00 |
| Root for the $\alpha$, $\beta$, and $\gamma$ groups | .086 | 2.36 |

Crenarchaeota radiating at 3.01 Ga and the Euryarchaeota radiating earlier, at 3.33 Ga. The halophiles diverged from the methanogens relatively recently (2.74 Ga), which is also around the time the cyanobacterial, Gram-positive, and Proteobacterial lineages arose in the Bacterial Domain (Table 2).

The Bacterial domain's first divergence occurred at least 3.46 Ga (Table 2), leading to the *Aquifex* lineage. Another divergence at 3.30 Ga led to the *Thermotoga* group, whereas the *Chloroflexus* and *Deinococcus* groups split at 3.10 and 2.85 Ga, respectively. The *Arthrobacter* relatives diverged at 2.76 Ga, and the *Chlamydia* and *Spirochete* groups split from the rest of the Bacteria shortly after this, diverging from each other at 2.70 Ga. Interestingly, a number of radiations occurred at about the time the cyanobacterial lineage diverged (2.65 Ga), which would also be consistent with the onset of the rise of molecular oxygen in the atmosphere (Canfield 1998; Rye and Holland 1998; Canfield and Raiswell 1999; Farquhar et al. 2000). The *Chlorobium*/*Flexibacter*/*Cytophaga* group diverged at about the same time, while the divergences of the Proteobacteria (2.54 Ga) and the Fusobacterium Group (2.38 Ga), which contains the *Bacillus*/*Clostridium*/*Mycoplasma* group, occurred a few tens to hundreds of millions years later, which may be possibly correlated to further increases in atmospheric $O_2$. Although the $\alpha$, $\beta$, $\gamma$, and $\delta$ Proteobacteria form a clade on our

tree, the $\varepsilon$ Proteobacteria appear to be more closely related to *Chlorobium limicola*, a green sulfur bacterium. From our tree, we are unable to propose with confidence an identity for the filamentous prokaryotic fossils found in the 3.5 Ga rocks (Schopf 1993). Filamentous microfossils believed to be the remnants of chemoautotrophic thermophilic prokaryotes that lived in subseafloor hydrothermal vents 3.235 Ga (Rasmussen 2000) have recently been reported. The discovery of these fossils in addition to the lack of evidence for oxygenic photosynthesis before 3 Ga make it unlikely that the Schopf microfossils are cyanobacteria. Although our proposed divergence dates are minimum estimates and actual divergences probably took place earlier, it is unlikely that cyanobacteria diverged almost a billion years before leaving any evidence in the geologic record. Additional geochemical analyses are required to establish the first occurrence and geographic range of cyanobacterial biomarkers during the Archean.

Bootstrap values determined from 10,000 replicates (Figure 2) indicate that most of the major divergences reported here are well supported. The earliest diverging lineages in the Bacterial Domain have high bootstrap values. Most of the major groupings (the cyanobacteria, the Proteobacterial subgroups, the *Arthrobacter* group, the *Flexibacter* group, the *Treponema* group, and the *Bacillus*/*Clostridium* group) are supported by high bootstrap values, although the divergence points of these groups are not well supported. This may be indicative of a massive radiation, which occurred around the time of the cyanobacterial divergence, and may be associated with the appearance of significant amounts of molecular oxygen in the environment. A similar case is seen in the Archaeal Domain, with support for the major groupings (the *Archaeoglobus* group, the *Methanosarcina*/*Haloferax* group, and the *Sulfolobus*/*Thermoproteus* group), but not for their divergence points. It is possible that the radiation of the Archaea that also occurs $\sim$2.65 Ga may be due to the increasing presence of oxygen due to cyanobacterial metabolism.

The original Tree of Life based on SSU rRNA data was unrooted (since all sequences were related, there could be no outgroup to root the tree), and it was by using paralogous genes (the result of gene duplication events which would have occurred before the LCA split into separate lineages) that the LCA was placed between the Bacteria and the Archaea (Gogarten et al. 1989; Iwabe et al. 1989; Brown and Doolittle 1995). The overlaying of phenotypic data onto Woese's original phylogenetic tree indicated that the deepest branching organisms were most likely prokaryotic, anaerobic, and thermophilic (Kandler 1994). Extrapolation of this phenotype to the Last Common Ancestor is consistent with the evidence for high temperatures and high levels of $CO_2$ in the early Archean atmosphere (Kasting 1993; Kasting et al. 1993).

We wished to determine the timing of the LCA of the unrooted tree, using our original data set merged with a data set of 28 representative Eukaryotic SSU rRNA sequences (while keeping in mind the vast differences between the selective pressures operational in the prokaryotic and eukaryotic domains). Our calculations give estimates for the LCA of all three domains of 4.29 Ga when using the cyanobacterial calibration. By comparison, estimates for the LCA of all three domains based on the Reptile-Mammal split were 3.62 Ga, 3.77 Ga, and 5.02 Ga, using fossil record times for the synapsid-diapsid divergence of 288 Ma, 300 Ma, and 400 Ma, respectively (data not shown). While most previous calibrations have used a mammal-bird divergence time estimate from the fossil record of 300 Ma, recent criticism of this date (Lee 1999) has led to the proposal of 288 Ma as the minimum unambiguous fossil record evidence for this divergence.

Our estimated times of divergence for the Bacterial-Archaeal LCA and the Bacterial-Archaeal-Eukaryote LCA are quite similar (4.29 Ga). This very early date is consistent with recent evidence for a habitable planet surface after 4.4 Ga (Nisbet and Sleep 2001; Wilde et al. 2001). This ancestor may reflect a "primitive" (i.e., more prokaryotic-like)

nature of the very early Eukaryotes, which were likely unicellular and thermophilic. The genesis of the Eukaryotic lineage has long been problematic (Gupta 1998; Lopez-Garcia and Moreira 1999; Vellai and Vida 1999). Competing theories run the gamut from the Ur-Karyote's formation due to fusion (Karlin et al. 1999), phagocytosis (Lake and Rivera 1994; Gupta and Golding 1996), or syntrophy (Martin and Muller 1998; Moreira and Lopez-Garcia 1998) between members of the Bacterial and Archaeal domains after they were well established to the simultaneous formation of the Eukaryotic domain alongside the Bacterial and Archaeal domains (Kandler 1994; Woese 1998). The presence of a complex distribution of steranes and methylated steranes are observed in the same Late Archean shales in which the 2-methylhopanes were found (Eigenbrode et al., unpublished observations; Brocks et al. 1999). These compounds indicate that the biochemistry of the Eukaryotic domain was established at least by this date, although we note that the presence of these outer membrane lipids does not require the presence of mitochondria at this time. The chimeric nature of the Eukaryotic genome can be explained by an early symbiotic event, massive gene transfers after distinct domains had become established, or promiscuous horizontal transfer of genetic material in an early community consisting of all three domains before increasing cellular integrity decreased the frequency of such events. Our time estimates of the LCAs do not allow us to reject any of these hypotheses.

## Discussion

### *The Timing of Major Symbiotic Events*

Our divergence time estimates allow certain speculations about the timing of important symbiotic events. A major endosymbiosis was the capture of an $\alpha$-Proteobacterium by an early Eukaryotic cell, which led to an obligate symbiosis with the bacterium becoming the organelle known as the mitochondrion. This event could not have taken place before the split between the $\alpha$-proteobacterial ancestor and the $\beta$-/$\gamma$-proteobacterial ancestor, which occurred 2.36 Ga. As mitochondrial SSU rRNA gene sequences are most closely related to those of the $\alpha$-Proteobacteria, a more accurate date for the origin of mitochondria might be the date of the radiation of the $\alpha$-Proteobacteria ($\sim$2.00 Ga).

A second major endosymbiotic event was the capture of a cyanobacterium by a Eukaryotic cell with the cyanobacterium eventually becoming the chloroplast. There is considerable debate as to how many endosymbiotic events are necessary to explain the origins of chloroplasts in plants, green algae, red algae, and brown and yellow algae. Chloroplasts could not have originated earlier than $\sim$1.90 Ga, if one uses the split between the free-living cyanobacteria and the chloroplasts as a minimum age of the origin of chloroplasts. The timing of endosymbiosis events between Eukaryotic cells and an $\alpha$-Proteobacterium and a cyanobacterium has been subject to debate. Comparison of rRNA trees with those generated from *cpn* gene sequences led Sogin (1997) to propose a date for the origins of mitochondria and chloroplasts at $\sim$850 Ma, which is $\sim$1 billion years later than the 1.9 Ga date from our calibration.

Although our results suggest the cyanobacterial lineage diverged before the $\alpha$-proteobacterial radiation, we propose that mitochondria arose in Eukaryotes before chloroplasts. Most early Eukaryotes are believed to have been anaerobes. With the increasing levels of toxic (to them) molecular oxygen resulting from the oxygenic photosynthesis of cyanobacteria, they would have found themselves increasingly marginalized. However, in a few tens of millions of years after the divergence of the cyanobacterial lineage, the Proteobacteria emerged capable of using molecular oxygen as a terminal electron acceptor.

The endosymbiotic capture of an $\alpha$-Proteobacterium by a Eukaryotic cell would then be favored not only by selective pressures due to the increased energy yield from an electron transport system able to use molecular oxygen, but also by its ability to remove molecular oxygen that is toxic to its anaerobic processes. Having solved the problem of molecular oxygen toxicity through endosymbiosis, and reaping the benefits of increased ATP from their new endosymbionts, Eukaryotic cells could then incorporate oxygen-producing photosynthetic endosymbionts at their leisure. The timing (Figure 2) of the $\alpha$-Proteobacterial radiation ($\sim$2.00 Ga) and the cyanobacterial-chloroplast divergence ($\sim$1.90 Ga) support this hypothesis.

### *Sulfur and Carbon Isotope Ratios*

The period around 2.7 to 2 Ga is marked by a significant rise in atmospheric oxygen as evidenced by a number of geochemical indicators, including sulfur isotopic data. These data suggest the onset of the rise from near-zero levels at 2.7 Ga, with sufficient oxygenation to support sulfate production (and subsequent activity of sulfate-reducing bacteria) by 2.3 to 2.2 Ga (Schidlowski et al. 1983; Thode and Goodwin 1983; Canfield and Raiswell 1999; Canfield et al. 2000). Consistent with this geochemical evidence, we note that our tree indicates a dramatic radiation of organisms capable of sulfate reduction within the $\delta$-Proteobacteria around 2.38 Ga. Interestingly, the sulfate reducers *Thermodesulfovibrio*, *Thermodesulfobacterium*, *Desulfothiovibrio*, and *Desulfotomaculum* were present much earlier, although we caution that this is not necessarily evidence for significant sulfate reduction at 3 Ga, since this ability could have been acquired horizontally at a later time. Detailed sequence comparisons of the enzymes involved in sulfate reduction from all these organisms could identify the taxa in which this pathway originated. Although a recent report (Shen et al. 2001) indicated evidence for sulfate reduction at 3.47 Ga, what was actually measured was sulfide production in these samples. Genera able to produce sulfide from sulfur anaerobically are widely distributed throughout the Crenarchaeotal and Euryarchaeotal kingdoms of the Archaeal Domain. Therefore, it is not surprising to find evidence of sulfide production at 3.47 Ga, and as this ability is found extensively in both Archaeal kingdoms, it likely predates their divergence at 3.46 Ga.

An anomalous set of low $\delta$-$^{13}$C values for Archean sedimentary organic matter is recorded at around 2.7 Ga. The widespread occurrence of this signature has led to the suggestion that it records a global ecosystem that included significant recycling of methane carbon into preserved organic matter (Hayes 1983, 1994). The signature potentially provides the earliest record of aerobic metabolism because all identified methane-consuming bacteria require $O_2$. However, methanotrophic bacteria fall within the $\alpha$- and $\gamma$-Proteobacteria, which diverged from the $\delta$-Proteobacteria by 2.54 Ga. Determining whether any $\delta$- or $\beta$-Proteobacteria are capable of methane oxidation or contain cryptic copies of the genes encoding this pathway will give a more precise date as to when this ability arose. Current genome sequencing efforts will help in this regard. As the divergence date of the known methane oxidizing bacteria (2.54 Ga) is significantly later than the date (2.7 Ga) determined by the isotopic record, we suggest another mechanism such as anaerobic methane recycling by Archaea (Hinrich et al. 1999) may be responsible for the low isotopic signature. Organisms able to anaerobically utilize methane as yet have not been isolated, although preliminary SSU rRNA results indicate they may be Archaea (Hinrich et al. 1999; Thomsen et al. 2001). The unambiguous identification of these organisms may enable us to resolve the discrepancy between the divergence dates and the isotopic signature dates. The earliest date at which the 2.7 Ga $\delta$-$^{13}$C values can be due to aerobic methanotrophy is the latest date at which the cyanobacteria diverged, due to the need for significant amounts

of molecular oxygen. The discovery of cyanobacterial markers in rocks older than 2.7 Ga will resolve the discrepancy as to the origin of these low $\delta$-$^{13}$C values, and may necessitate a slight recalibration of our tree. Biomarker compounds for aerobic methanotrophs (3-methylhopanes) have not been found in late Archean sedimentary rocks (Brocks et al. 1999), although this may be due to limited sampling.

## Conclusion

Although the poor microbial fossil record from the early Archean does not permit definitive determination of either phylogeny or metabolism, geochemical methods enable minimum dates of signature metabolic processes to be established. The presence of 2-methylhopanoids in 2.65 Ga rock samples has enabled us to estimate divergence times for the major prokaryotic phyla in the Bacterial and Archaeal domains using a large representative sample of SSU rRNA sequences. We propose a date of 4.29 Ga for the Last Common Ancestor of the Bacterial and Archaeal Domains and a date of 3.46 Ga for the Bacterial and Archaeal Domain individual radiations. These divergence times do not conflict with geological evidence but call into question the interpretation of early Archean ($\sim$3.5 Ga) microfossils as evidence of cyanobacterial lineages. These divergence time estimates are undoubtedly minimum times for these events, and may be pushed backward as more molecular fossils are found in older geological formations. In fact, we hope that our use of the metabolic fossil data to increase our understanding of the interplay between life and the early Archean environment will lead to increased attempts by microbiologists and geologists to discover additional molecular fossils. Our distance tables could be useful for future calculations of events based on new discoveries and time calibrations.

## References

Arndt NT, Nelson DR, Compson W, Trendall AE, Thorne AM. 1991. The age of the Fortescue Group, Hamersley Basin, Western Australia, from ion microprobe zircon U-Pb results. Austral J Earth Sci 38:261–281.

Ayala FJ, Barrio E, Kwiatowski J. 1996. Molecular clock or erratic evolution? A tale of two genes. Proc Natl Acad Sci USA 93:11729–11734.

Ayala FJ, Rzhetsky A, Ayala FJ. 1998. Origin of the metazoan phyla: molecular clocks confirm paleontological estimates. Proc Natl Acad Sci USA 95:606–611.

Brocks JJ, Logan GA, Buick R, Summons RE. 1999. Archean molecular fossils and the early rise of eukaryotes. Science 285:1033–1036.

Brown JR, Doolittle WF. 1995. Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. Proc Natl Acad Sci USA 92:2441–2445.

Canfield DE. 1998. A new model for Proterozoic ocean chemistry. Nature 396:450–453.

Canfield DE, Habicht KS, Thamdrup B. 2000. The archean sulfur cycle and the early history of atmospheric oxygen. Science 288:658–661.

Canfield DE, Raiswell R. 1999. The evolution of the sulfur cycle. Am J Sci 299:697–723.

Delwiche CF, Palmer JD. 1996. Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. Mol Biol Evol 13:873–882.

Dennis PP, Ziesche S, Mylvaganam S. 1998. Transcription analysis of two disparate rRNA operons in the halophilic archaeon *Haloarcula marismortui*. J Bacteriol 180:4804–4813.

Doolittle RF, Feng D-F, Tsang S, Cho G, Little E. 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science 271:470–477.

Farquhar J, Bao HM, Thiemens M. 2000. Atmospheric influence of Earth's earliest sulfur cycle. Science 289:756–758.

Feng D-F, Cho G, Doolittle RF. 1997. Determining divergence times with a protein clock: update and reevaluation. Proc Natl Acad Sci USA 94:13028–13033.

Gogarten JP, Kibak H, Dittrich P, Taiz L, Bowman EJ, Bowman BJ, Manolson MF, Poole RJ, Date T, Oshima T, Konishi J, Denda K, Yoshida M. 1989. Evolution of the vacuolar $H^+$-ATPase: implications for the origin of eukaryotes. Proc Natl Acad Sci USA 86:6661–6665.

Gupta RS. 1998. Life's third domain (*Archaea*): an established fact or an endangered paradigm? Theoret Popul Biol 554:91–104.

Gupta RS, Golding GB. 1996. The origin of the eukaryotic cell. TIBS 21:166–171.

Hayes JM. 1983. Geochemical evidence bearing on the origin of aerobiosis, a speculative hypothesis. In: Schopf JW, editor. Earth's Earliest Biosphere, its Origin and Evolution. Princeton, NJ: Princeton University Press, p 291–301.

Hayes JM. 1994. Global methanotrophy at the Archean-Proterozoic transition. In: Bengtson S, editor. Early Life on Earth. New York: Columbia University Press, p 220–236.

Hinrich KU, Hayes JM, Sylva SP, Brewer PG, DeLong EF. 1999. Methane-consuming archaebacteria in marine sediments. Nature 398:802–805.

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 Natl Acad Sci USA 86:9355–9359.

Kandler O. 1994. The early diversification of life. In: Bengtson S, editor. Early Life on Earth. Nobel Symposium No. 84: New York: Columbia University Press, p 152–160.

Karlin S, Brocchieri L, Mrazek J, Campbell AM, Spormann AM. 1999. A chimeric prokaryotic ancestry of mitochondria and primitive eukaryotes. Proc Natl Acad Sci USA 96:9190–9195.

Kasting JF. 1993. Earth's early atmosphere. Science 259:920–926.

Kasting JF, Eggler DH, Raeburn SP. 1993. Mantle redox evolution and the oxidation state of the Archean atmosphere. J Geol 101:245–257.

Knoll AH. 1999. A new molecular window on early life. Science 285:1025–1026.

Lake JA, Rivera MC. 1994. Was the nucleus the first endosymbiont? Proc Natl Acad Sci USA 91:2880–2881.

Lee MSY. 1999. Molecular clock calibrations and metazoan divergence dates. J Mol Evol 49:385–391.

Lopez-Garcia P, Moreira D. 1999. Metabolic symbiosis at the origin of eukaryotes. TIBS 24:88–93.

Maidak BL, Cole JR, Lilburn TG, Parker Jr CT, Saxman PR, Stredwick JM, Garrity GM, Li B, Olsen GJ, Pramanik S, Schmidt TM, Tiedje JM. 2000. The RDP (Ribosomal Database Project) continues. Nucleic Acids Res 28:173–174.

Martin W, Muller M. 1998. The hydrogen hypothesis for the first eukaryote. Nature 392:37–41.

Moreira D, Lopez-Garcia P. 1998. Symbiosis between methanogenic archaea and $\delta$-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. J Mol Evol 47:517–530.

Nelson KE, Clayton RE, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, White O, Salzberg SL, Smith HO, Venter JC, Fraser CM. 1999. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritma*. Nature 399:323–329.

Nisbet EG, Sleep NH. 2001. The habitat and nature of early life. Nature 409:1083–1091.

Ochman H, Wilson AC. 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J Mol Evol 26:74–86.

Rasmussen B. 2000. Filamentous microfossils in a 3,235-million-year-old volcanogenic massive sulphide deposit. Nature 405:676–679.

Rye R, Holland HD. 1998. Paleosols and the evolution of atmospheric oxygen: a critical review. Amer J Sci 298:621–672.

Schidlowski M. 2001. Carbon isotopes as biogeochemical recorders of life over 3.8 Ga of Earth history: evolution of a concept. Precambrian Res 106:117–134.

Schidlowski M, Hayes JM, Kaplan IR. 1983. Isotopic inferences of ancient biochemistries: carbon, sulfur, hydrogen, and nitrogen. In: Schopf JW, editor. Earth's Earliest Biosphere, its Origin and Evolution. Princeton, NJ: Princeton University Press, p 149–186.

Schopf JW. 1993. Microfossils of the early archean apex chert: new evidence of the antiquity of life. Science 260:640–646.

Schopf JW. 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 JW, Packer BM. 1987. Early archean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona Group, Australia. Science 237:70–73.

Shen Y, Buick R, Canfield DM. 2001. Isotopic evidence for microbial sulphate reduction in the early Archaean era. Nature 410:77–81.

Sogin ML. 1997. History assignment: when was the mitochondrion founded? Curr Opin Genet Dev 7:792–799.

Summons RE, Jahnke LL, Hope JM, Logan GA. 1999. 2-methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. Nature 400:554–557.

Swofford DL. 2000. PAUP*. Phylogenetic analysis using parsimony (*and other methods), PAUP, Version 4.0b3a. Sinauer Associates, Sunderland, MA.

Takahashi K, Nei M. 2000. Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. Mol Biol Evol 17:1251–1258.

Thode HG, Goodwin AM. 1983. Further sulfur and carbon isotope studies of late archean iron-formations of the Canadian shield and the rise of sulfate reducing bacteria. Precambrian Res 20:337–356.

Thomsen TR, Finster K, Ramsing NB. 2001. Biogeochemical and molecular signatures of anaerobic methane oxidation in a marine sediment. Appl Environ Microbiol 67:1646–1656.

Trendall AE, Nelson DR, DeLaeter JR, Hassler SW. 1998. Precise zircon U-Pb ages from the Marra Mamba iron formation and Wittenoom Formation, Hamersley Group, Western Australia. Austral J Earth Sci 45:137–142.

Ueda K, Seki T, Kudo T, Yoshida T, Kataoka M. 1999. Two distinct mechanisms cause heterogeneity of 16S rRNA. J Bacteriol 181:78–82.

Vellai T, Vida G. 1999. The origin of eukaryotes: the difference between prokaryotic and eukaryotic cells. Proc Roy Soc Lond B 266:1571–1577.

Wang DY-C, Kumar S, Hedges SB. 1999. Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. Proc Roy Soc Lond B 266:163–171.

Wang Y, Zhang Z, Ramanan N. 1997. The actinomycete *Thermobispora bispora* contains two distinct types of transcriptionally active 16S rRNA genes. J Bacteriol 179:3270–3276.

Watanabe Y, Martini JE, Ohmoto H. 2000. Geochemical evidence for terrestrial ecosystems 2.6 billion years ago. Nature 408:574–578.

Wilde S, Valley JW, Peck WH, Graham CM. 2001. Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. Nature 409:175–178.

Woese CR. 1998. The universal ancestor. Proc Natl Acad Sci USA 95:6854–6859.

Woese CR, Kandler O, Wheelis ML. 1990. Towards a natural system of organisms: proposal for the domains archaea, bacteria, and eucarya. Proc Natl Acad Sci USA 87:4576–4579.

Wray GA, Levinton JS, Shapiro LH. 1996. Molecular evidence for deep precambrian divergences among metazoan phyla. Science 274:568–572.

Yap WH, Zhang Z, Wang Y. 1999. Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. J Bacteriol 181:5201–5209.