

Research article

Open Access

## A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land

Fabia U Battistuzzi<sup>1</sup>, Andreia Feijao<sup>2</sup> and S Blair Hedges\*<sup>1</sup>

Address: <sup>1</sup>NASA Astrobiology Institute and Department of Biology, 208 Mueller Laboratory, The Pennsylvania State University, University Park, PA 16802, USA and <sup>2</sup>European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

Email: Fabia U Battistuzzi - fxb142@psu.edu; Andreia Feijao - feijao@embl.de; S Blair Hedges\* - sbh1@psu.edu

\* Corresponding author

Published: 09 November 2004

Received: 07 August 2004

*BMC Evolutionary Biology* 2004, 4:44 doi:10.1186/1471-2148-4-44

Accepted: 09 November 2004

This article is available from: <http://www.biomedcentral.com/1471-2148/4/44>

© 2004 Battistuzzi et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** The timescale of prokaryote evolution has been difficult to reconstruct because of a limited fossil record and complexities associated with molecular clocks and deep divergences. However, the relatively large number of genome sequences currently available has provided a better opportunity to control for potential biases such as horizontal gene transfer and rate differences among lineages. We assembled a data set of sequences from 32 proteins (~7600 amino acids) common to 72 species and estimated phylogenetic relationships and divergence times with a local clock method.

**Results:** Our phylogenetic results support most of the currently recognized higher-level groupings of prokaryotes. Of particular interest is a well-supported group of three major lineages of eubacteria (Actinobacteria, *Deinococcus*, and Cyanobacteria) that we call Terrabacteria and associate with an early colonization of land. Divergence time estimates for the major groups of eubacteria are between 2.5–3.2 billion years ago (Ga) while those for archaeobacteria are mostly between 3.1–4.1 Ga. The time estimates suggest a Hadean origin of life (prior to 4.1 Ga), an early origin of methanogenesis (3.8–4.1 Ga), an origin of anaerobic methanotrophy after 3.1 Ga, an origin of phototrophy prior to 3.2 Ga, an early colonization of land 2.8–3.1 Ga, and an origin of aerobic methanotrophy 2.5–2.8 Ga.

**Conclusions:** Our early time estimates for methanogenesis support the consideration of methane, in addition to carbon dioxide, as a greenhouse gas responsible for the early warming of the Earth's surface. Our divergence times for the origin of anaerobic methanotrophy are compatible with highly depleted carbon isotopic values found in rocks dated 2.8–2.6 Ga. An early origin of phototrophy is consistent with the earliest bacterial mats and structures identified as stromatolites, but a 2.6 Ga origin of cyanobacteria suggests that those Archean structures, if biologically produced, were made by anoxygenic photosynthesizers. The resistance to desiccation of Terrabacteria and their elaboration of photoprotective compounds suggests that the common ancestor of this group inhabited land. If true, then oxygenic photosynthesis may owe its origin to terrestrial adaptations.

## Background

The evolutionary history of prokaryotes includes both horizontal and vertical inheritance of genes [1-3]. Horizontal gene transfer (HGT) events are of great interest in themselves, for their roles in creating functionally new combinations of genes [4], but they pose problems for investigating the phylogenetic history and divergence times of organisms. The existence of a core of genes that has not been transferred is still under debate as HGTs have been detected in genes previously considered to be immune to these events [2,5-11]. Although a complete absence of HGT appears to be unlikely, genes belonging to different functional categories seem to be horizontally transferred with different frequencies [11-13]. Genes forming complex interactions with other cellular components (e.g. translational proteins) have a lower frequency of HGT and are generally more conserved among organisms. Recent studies based on analyses of these genes have obtained similar phylogenies suggesting an underlying phylogenetic signal [3,14-17]. If we accept the use of core genes for phylogeny reconstruction then they should also be of use for time estimation with molecular clocks. Moreover the increasing number of prokaryotic genomes available has facilitated the detection of HGT through more accurate detection of orthology, paralogy, and monophyletic groups, and the concatenation of gene and protein sequences has helped increase the confidence of nodes and decrease the variance of time estimates [14,16,18,19].

Temporal information concerning prokaryote evolution has come from diverse sources. For eukaryotes, the fossil record provides an abundant source of such data, but this has not been true for prokaryotes, which are difficult to identify as fossils [20,21]. Limited information on specific groups or metabolites has been obtained from analyses of isotopic concentrations [22] and detection of biomarkers [23,24]. By making some simple assumptions – e.g., that aerobic organisms evolved after oxygen became available [25] – it is possible to constrain some nodes in the prokaryote timescale, but only in a coarse sense. However, most information on the timescale of prokaryote evolution has come from analysis of DNA and amino acid sequence data with molecular clocks [26-30]. The detection of evolutionary patterns in metabolic innovations, as a consequence of a phylogeny not dominated by HGT events, allows more detailed constraints on a prokaryote timescale.

In contrast to conventional interpretations of cyanobacteria as being among the most ancient of life forms on Earth [31], these studies have consistently found a late origin of cyanobacteria [28,30], nearly contemporaneous with the major Proterozoic rise in oxygen at 2.3 billion years ago (Ga), termed the Great Oxidation Event (GOE) [32].

In this study we have assembled a data set of amino acid sequences from 32 proteins common to 72 species of prokaryotes and eukaryotes and estimated phylogenetic relationships and divergence times with a local clock method. These results in turn have been used to investigate the origin of metabolic pathways of importance in evolution of the biosphere.

## Results

### Data set

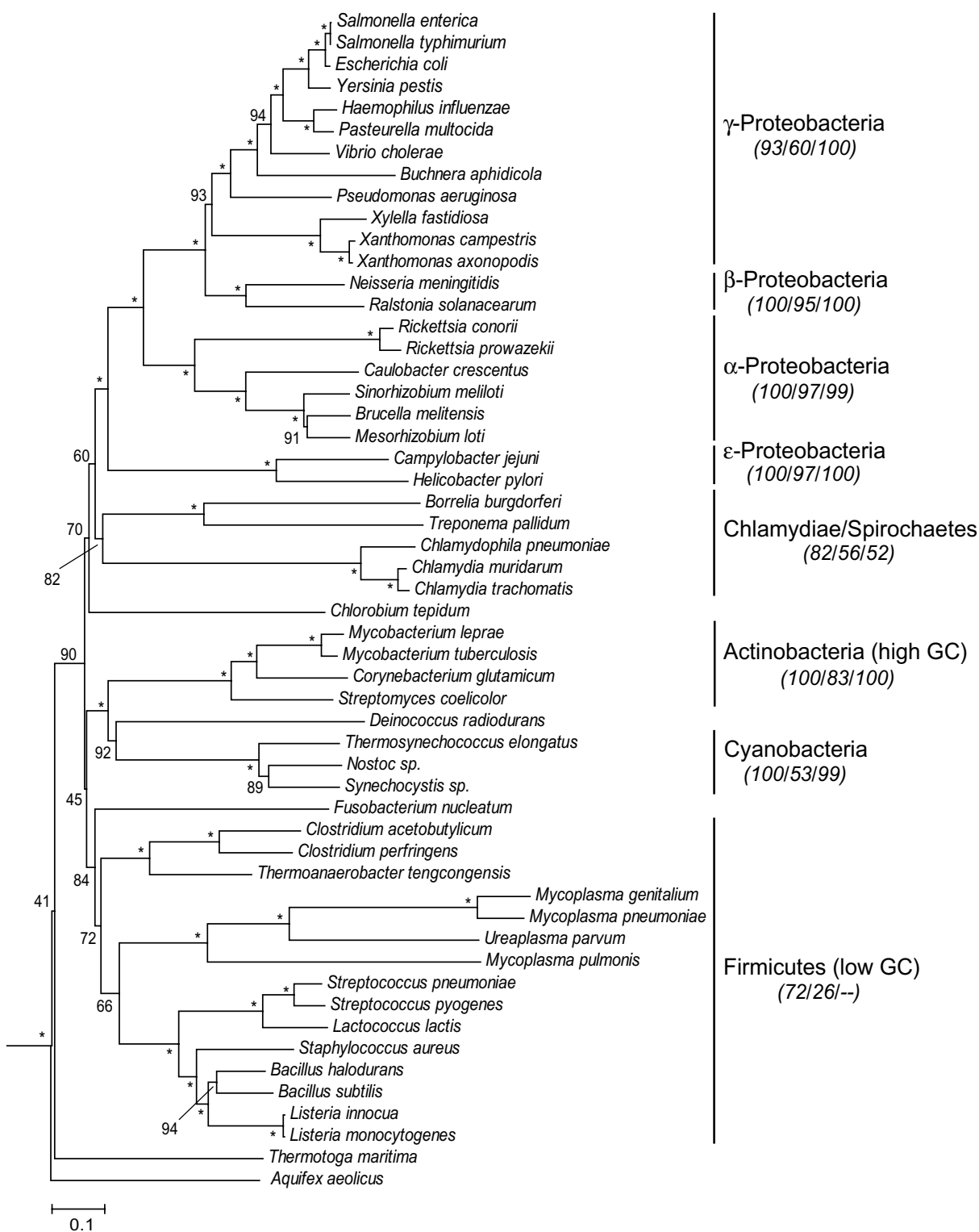
The majority (81%) of the 32 proteins that were used are classified in the "information storage and processes" functional category of the COG. The other categories represented are "cellular processes" (10%), "metabolism" (3%), and "information storage and processing" + "metabolism" (proteins with combined functions; 6%). Other studies that have analyzed prokaryote genome sequence data for phylogeny have found a similar high proportion of proteins in the "information storage and processes" functional category, presumably because HGT is more difficult with such genes that are vital for the survival of the cell [3,18,33,34].

The concatenated and aligned data set of 32 proteins contained 27,205 amino acid sites (including insertions and deletions). With alignment gaps removed, the two data sets analyzed were 7,338 amino acid sites (Archaeobacteria) and 7,597 amino acid sites (Eubacteria). The data sets were complete in the sense that sequences of all taxa were present for all proteins.

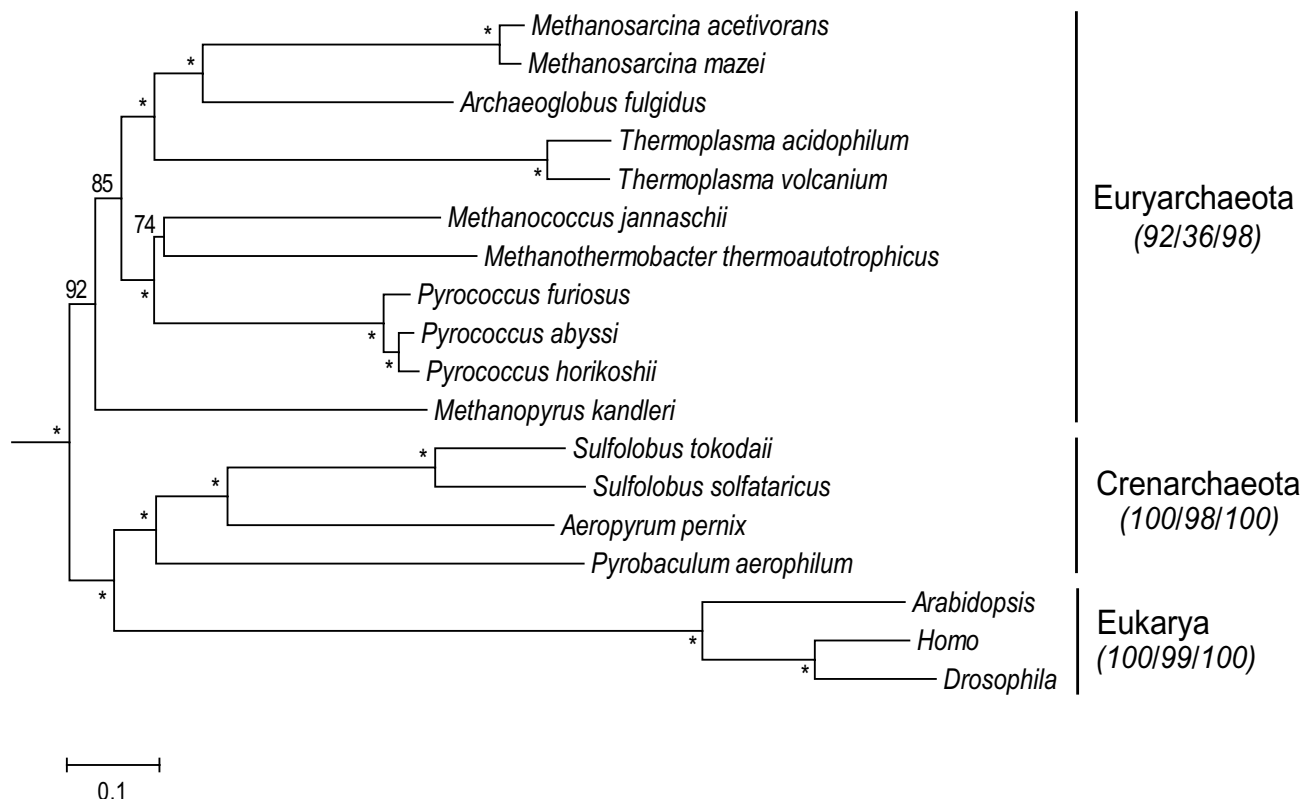
### Phylogeny

The phylogeny of eubacteria (Fig. 1) shows significant bootstrap support for most of the major groups and subgroups. All proteobacteria form a monophyletic group (support values 95/47/99 for ME, ML and Bayesian respectively) with the following relationships of the subgroups: (epsilon (alpha (beta, gamma))). There has been debate about the effect of base composition and substitution rate on the phylogenetic position of the endosymbiont *Buchnera* among  $\gamma$ -proteobacteria [35,36]. Its position here (Fig. 1) differs slightly from both studies; accordingly, any conclusions concerning its divergence time should be treated with caution. Spirochaetes cluster with Chlamydiae, Actinobacteria with Cyanobacteria and *Deinococcus* (support values for Cyanobacteria + *Deinococcus* are 92/80/99) and the hyperthermophiles (*Thermotoga*, *Aquifex*) branch basally in the tree. These groups and relationships are similar to those found previously with analyses of prokaryote genome sequences [3,18,33,34].

The phylogeny of archaeobacteria (Fig. 2) agrees with some but not all aspects of previous phylogenetic analyses of prokaryote genomes using sequence data [3,14,18,30,37,38] and the presence and absence of genes



**Figure 1**  
 Phylogenetic tree (ME;  $\alpha = 0.94$ ) of eubacteria rooted with archaeobacteria, using sequences of 32 proteins (7,597 amino acids). Bootstrap values are shown on nodes; asterisks indicate support values > 95%. For major groups, support values from three phylogenetic methods (ME/ML/Bayesian) are indicated in italics (dash indicates a group was not present).

**Figure 2**

Phylogenetic tree (ME;  $\alpha = 1.20$ ) of archaeobacteria rooted with eubacteria, using sequences of 32 proteins (7,338 amino acids). Bootstrap values are shown on nodes; asterisks indicate support values > 95%. For major groups, support values from three phylogenetic methods (ME/ML/Bayesian) are indicated in italics.

[37,39-41]. For example, each of the two major clades of Archaeobacteria (excluding Korarchaeota, which was not represented) is monophyletic. This is consistent with some analyses [14,18] but not others [3]. Also, the position of Crenarchaeota as closest relatives of eukaryotes (Fig. 2), instead of Euryarchaeota, has been debated [14,18,30,42,43]. The faster rate of evolution in eukaryotes (Fig. 2), as noted elsewhere [30,44], requires some caution in drawing conclusions regarding their phylogenetic position. Methanogens were found to be monophyletic in some previous analyses [3,41] but were paraphyletic in other analyses [38,45,46] and in our analysis (Fig. 2). The phylogenetic position of one species of methanogen in particular, *Methanopyrus kandleri*, has differed among previous studies [47-49]. However, it is difficult to make direct comparisons among various studies because they have included different sets of taxa.

#### Time estimation

Times of divergence were estimated for all nodes in the phylogenies of eubacteria (Fig. 1) and archaeobacteria (Fig.

2) using the alternative constraints (calibrations) described in the Methods. The eubacteria time estimates show an average 7% increase from the molecular to the geologic (2.3 Ga minimum) calibration point. Two other additional geologic calibration points were used in the analyses (see Methods), 2.3 Ga fixed and 2.7 Ga minimum, which showed respectively 10% younger and 11% older time estimates compared with the 2.3 Ga minimum calibration point.

The times estimated with the fossil calibration point in the archaeobacteria data set were on average only 10% younger than the ones estimated with the molecular calibration. Moreover there was even a smaller effect on the time estimates of the deepest nodes, which were the ones of interest in this study (node M 3.2%, node N 2.1%, node O 1.8% and node P 1.3%). This variation is due not only to the different calibration times but also to the type of constraints used (i.e. minimum boundaries only vs. minimum and maximum bounds).

A single timetree (Fig. 3) was constructed from the phylogenetic and divergence time data. The time estimates summarized in that tree derive only from the best-justified calibrations. For eubacteria, the 2.3 Ga minimum calibration (constraint), from the geologic record, was chosen because it encompasses all of the hypothesized time estimates for the origin of cyanobacteria. For archaeobacteria, the 1.2 Ga calibration (minimum 1.174 Ga, maximum 1.222 Ga), from the red algae fossil record, was selected because it provides a conservative constraint on the divergence of plants and animals. Time estimates and 95% credibility intervals for all nodes under all calibrations are presented elsewhere [see Additional file 1, Additional file 2, and Additional file 3], and those data are summarized for selected nodes and calibrations for eubacteria and archaeobacteria (Table 1). Although some undetected HGT could be a source of bias in the time estimates, the direction of the bias (raising or lowering the estimate) would depend on the specific node and groups involved, and it is unlikely to have had a major affect on the results, even if present.

Divergence times within eubacteria (Fig. 3, Table 1, nodes A-K) show a pattern seen previously [30] whereby most major groups diverge from one another (nodes B-I excluding node D) in a relatively limited time interval, approximately between 2.5–3.2 Ga. The position of the hyperthermophiles has been debated, with some studies showing them in a basal position whereas others place them more derived. The high G-C composition of these taxa is believed to be responsible for this difficulty in phylogenetic placement. Here, they branch basally (node J, 3.17–4.13 Ga and node K, 3.43–4.46 Ga), but this should be interpreted with caution for this reason. The divergence of *Escherichia coli* from *Salmonella typhimurium* (Fig. 3, Table 1, node A; 0.06–0.18 Ga) is consistent with the time estimated previously from consideration of mammalian host evolution (0.12–0.16 Ga) [26]. On the other hand an inconsistency with the fossil record is represented by the divergence of unicellular (*Thermosynechococcus elongatus*) and heterocyst-forming (*Nostoc sp.*) cyanobacteria. Our time estimate for this divergence is 0.70–1.41 Ga (Fig. 3, Table 1, node D) while microfossils of both groups have been identified in Mesoproterozoic (1.5–1.3 Ga) and Paleoproterozoic (2.12–2.02 Ga) rocks [50–52]. However the identification of these latter fossils has been debated [51]. Branch lengths of cyanobacteria in our protein tree and in 16S ribosomal RNA trees [34] do not suggest obvious substitutional biases or rate changes as they are neither unusually long nor unusually short. The reason for the discrepancy between the molecular and fossil times remains unclear but a possible misinterpretation of the fossil record cannot be dismissed.

Divergence times of most internal nodes among archaeobacteria (Fig. 3, Table 1, nodes L-P) are closely spaced in

time and relatively ancient, approximately between 3.1–4.1 Ga, regardless of the initial setting (prior) for the ingroup root. Node P is the earliest divergence, separating Euryarchaeota from Crenarchaeota+eukaryotes. Node O represents the common ancestor of the methanogens in our analysis (*Methanopyrus kandleri*, *Methanothermobacter thermoautotrophicus*, *Methanococcus jannaschii*, *Archaeoglobus fulgidus*, *Methanosarcina mazei* and *M. acetivorans*). Therefore, methanogenesis presumably arose between nodes P and O, or between 4.11 Ga (3.31–4.49 Ga) and 3.78 Ga (3.05–4.16 Ga) (Fig. 3, Table 1). If the position of *Methanopyrus kandleri* is not considered, in lieu of the current debate concerning its relationships (noted above), node N (Fig. 3, Table 1), the minimum time for the origin of methanogenesis drops only slightly, from 3.78 Ga (3.05–4.16 Ga) to 3.57 Ga (2.88–3.95 Ga).

## Discussion

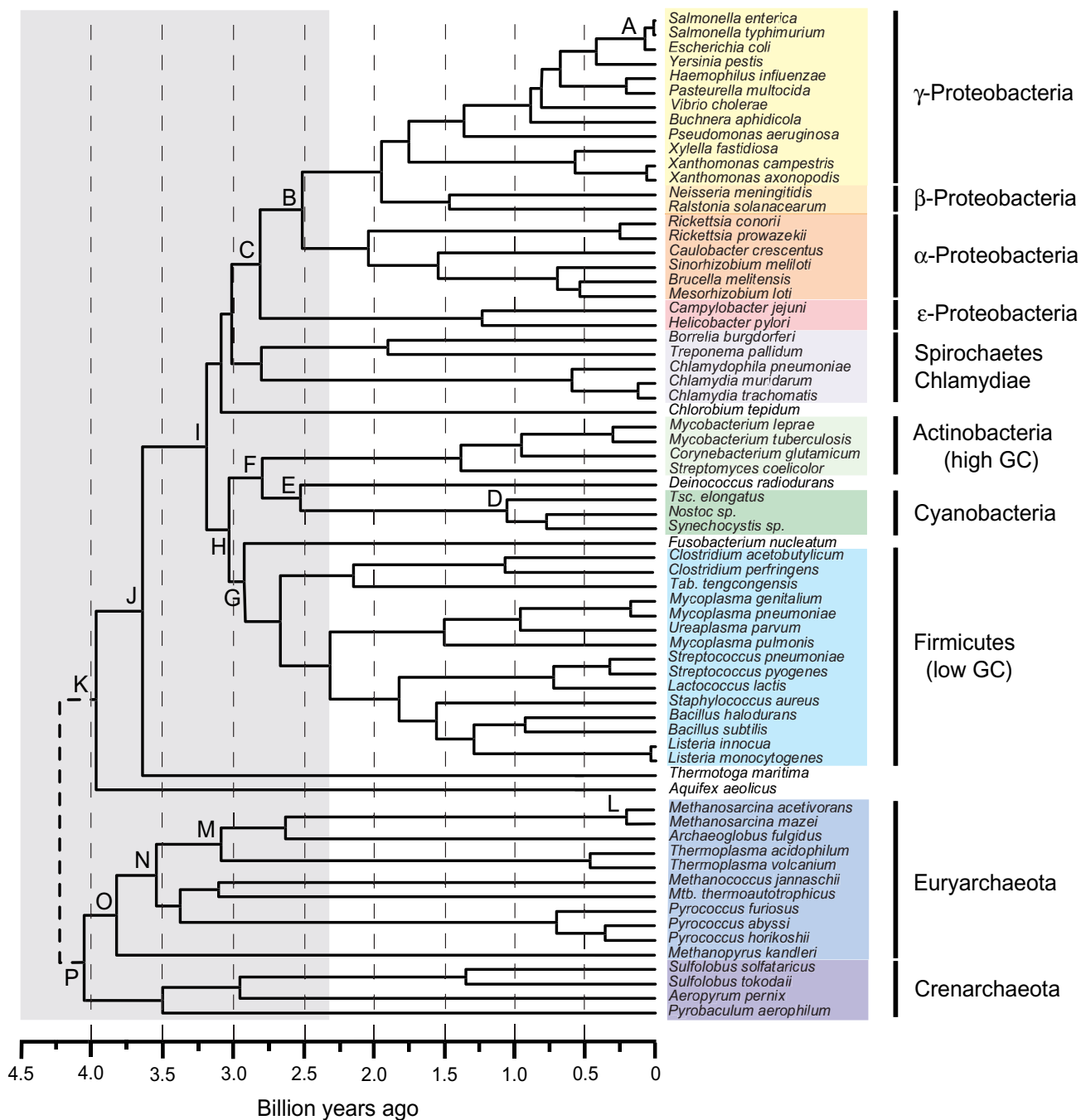
### Origin of life on Earth

Neither the time for the origin of life, nor the divergence of archaeobacteria and eubacteria, was estimated directly in this study. Nonetheless, one divergence within archaeobacteria was estimated to be as old as 4.11 Ga (Node P), suggesting even earlier dates for the last common ancestor of living organisms and the origin of life. This is in agreement with previous molecular clock analyses using mostly different data sets and methodology [28,30]. A Hadean (4.5–4.0 Ga) origin for life on Earth is also consistent with the early establishment of a hydrosphere [31,53]. Nevertheless, the earliest geologic and fossil evidence for life has been debated [21,54–59] leaving no direct support for such old time estimates.

### Methanogenesis

The lower luminosity of the sun during the Hadean and Archean predicts that surface water would have been frozen during that time. Instead there is evidence of liquid water and moderate to high surface temperatures [60,61]. The long term carbon cycle (carbonate-silicate cycle), which acts as a temperature buffer, combined with greenhouse gases, probably explain this "Faint Young Sun Paradox" [61]. Arguments have been made in support of either methane [62–64] or carbon dioxide [65] as the major greenhouse gas involved. If methane was important, it would have necessarily come from organisms (methanogens), given the volume required.

Archaeobacteria are the only prokaryotes known to produce methane. Our time estimate of between 4.11 Ga (3.31–4.49 Ga) and 3.78 Ga (3.05–4.16 Ga) for the origin of methanogenesis suggests that methanogens were present on Earth during the Archean, consistent with the methane greenhouse theory [64]. Nonetheless, this does not rule out the alternative (carbon dioxide) explanation [65].



**Figure 3**

A timescale of prokaryote evolution. Letters indicate nodes discussed in the text. The last common ancestor was arbitrarily placed at 4.25 Ga in the tree, although this placement was not part of the analyses. The grey rectangle shows the time prior to the initial rise in oxygen (presumably anaerobic conditions). Mtb: *Methanothermobacter*, Tab: *Thermoanaerobacter*, Tsc: *Thermosynechococcus*.

**Table 1: Time estimates for selected nodes in the tree of eubacteria (A-K) and archaeobacteria (L-P). Letters refer to Fig. 3.**

	Time (Ma) <sup>a</sup>	CI <sup>b</sup>
<b>Node A</b>	102	57–176
<b>Node B</b>	2508	2154–2928
<b>Node C</b>	2800	2452–3223
<b>Node D</b>	1039	702–1408
<b>Node E</b>	2558	2310–2969
<b>Node F</b>	2784	2490–3203
<b>Node G</b>	2923	2587–3352
<b>Node H</b>	3054	2697–3490
<b>Node I</b>	3186	2801–3634
<b>Node J</b>	3644	3172–4130
<b>Node K</b>	3977	3434–4464
<b>Node L</b>	233	118–386
<b>Node M</b>	3085	2469–3514
<b>Node N</b>	3566	2876–3948
<b>Node O</b>	3781	3047–4163
<b>Node P</b>	4112	3314–4486

<sup>a</sup> Averages of the divergence times estimated using the 2.3 Ga minimum constraint and the five ingroup root constraints (nodes A-K) and using the  $1.198 \pm 0.022$  Ga constraint and the five ingroup root constraints (nodes L-P).

<sup>b</sup> Credibility interval (minimum and maximum averages of the analyses under the five ingroup root constraints)

### Anaerobic methanotrophy

Anaerobic methanotrophy, or anaerobic oxidation of methane (AOM), is a metabolism associated with anoxic marine sediments rich in methane. This metabolism is characterized by the coupling of two reactions, oxidation of methane and sulfate reduction. The methane oxidizers are represented by archaeobacteria phylogenetically related to the Methanosarcinales, while the sulfate reducers, when present, are eubacterial members of the  $\delta$ -proteobacteria division [66]. These two groups of prokaryotes have been found associated in syntrophies, thus suggesting the coupling of these two pathways [66-69]. Archaeobacteria have been found also isolated in monospecific clusters, oxidizing methane through an unknown reaction. It has been suggested that they may use elements of both the methanogenesis and sulfate-reducing pathways [70]. An example of coexistence of genes from both of these pathways is *Archaeoglobus fulgidus*. The particular condition of this archaeobacterium has been explained with an ancient horizontal gene transfer from an eubacterial lineage, most likely a  $\delta$ -proteobacterium [71,72].

The phylogenetic position of the anaerobic methanotrophs with the Methanosarcinales places the maximum date for the origin of this metabolism at 3.09 (2.47–3.51) Ga (Fig. 3, Table 1, node M). The minimum time estimate

of 0.23 Ga (0.12–0.39 Ga) (Fig. 3, Table 1, node L), probably a substantial underestimate of the true time, results from the limited phylogenetic sampling available for this group.

### Aerobic methanotrophy

Aerobic methanotrophs are represented in the  $\alpha$  and  $\gamma$  divisions of the proteobacteria. This suggests an origin for this metabolism between node C (2.80 Ga; 2.45–3.22 Ga) and node B (2.51 Ga; 2.15–2.93 Ga) (Fig. 3, Table 1). Shared genes from this pathway and from methanogenesis also have been found in the Planctomycetales [73]. This has suggested a revision of the direction of the HGT, usually considered from archaeobacteria to eubacteria [1], that presumably has spread these genes in the two domains. However the absence of Planctomycetales from our dataset and its controversial phylogenetic position [74] does not allow us to discriminate among these possibilities.

Both anaerobic and aerobic methanotrophy have been used to explain the highly depleted carbon isotopic values found in 2.8–2.6 Ga geologic formations [22,75]. Our time estimates for these two metabolisms are both compatible with the isotopic record. Molecular clock methods have estimated the origin of cyanobacteria at 2.56 Ga (2.04–3.08 Ga) [30]. Because oxygenic photosynthesis would have been necessary for aerobic methanotrophy [75], an anaerobic metabolism seems more likely to explain the isotopic record.

### Phototrophy

The ability to utilize light as an energy source (phototrophy, photosynthesis) is restricted to eubacteria among prokaryotes. Phototrophic eubacteria are found in five major phyla (groups), including proteobacteria, green sulfur bacteria, green filamentous bacteria, gram positive heliobacteria, and cyanobacteria [4,76]. Only cyanobacteria produce oxygen.

There are three explanations for this broad taxonomic distribution of phototrophic metabolism; it evolved in one lineage of eubacteria and spread at a later time to other lineages by horizontal transfer, the common ancestor of these groups possessed this metabolism and genetic machinery, or there was a combination of horizontal transfer and vertical inheritance [4]. Because two of the three explanations require a phototrophic common ancestor, and because some features of the Archean geologic record require this metabolism if biologically produced [77], we have assumed here that the common ancestor (Node I) was phototrophic.

Therefore, we estimate that phototrophy evolved prior to 3.19 (2.80–3.63) Ga (Fig. 3, Table 1, node I). Because the

hyperthermophiles *Aquifex* and *Thermotoga* are not phototrophic and branch more basally, 3.64 (3.17–4.13) Ga (Node J) can be considered a maximum date for phototrophy. However, if those hyperthermophiles instead occupy a more derived position on the tree, as some analyses have indicated [33], then the maximum date is no longer constrained in this analysis.

### The colonization of land

The evolution of phototrophy was most likely linked to the evolution of other features essential to survival in stressful environments. Considerable biological damage can occur from exposure to ultraviolet radiation, especially prior to the GOE and later formation of the protective ozone layer [78]. The synthesis of pigments such as carotenoids, which function as photoprotective compounds against the reactive oxygen species created by UV radiation [79], is an ability present in all the photosynthetic eubacteria and in groups that are partly or mostly associated with terrestrial habitats such as the actinobacteria, cyanobacteria, and *Deinococcus-Thermus*.

Pigmentation was probably a fundamental step in the colonization of surface environments [80]. Besides the sharing of photoprotective compounds, these three groups (cyanobacteria, actinobacteria, and *Deinococcus*) also share a high resistance to dehydration [81–84], which further suggests that their common ancestor was adapted to land environments. Therefore we propose the name Terrabacteria (*L. terra*, land or earth) for the group that includes the bacterial phyla Actinobacteria, Cyanobacteria, and *Deinococcus-Thermus*. An early colonization of land is inferred to have occurred after the divergence of this terrestrial lineage with Firmicutes (Fig. 3, Table 1, node H), 3.05 (2.70–3.49) Ga, and prior to the divergence of Actinobacteria with Cyanobacteria + *Deinococcus* (Fig. 3, Table 1, node F), 2.78 (2.49–3.20) Ga. These molecular time estimates are compatible with time estimates (2.6–2.7 Ga) based on geological evidence for the earliest colonization of land by organisms (prokaryotes) [85]. Many groups of prokaryotes currently inhabit terrestrial environments, indicating that land has been colonized multiple times in different lineages.

### Oxygenic photosynthesis

From the above analyses and discussion, some of the early steps leading to oxygenic photosynthesis apparently were acquisition of protective pigments, phototrophy, and the colonization of land. Currently, hundreds of terrestrial species of cyanobacteria are known, broadly distributed among the orders, with species occurring in some of the driest environments on Earth. It is possible that a terrestrial ancestry of cyanobacteria, where stresses resulting from desiccation and solar radiation were severe, may have played a part in the evolution of oxygenic photosyn-

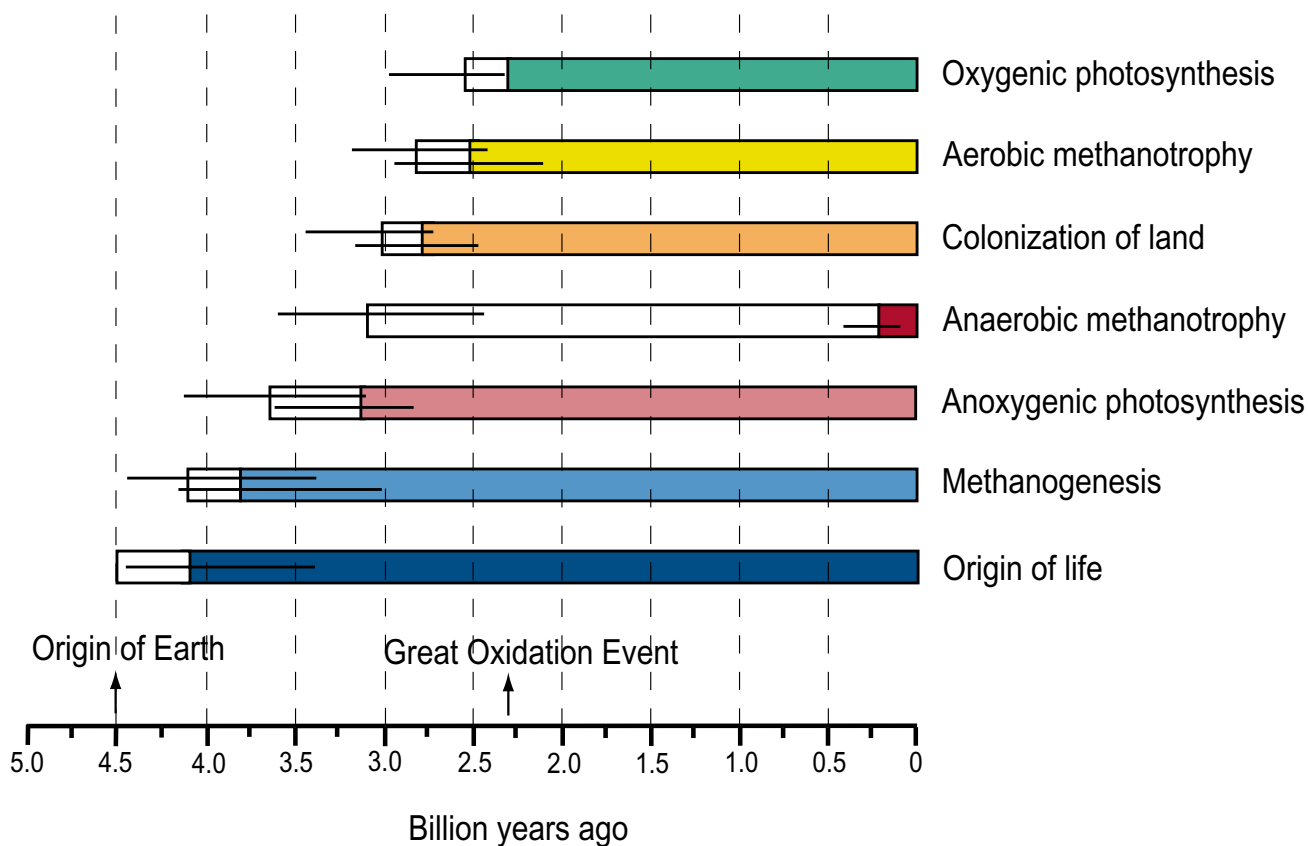
thesis. Nonetheless, there is ample evidence that horizontal gene transfer also has played an important role in the assembly of photosynthetic machinery [4].

Although we have used the origin of cyanobacteria as a calibration (2.3 Ga, geologic time based on GOE), such minimum constraints permit the estimated time to be much older in a Bayesian analysis. However, in this case, the time estimated for node E (2.56 Ga; 2.31–2.97 Ga; Fig. 3, Table 1) was not much older than the constraint itself. It also agrees with an earlier molecular time estimate (2.56 Ga; 2.04–3.08 Ga) based on a largely different data set and methods [30]. When we used the older minimum constraint of 2.7 Ga, corresponding to 2 $\alpha$ -methyl-hopane evidence considered to represent a biomarker of cyanobacteria [86], the estimated time was likewise only slightly older [see Additional file 1]. The oldest time estimates for oxygenic photosynthesis that we obtained are still considerable younger than has been assumed – generally – in the geologic literature [31,32,87]. This suggests that carbon isotope excursions, microfossils, microbial mats, stromatolites, and other pre-3 Ga evidence ascribed to cyanobacteria should be re-evaluated.

### Conclusions

The analyses presented here are based on the assumption, still under debate, that historical information (phylogenies and divergence times) can be retrieved from genes in the prokaryote genome that have not been affected by horizontal gene transfer. Our prokaryotic timeline shows deep divergences within both the eubacterial and archaeobacterial domains indicating a long evolutionary history. The early evolution of life (>4.1 Ga) and early origin of several important metabolic pathways (phototrophy, methanogenesis; but not oxygenic photosynthesis) suggests that organisms have influenced the Earth's environment since early in the history of the planet (Fig. 4). An inferred early presence of methanogens (3.8–4.1 Ga) is consistent with models suggesting that methane was important in keeping the Earth's surface warm in the Archean but does not rule out the possibility that carbon dioxide may have been equally (or more) important. In contrast to many classical interpretations of the early evolution of life, we find no compelling evidence for a pre-3 Ga evolution of cyanobacteria and oxygenic photosynthesis. This unique metabolism apparently evolved relatively late in the radiation of eubacterial clades, shortly before the Great Oxidation event (~2.3 Ga). The evolution of oxygenic photosynthesis may have involved a combination of adaptations to stressful terrestrial environments as well as acquisition of genes through horizontal transfer.





**Figure 4**  
 A time line of metabolic innovations and events on Earth. The minimum time for oxygenic photosynthesis is constrained by the Great Oxidation Event (2.3 Ga) whereas the maximum time for the origin of life is constrained by the origin of Earth (4.5 Ga). Horizontal lines indicate credibility intervals, white boxes indicate minimum and maximum time constraints on the origin of a metabolism or event, and colored boxes indicate the presence of the metabolism or event.

**Methods**

**Data assembly**

We assembled a dataset that maximized the number of taxa and proteins from available organisms with complete genome sequences of prokaryotes and selected eukaryotes. In doing so, we omitted a few taxa (e.g., *Agrobacterium tumefaciens Cereon str C58* and *Halobacterium sp. NRC-1*) whose addition to the data set would have resulted in a substantial reduction in the total number of proteins. Data assembly began with the Clusters of Orthologous Groups of Proteins (COG) [88], which consisted of 84 proteins common to 43 species. With that initial dataset we added other species from among completed microbial genomes (NCBI; National Center for Biotechnology Information), assisted by BLAST and PSI-BLAST [89]. In total 72 species were included in the study (54 eubacteria, 15 archaeobacteria and three eukaryotes).

The species of Archaeobacteria and their accession numbers are: *Aeropyrum pernix K1* (NC\_000854), *Archaeoglobus fulgidus* (NC\_000917), *Methanothermobacter thermoautotrophicus str. Delta H* (NC\_000916), *Methanococcus jannaschii* (NC\_000909), *Methanopyrus kandleri* AV19 (NC\_003551), *Methanosarcina acetivorans str. C2A* (NC\_003552), *Methanosarcina mazei Goe1* (NC\_003901), *Pyrobaculum aerophilum* (NC\_003364), *Pyrococcus abyssi* (NC\_000868), *Pyrococcus furiosus* DSM 3638 (NC\_003413), *Pyrococcus horikoshii* (NC\_000961), *Sulfolobus solfataricus* (NC\_002754), *Sulfolobus tokodaii* (NC\_003106), *Thermoplasma acidophilum* (NC\_002578), *Thermoplasma volcanium* (NC\_002689).

The species of Eubacteria are: *Aquifex aeolicus* (NC\_000918), *Bacillus halodurans* (NC\_002570), *Bacillus subtilis* (NC\_000964), *Borrelia burgdorferi* (NC\_001318), *Brucella melitensis* (NC\_003317, NC\_003318), *Buchnera*

*aphidicola* str. APS (*Acyrtosiphon pisum*) (NC\_002528), *Campylobacter jejuni* (NC\_002163), *Caulobacter crescentus* CB15 (NC\_002696), *Chlamydia muridarum* (NC\_002620), *Chlamydia trachomatis* (NC\_000117), *Chlamydomonas pneumoniae* CWL029 (NC\_000922), *Chlorobium tepidum* str. TLS (NC\_002932), *Clostridium acetobutylicum* (NC\_003030), *Clostridium perfringens* (NC\_003366), *Corynebacterium glutamicum* ATCC 13032 (NC\_003450), *Deinococcus radiodurans* (NC\_001263, NC\_001264), *Escherichia coli* O157:H7 EDL933 (NC\_002655), *Fusobacterium nucleatum* subsp. *nucleatum* ATCC 25586 (NC\_003454), *Haemophilus influenzae* Rd (NC\_000907), *Helicobacter pylori* 26695 (NC\_000915), *Lactococcus lactis* subsp. *lactis* (NC\_002662), *Listeria innocua* (NC\_003212), *Listeria monocytogenes* EGD-e (NC\_003210), *Mesorhizobium loti* (NC\_002678), *Mycobacterium leprae* (NC\_002677), *Mycobacterium tuberculosis* H37Rv (NC\_000962), *Mycoplasma genitalium* G-37 (NC\_000908), *Mycoplasma pneumoniae* (NC\_000912), *Mycoplasma pulmonis* (NC\_002771), *Neisseria meningitidis* MC58 (NC\_003112), *Nostoc* sp. PCC7120 (NC\_003272), *Pasteurella multocida* (NC\_002663), *Pseudomonas aeruginosa* PA01 (NC\_002516), *Ralstonia solanacearum* (NC\_003295), *Rickettsia conorii* (NC\_003103), *Rickettsia prowazekii* (NC\_000963), *Salmonella enterica* subsp. *enterica* serovar *Typhi* (NC\_003198), *Salmonella typhimurium* LT2 (NC\_003197), *Sinorhizobium meliloti* (NC\_003047), *Staphylococcus aureus* Mu50 (NC\_002758), *Streptococcus pneumoniae* TIGR4 (NC\_003028), *Streptococcus pyogenes* M1 GAS (NC\_002737), *Streptomyces coelicolor* A3(2) (NC\_003888), *Synechocystis* PCC6803 (NC\_000911), *Thermoanaerobacter tengcongensis* (NC\_003869), *Thermosynechococcus elongatus* BP-1 (NC\_004113), *Thermotoga maritima* (NC\_000853), *Treponema pallidum* subsp. *pallidum* str. *Nichols* (NC\_000919), *Ureaplasma parvum* serovar 3 str. ATCC 700970 (NC\_002162), *Vibrio cholerae* O1 biovar *eltor* str. N16961 (NC\_002505, NC\_002506), *Xanthomonas campestris* pv. *campestris* str. ATCC 33913 (NC\_003902), *Xanthomonas axonopodis* pv. *citri* str. 306 (NC\_003919), *Xylella fastidiosa* 9a5c (NC\_002488), *Yersinia pestis* (NC\_003143).

The eukaryotes were *Arabidopsis thaliana*, *Drosophila melanogaster*, *Homo sapiens*. Accession numbers for eukaryote proteins are presented elsewhere [90].

This dataset consisted of 60 proteins that were individually analysed as a step in orthology determination. The proteins were aligned with CLUSTALW [91]. Then phylogenetic trees of each protein were built and visually inspected. Initial trees were constructed using Minimum Evolution (ME), with MEGA version 2.1 [92]. The major criterion that we used in determining which genes to include or exclude was the monophyly of domains. We rejected genes with domains (archaeobacteria and eubacte-

ria) that were non-monophyletic, as these would be the best examples of HGT; this amounted to 61% of the genes rejected. Some other genes were omitted if there were detectable cases of HGT within a domain, such as the deep nesting of a species from one Phylum within a clade of another Phylum. Otherwise we did not eliminate genes that had a different branching order of phyla within a domain or different relationships of groups of lower taxonomic categories. Admittedly, ancient cases of HGT might be an explanation for some of those topological differences, but they are not detectable. However, we further tested the effectiveness of our criteria by examining the stability of individual protein trees, using different gamma values ( $\alpha = 1, 0.5$  and  $0.3$ ). We kept only the genes that were stable to such perturbations (in terms of remaining in that category of non-HGT genes). The position of eukaryotes, which varies depending on the gene, was not considered in assessing monophyly of eubacteria and archaeobacteria.

The 32 remaining proteins were concatenated for analysis. The  $\alpha$  parameters used during the tree building process were estimated with the program PamL (JTT+gamma model) [93]. From the concatenation, trees were constructed with ME, Maximum Likelihood (ML) [94] and Bayesian [95] methods. The phylogenies obtained with ME, ML and Bayesian were similar, differing only at non-significant nodes assessed by the bootstrap method [96], with one only significant exception on the position of *M. kandleri* in the Bayesian phylogeny. The sequence alignments and other supplementary data are presented elsewhere [90].

#### Time estimation

Time estimation was conducted separately within each domain (Archaeobacteria and Eubacteria) using reciprocal rooting and several calibration points. All time estimates were calculated with a Bayesian local clock approach [97] utilizing concatenated data sets of multiple proteins and a JTT+gamma model of substitution [19,98,99]. The following settings were used: numsamp (10,000), burnin (100,000), and sampfreq (100). This method permitted rates to vary on different branches, which was necessary given the known rate variation among prokaryote and eukaryote nuclear protein sequences [30,44]. Calibration of rate in this method was implemented by assigning constraints to nodes in the phylogeny. Five different initial settings (prior distributions) were used in each domain [see Additional file 4]. These were chosen at intervals of 0.5 Ga starting from 4.5 Ga, which is approximately the age of the Earth and Solar System, to 2.5 Ga, which is slightly before the major rise in oxygen (Great Oxidation Event; GOE) as recorded in the geologic record [32] and related to the presence of oxygenic cyanobacteria. Those constraints pertained to the ingroup root, or deepest

divergence in the tree excluding the outgroup. Because of the relatively small number of duplicate genes available for rooting the tree of life, we were unable to estimate the time of the last common ancestor (the divergence of eubacteria and archaeobacteria).

For the archaeobacterial data set, we included eukaryotes for calibration purposes because reliable calibration points were unavailable among those prokaryotes. In doing so, only proteins in which eukaryotes clustered with archaeobacteria were included [30]. An outgroup was used that consisted of representatives of the major groups of eubacteria [90]. We used the fossil and molecular times (separately) of the plant-animal divergence as calibration points, for comparison. The fossil calibration was the first appearance of a representative of the plant lineage (red algae) at  $1.198 \pm 0.022$  Ga [100]. The molecular time estimate for this divergence was  $1.609 \pm 0.060$  Ga from a study of 143 rate-constant proteins [98]. We used the minimum and maximum bounds for these calibration times as constraints in the Bayesian analysis. Although the results of these two different calibrations are provided for comparison, our preferred calibration is the 1.2 Ga fossil calibration because it has the best justification (supporting evidence). Therefore, our summary time estimates for archaeobacteria, presented in the timetree (Fig. 3), use only this fossil calibration.

For the eubacterial data set, we used four internal time constraints in separate analyses, all involving the origin of cyanobacteria. The first and most conservative constraint was a fixed origin (minimum and maximum bounds) at 2.3 Ga, which corresponds to the GOE. For the second constraint we used 2.3 Ga as a minimum bound, with no maximum bound. For the third constraint we used a previous molecular time estimate (2.56 Ga) for the divergence of cyanobacteria from closest living relatives among eubacteria, and fixed the minimum (2.04 Ga) and maximum (3.08 Ga) values to the 95% confidence limits of that time estimate [30]. The fourth constraint for the origin of cyanobacteria was set at 2.7 Ga (minimum constraint) based on biomarker evidence for the presence of 2 $\alpha$ -methylhopanes [86]. We did not consider the fossil record of cyanobacteria because the earliest indisputable fossils [52] are younger (2000 Ma) than the indirect evidence (GOE) for the presence of these oxygen-producing organisms. Older fossils of cyanobacteria are known but are disputed [52,101]. The use of these four alternative constraints for the origin of cyanobacteria considers most of the widely discussed hypotheses but does not rule out an origin prior to 2.7 Ga. Although the results of the four different calibrations are provided for comparison, our preferred calibration is the 2.3 (minimum) geologic calibration because it has the best justification (supporting evidence). Therefore, our summary time estimates for

eubacteria, presented in the timetree (Fig. 3), use only this geologic calibration.

For each of these calibration points, all five initial settings were applied, resulting in 15 and 20 analyses for the Archaeobacteria and Eubacteria (respectively). The effects of the different initial settings on the analyses were found to be minimal. A 44% difference in the priors, in fact, generated a maximum 2.7% (average of all significant nodes) difference in the time estimates (fossil calibration point) in the archaeobacteria and a maximum 3.5% (average of all significant nodes) difference in the eubacteria (molecular calibration point) [see Additional file 5].

### Authors' contributions

AF assembled and aligned the dataset and conducted initial analyses. FUB conducted phylogenetic and molecular clock analyses and co-drafted the manuscript. SBH directed the research and co-drafting the manuscript.

### Additional material

#### Additional File 1

*Complete time estimation analyses. Estimated times for each node and calibration for Eubacteria and Archaeobacteria. The node numbers refer to additional files 1 (eubacteria) and 2 (archaeobacteria).*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-4-44-S1.xls>]

#### Additional File 2

*Eubacteria tree. Phylogenetic tree of eubacteria (ME;  $\alpha = 0.94$ ). Node numbers assigned during the time estimation analyses are represented in italics.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-4-44-S2.pdf>]

#### Additional File 3

*Archaeobacteria tree. Phylogenetic tree of archaeobacteria (ME;  $\alpha = 1.20$ ). Node numbers assigned during the time estimation analyses are represented in italics.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-4-44-S3.pdf>]

#### Additional File 4

*Prior distribution values. Mean of the prior distribution for the rate of molecular evolution of the ingroup root node ( $r_{rate}$ ) in Eubacteria and Archaeobacteria.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-4-44-S4.doc>]

### Additional File 5

Percentage difference. Divergence time estimates and percentage difference due to different ingroup root constraints used under each calibration point. Node numbers refer to additional file 2 (eubacteria) and additional file 3 (archaeobacteria).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-4-44-S5.doc>]

### Acknowledgements

We thank Prachi Shah for programming assistance, Hidemi Watanabe for providing alignment tools, and Jaime E. Blair, Robert E. Blankenship, James G. Ferry, Davide Pisani and Fabienne Thomar for discussion. This work was supported by grants to SBH from the NASA Astrobiology Institute and the National Science Foundation. AF was supported by a Director's Travel Scholar grant from NASA Astrobiology Institute.

### References

- Boucher Y, Douady CJ, Papke RT, Walsh DA, Boudreau ME, Nesbo CL, Case RJ, Doolittle WF: **Lateral gene transfer and the origins of prokaryotic groups.** *Annu Rev Genet* 2003, **37**:283-328.
- Gogarten JP, Doolittle WF, Lawrence JG: **Prokaryotic evolution in light of gene transfer.** *Mol Biol Evol* 2002, **19**:2226-2238.
- Wolf YI, Rogozin IB, Grishin NV, Koonin EV: **Genome trees and the tree of life.** *Trends in Genetics* 2002, **18**:472-479.
- Raymond J, Zhaxybayeva O, Gogarten JP, Gerdes SY, Blankenship RE: **Whole-genome analysis of photosynthetic prokaryotes.** *Science* 2002, **298**:1616-1620.
- Nesbo CL, Boucher Y, Doolittle WF: **Defining the core of non-transferable prokaryotic genes: the euryarchaeal core.** *J Mol Evol* 2001, **53**:340-350.
- Olsen GJ, Woese CR, Overbeek R: **The winds of (evolutionary) change: breathing new life into microbiology.** *Journal of Bacteriology* 1994, **176**:1-6.
- Doolittle WF: **Phylogenetic classification and the universal tree.** *Science* 1999, **284**:2124-2128.
- Brochier C, Philippe H, Moreira D: **The evolutionary history of ribosomal protein RpS14: horizontal gene transfer at the heart of the ribosome.** *Trends Genet* 2000, **16**:529-533.
- Koonin EV: **Horizontal gene transfer: the path to maturity.** *Mol Microbiol* 2003, **50**:725-727.
- Philippe H, Douady CJ: **Horizontal gene transfer and phylogenetics.** *Curr Opin Microbiol* 2003, **6**:498-505.
- Lawrence JG, Hendrickson H: **Lateral gene transfer: when will adolescence end?** *Mol Microbiol* 2003, **50**:739-749.
- Jain R, Rivera MC, Lake JA: **Horizontal gene transfer among genomes: The complexity hypothesis.** *P Natl Acad Sci USA* 1999, **96**:3801-3806.
- Hansmann S, Martin W: **Phylogeny of 33 ribosomal and six other proteins encoded in an ancient gene cluster that is conserved across prokaryotic genomes: influence of excluding poorly alignable sites from analysis.** *Int J Syst Evol Microbiol* 2000, **50 Pt 4**:1655-1663.
- Brown JR, Douady CJ, Italia MJ, Marshall WE, Stanhope MJ: **Universal trees based on large combined protein data sets.** *Nature Genetics* 2001, **28**:281-285.
- Daubin V, Gouy M, Perriere G: **Bacterial molecular phylogeny using supertree approach.** *Genome Inform Ser Workshop Genome Inform* 2001, **12**:155-164.
- Brown JR: **Ancient horizontal gene transfer.** *Nat Rev Genet* 2003, **4**:121-132.
- Daubin V, Moran NA, Ochman H: **Phylogenetics and the cohesion of bacterial genomes.** *Science* 2003, **301**:829-832.
- Hedges SB: **The origin and evolution of model organisms.** *Nature Reviews Genetics* 2002, **3**:838-849.
- Hedges SB, Kumar S: **Genomic clocks and evolutionary timescales.** *Trends in Genetics* 2003, **19**:200-206.
- Benton MJ: **The Fossil Record 2.** 1st edition. London, Chapman and Hall; 1993:845.
- Altermann W, Kazmierczak J: **Archean microfossils: a reappraisal of early life on Earth.** *Res Microbiol* 2003, **154**:611-617.
- Hinrichs KU: **Microbial fixation of methane carbon at 2.7 Ga: Was an anaerobic mechanism possible?** *Geochim Geophys Geosy* 2002, **3**.
- Brocks JJ, Buick R, Summons RE, Logan GA: **A reconstruction of Archean biological diversity based on molecular fossils from the 2.78 to 2.45 billion-year-old Mount Bruce Supergroup, Hamersley Basin, Western Australia.** *Geochim Cosmochim Acta* 2003, **67**:4321-4335.
- Summons RE, Jahnke LL, Hope JM, Logan GA: **2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis.** *Nature* 1999, **400**:554-557.
- Blank CE: **Evolutionary timing of the origins of mesophilic sulphate reduction and oxygenic photosynthesis: a phylogenomic dating approach.** *Geobiology* 2004, **2**:1-20.
- Ochman H, Wilson AC: **Evolution in bacteria: evidence for a universal substitution rate in cellular genomes.** *Journal of Molecular Evolution* 1987, **26**:74-86.
- Doolittle RF, Feng DF, Tsang S, Cho G, Little E: **Determining divergence times of the major kingdoms of living organisms with a protein clock.** *Science* 1996, **271**:470-477.
- Feng DF, Cho G, Doolittle RF: **Determining divergence times with a protein clock: update and reevaluation.** *Proceedings of the National Academy of Sciences (USA)* 1997, **94**:13028-13033.
- Sheridan PP, Freeman KH, Brenchley JE: **Estimated minimal divergence times of the major bacterial and archaeal phyla.** *Geomicrobiol J* 2003, **20**:1-14.
- Hedges SB, Chen H, Kumar S, Wang DY, Thompson AS, Watanabe H: **A genomic timescale for the origin of eukaryotes.** *BMC Evolutionary Biology* 2001, **1**:4.
- Nisbet EG, Sleep NH: **The habitat and nature of early life.** *Nature* 2001, **409**:1083-1091.
- Holland HD: **Volcanic gases, black smokers, and the Great Oxidation Event.** *Geochimica et Cosmochimica Acta* 2002, **21**:3811-3826.
- Brochier C, Babetste E, Moreira D, Philippe H: **Eubacterial phylogeny based on translational apparatus proteins.** *Trends in Genetics* 2002, **18**:1-5.
- Jackson CR, Dugas SL: **Phylogenetic analysis of bacterial and archaeal arsC gene sequences suggests an ancient, common origin for arsenate reductase.** *BMC Evol Biol* 2003, **3**:18.
- Itoh T, Martin W, Nei M: **Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts.** *Proc Natl Acad Sci U S A* 2002, **99**:12944-12948.
- Canback B, Tamas I, Andersson SG: **A phylogenomic study of endosymbiotic bacteria.** *Mol Biol Evol* 2004, **21**:1110-1122.
- Wolf YI, Rogozin IB, Grishin NV, Tatusov RL, Koonin EV: **Genome trees constructed using five different approaches suggest new major bacterial clades.** *BMC Evolutionary Biology* 2001, **1**:8.
- Brochier C, Forterre P, Gribaldo S: **Archaeal phylogeny based on proteins of the transcription and translation machineries: tackling the Methanopyrus kandleri paradox.** *Genome Biol* 2004, **5**:R17.
- Snel B, Bork P, Huynen MA: **Genome phylogeny based on gene content.** *Nature Genetics* 1999, **21**:108-110.
- Tekaia F, Lazcano A, Dujon B: **The genomic tree as revealed from whole proteome comparisons.** *Genome Research* 1999, **9**:550-557.
- House CH, Runnegar B, Fitz-Gibbon ST: **Geobiological analysis using whole genome-based tree building applied to the Bacteria, Archaea and Eukarya.** *Geobiology* 2003, **1**:15-26.
- Rivera MC, Lake JA: **Evidence that eukaryotes and eocyte prokaryotes are immediate relatives.** *Science* 1992, **257**:74-76.
- Cammarano P, Creti R, Sanangelantoni AM, Palm P: **The Archaea monophyly issue: a phylogeny of translational elongation factor G(2) sequences inferred from an optimized selection of alignment positions.** *Journal of Molecular Evolution* 1999, **49**:524-537.
- Kollman JM, Doolittle RF: **Determining the relative rates of change for prokaryotic and eukaryotic proteins with anciently duplicated paralogs.** *Journal of Molecular Evolution* 2000, **51**:173-181.

45. Forterre P, Brochier C, Philippe H: **Evolution of the archaea.** *Theor Popul Biol* 2002, **61**:409-422.
46. Matte-Tailliez O, Brochier C, Forterre P, Philippe H: **Archaeal phylogeny based on ribosomal proteins.** *Molecular Biology and Evolution* 2002, **19**:631-639.
47. Burggraf S, Stetter KO, Rouviere P, Woese CR: **Methanopyrus-Kandleri - an Archaeal Methanogen Unrelated to All Other Known Methanogens.** *Syst Appl Microbiol* 1991, **14**:346-351.
48. Rivera MC, Lake JA: **The phylogeny of Methanopyrus kandleri.** *International Journal of Systematic Bacteriology* 1996, **46**:348-351.
49. Slesarev AI, Mezhevaya KV, Makarova KS, Polushin NN, Shcherbinina OV, Shakhova VV, Belova GI, Aravind L, Natale DA, Rogozin IB, Tatusov RL, Wolf YI, Stetter KO, Malykh AG, Koonin EV, Kozyavkin SA: **The complete genome of hyperthermophile Methanopyrus kandleri AV19 and monophyly of archaeal methanogens.** *Proceedings of the National Academy of Sciences USA* 2002, **99**:4644-4649.
50. Golubic S, Sergeev VN, Knoll AH: **Mesoproterozoic Archaeoelipsoidea: akinetes of heterocystous cyanobacteria.** *Lethaia* 1995, **28**:285-298.
51. Amard B, BertrandSarfati J: **Microfossils in 2000 Ma old cherty stromatolites of the Franceville Group, Gabon.** *Precambrian Res* 1997, **81**:197-221.
52. Sergeev VN, Gerasimenko LM, Zavarzin GA: **The Proterozoic history and present state of cyanobacteria.** *Microbiology* 2002, **71**:623-637.
53. Mojzsis SJ, Harrison TM, Pidgeon RT: **Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago.** *Nature* 2001, **409**:178-181.
54. Schopf JW: **Microfossils of the Early Archean Apex Chert - New Evidence of the Antiquity of Life.** *Science* 1993, **260**:640-646.
55. Mojzsis SJ, Arrhenius G, McKeegan KD, Harrison TM, Nutman AP, Friend CR: **Evidence for life on Earth before 3,800 million years ago.** *Nature* 1996, **384**:55-59.
56. Brasier MD, Green OR, Jephcoat AP, Kleppe AK, Van Kranendonk MJ, Lindsay JF, Steele A, Grassineau NV: **Questioning the evidence for Earth's oldest fossils.** *Nature* 2002, **416**:76-81.
57. Kazmierczak J, Altermann W: **Neoarchean biomineralization by benthic cyanobacteria.** *Science* 2002, **298**:2351-2351.
58. Schopf JW, Kudryavtsev, Agresti DG, Wdowiak TJ, Czaja AD: **Laser-Raman imagery of Earth's earliest fossils.** *Nature* 2002, **416**:73-76.
59. Brasier M, Green O, Lindsay J, Steele A: **Earth's oldest (similar to 3.5 Ga) fossils and the 'Early Eden hypothesis': Questioning the evidence.** *Origins Life Evol B* 2004, **34**:257-269.
60. Schwartzman DW: **Life, temperature, and the Earth.** New York, Columbia University Press; 1999:241.
61. Kasting JF, Catling D: **Evolution of a habitable planet.** *Annu Rev Astron Astr* 2003, **41**:429-463.
62. Pavlov AA, Kasting JF, Brown LL, Rages KA, Freedman R: **Greenhouse warming by CH<sub>4</sub> in the atmosphere of early Earth.** *J Geophys Res-Planet* 2000, **105**:11981-11990.
63. Kasting JF, Pavlov AA, Siefert JL: **A coupled ecosystem-climate model for predicting the methane concentration in the archaean atmosphere.** *Origins Life Evol B* 2001, **31**:271-285.
64. Pavlov AA, Hurtgen MT, Kasting JF, Arthur MA: **Methane-rich Proterozoic atmosphere?** *Geology* 2003, **31**:87-90.
65. Ohmoto H, Watanabe Y, Kumazawa K: **Evidence from massive siderite beds for a CO<sub>2</sub>-rich atmosphere before ~ 1.8 billion years ago.** *Nature* 2004, **429**:395-399.
66. Orphan VJ, Hinrichs KU, Ussler W, Paull CK, Taylor LT, Sylva SP, Hayes JM, DeLong EF: **Comparative analysis of methane-oxidizing archaea and sulfate-reducing bacteria in anoxic marine sediments.** *Applied and Environmental Microbiology* 2001, **67**:1922-1934.
67. Boetius A, Ravenschlag K, Schubert CJ, Rickert D, Widdel F, Gieseke A, Amann R, Jorgensen BB, Witte U, Pfannkuche O: **A marine microbial consortium apparently mediating anaerobic oxidation of methane.** *Nature* 2000, **407**:623-626.
68. DeLong EF: **Microbiology - Resolving a methane mystery.** *Nature* 2000, **407**:577-579.
69. Orphan VJ, House CH, Hinrichs KU, McKeegan KD, DeLong EF: **Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis.** *Science* 2001, **293**:484-487.
70. Orphan VJ, House CH, Hinrichs KU, McKeegan KD, DeLong EF: **Multiple archaeal groups mediate methane oxidation in anoxic cold seep sediments.** *Proceedings of the National Academy of Sciences (USA)* 2002, **99**:7663-7668.
71. Klenk HP, Clayton RA, Tomb JF, White O, Nelson KE, Ketchum KA, Dodson RJ, Gwinn M, Hickey EK, Peterson JD, Richardson DL, Kerlavage AR, Graham DE, Kyrpides NC, Fleischmann RD, Quackenbush J, Lee NH, Sutton GG, Gill S, Kirkness EF, Dougherty BA, McKenney K, Adams MD, Loftus B, Peterson S, Reich CI, McNeil LK, Badger JH, Glodek A, Zhou LX, Overbeek R, Gocayne JD, Weidman JF, McDonald L, Utterback T, Cotton MD, Spriggs T, Artiach P, Kaine BP, Sykes SM, Sadow PW, D'Andrea KP, Bowman C, Fujii C, Garland SA, Mason TM, Olsen GJ, Fraser CM, Smith HO, Woese CR, Venter JC: **The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus.** *Nature* 1997, **390**:364-370.
72. Klein M, Friedrich M, Roger AJ, Hugenholtz P, Fishbain S, Abicht H, Blackall LL, Stahl DA, Wagner M: **Multiple lateral transfers of dissimilatory sulfite reductase genes between major lineages of sulfate-reducing prokaryotes.** *J Bacteriol* 2001, **183**:6028-6035.
73. Chistoserdova L, Jenkins C, Kalyuzhnaya M, Marx CJ, Lapidus A, Vorholt JA, Staley JT, Lidstrom ME: **The enigmatic Planctomycetes may hold a key to the origins of methanogenesis and methylotrophy.** *Mol Biol Evol* 2004, **21**:1234-1241.
74. Jenkins C, Fuerst JA: **Phylogenetic analysis of evolutionary relationships of the planctomycete division of the domain bacteria based on amino acid sequences of elongation factor Tu.** *J Mol Evol* 2001, **52**:405-418.
75. Hayes JM: **Global methanotrophy at the Archean-Proterozoic transition.** In *Early life on Earth* Edited by: S. B. New York, Columbia University Press; 1994:220-236.
76. Xiong J, Fischer WM, Inoue K, Nakahara M, Bauer CE: **Molecular evidence for the early evolution of photosynthesis.** *Science* 2000, **289**:1724-1730.
77. DesMarais DJ: **When did photosynthesis emerge on Earth?** *Science* 2000, **289**:1703-1705.
78. Cockell CS, Horneck G: **The history of the UV radiation climate of the earth - Theoretical and space-based observations.** *Photochem Photobiol* 2001, **73**:447-451.
79. Gotz T, Windhovel U, Boger P, Sandmann G: **Protection of photosynthesis against ultraviolet-B radiation by carotenoids in transformants of the cyanobacterium Synechococcus PCC7942.** *Plant Physiol* 1999, **120**:599-604.
80. Wynn-Williams DD, Edwards HG, Newton EM, Holder JM: **Pigmentation as a survival strategy for ancient and modern photosynthetic microbes under high ultraviolet stress on planetary surfaces.** *International journal of Astrobiology* 2002, **1**:39-49.
81. Potts M: **Desiccation tolerance of prokaryotes.** *Microbiol Rev* 1994, **58**:755-805.
82. Mattimore V, Battista JR: **Radioresistance of Deinococcus radiodurans: Functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation.** *J Bacteriol* 1996, **178**:633-637.
83. Rokitko PV, Romanovskaya VA, Malashenko YR, Chernaya NA, Gushcha NI, Mikheev AN: **Soil drying as a model for the action of stress factors on natural bacterial populations.** *Microbiology* 2001, **72**:756-761.
84. Shirley B, McMaster NJ, Smith SC, Wright DJ, Rodriguez H, Jaruga P, Birincioglu M, Helm RF, Potts M: **Genomic DNA of Nostoc commune (Cyanobacteria) becomes covalently modified during long-term (decades) desiccation but is protected from oxidative damage and degradation.** *Nucleic Acids Res* 2003, **31**:2995-3005.
85. Watanabe Y, Martini JE, Ohmoto H: **Geochemical evidence for terrestrial ecosystems 2.6 billion years ago.** *Nature* 2000, **408**:574-578.
86. Summons RE, Jahnke LL, Hope JM, Logan GA: **2-methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis.** *Nature* 1999, **400**:554-557.
87. Buick R: **The antiquity of oxygenic photosynthesis: evidence from stromatolites in sulphate deficient Archaean Lakes.** *Science* 1992, **255**:74-77.
88. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV: **The COG database: new developments in phylogenetic classifica-**

- tion of proteins from complete genomes. *Nucleic Acids Research* 2001, **29**:22-28.
89. Altschul SF, Madden TL, Schaeffer AA, Zhang J, Zhang Z, Miller W, Lipman D: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Research* 1997, **25**:3389-3402.
  90. genomics E: <http://www.evogenomics.org/publications/data/prokaryotes>.
  91. Thompson JD, Higgins DG, Gibson TJ: **CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Research* 1994, **22**:4673-4680.
  92. Kumar S, Tamura K, Jakobsen IB, Nei M: **MEGA2: molecular evolutionary genetics analysis software.** *Bioinformatics* 2001, **17**:1244-1245.
  93. Yang Z: **PAML: a program package for phylogenetic analysis by maximum likelihood.** *CABIOS* 1997, **13**:555-556.
  94. Strimmer K, vonHaeseler A: **Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies.** *Molecular Biology and Evolution* 1996, **13**:964-969.
  95. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**:1572-1574.
  96. Felsenstein J: **Confidence limits on phylogenies: an approach using the bootstrap.** *Evolution* 1985, **39**:783-791.
  97. Thorne JL, Kishino H, Painter IS: **Estimating the rate of evolution of the rate of molecular evolution.** *Molecular Biology and Evolution* 1998, **15**:1647-1657.
  98. Hedges SB, Blair JE, Venturi ML, Shoe JL: **A molecular timescale of eukaryote evolution and the rise of complex multicellular life.** *BMC Evol Biol* 2004, **4**:2.
  99. Hedges SB, Kumar S: **Precision of molecular time estimates.** *Trends Genet* 2004, **20**:242-247.
  100. Butterfield NJ: **Bangiomorpha pubescens n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes.** *Paleobiology* 2000, **26**:386-404.
  101. Knoll AH: **The geobiological consequences of evolution.** *Geobiology* 2003, **1**:3-14.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

