MBE Advance Access published November 10, 2008

A major clade of prokaryotes with ancient adaptations to life on land

Fabia U. Battistuzzi[1,2] and S. Blair Hedges[1*]

[1]Department of Biology, Pennsylvania State University, University Park, PA 16802-5301 USA;

[2]Current address: Center for Evolutionary Functional Genomics, The Biodesign Institute, Arizona State University, Tempe, AZ 85287-5301, USA.

**Author for correspondence: sbh1@psu.edu*

Intended as a Research Article

Address for correspondence:
Dr. S. Blair Hedges
Department of Biology
Pennsylvania State University
University Park, PA 16802-5301, USA
Tel: (814) 865-9991
Fax: (814) 865-3125
E-mail: sbh1@psu.edu

Key words: Phylogeny, Bacteria, Archaea, Terrabacteria, Hydrobacteria, Molecular clock

Running head: Terrestrial adaptations in prokaryotes

Abbreviations: HGT, horizontal gene transfer; LSU, large subunit; ME, minimum evolution; ML, maximum likelihood, rRNA, ribosomal RNA; SSU, small subunit.

Abstract

Evolutionary trees of prokaryotes usually define the known classes and phyla but less often agree on the relationships among those groups. This has been attributed to the effects of horizontal gene transfer, biases in sequence change, and large evolutionary distances. Furthermore, higher-level clades of prokaryote phyla rarely are supported by information from ecology and cell biology. Nonetheless, common patterns are beginning to emerge as larger numbers of species are analyzed with sophisticated methods. Here we show how combined evidence from phylogenetic, cytological, and environmental data support the existence of an evolutionary group that appears to have had a common ancestor on land early in Earth's history and includes two-thirds of known prokaryote species. Members of this terrestrial clade (Terrabacteria), which includes Cyanobacteria, the Gram-positive phyla (Actinobacteria and Firmicutes), and two phyla with cell walls that differ structurally from typical Gram-positive and Gram-negative phyla (Chloroflexi and *Deinococcus-Thermus*), possess important adaptations such as resistance to environmental hazards (e.g., desiccation, ultraviolet radiation, and high salinity) and oxygenic photosynthesis. Moreover, the unique properties of the cell wall in Gram-positive taxa, which likely evolved in response to terrestrial conditions, have contributed towards pathogenicity in many species. These results now leave open the possibility that terrestrial adaptations may have played a larger role in prokaryote evolution than currently understood.

Introduction

The evolutionary history of prokaryotes has been intensely studied using DNA and protein sequences, gene content, and sequence signatures (e.g., Gupta 1998; Wolf et al. 2001; Brochier et al. 2002; Battistuzzi, Feijão, and Hedges 2004; Ciccarelli et al. 2006; Lienau et al. 2006). Although the monophyly of most classes and phyla is well resolved, no consensus has been reached on relationships among those groups, especially among phyla. Horizontal gene transfer (HGT) has been considered at least part of the reason for this phylogenetic uncertainty (Doolittle and Bapteste 2007), although a working model holds that the tree can be resolved with a set of core genes (proteins) having reduced levels of HGT (Choi and Kim 2007). Core proteins are those shared by a set of species for which a major influence of HGT can be excluded. Based on different HGT detection methods and species sets, this core protein approach has identified overlapping sets of 20–40 proteins from complete genomes that are shared by eubacteria (also called "Bacteria"), archaebacteria (also called "Archaea"), and eukaryotes (e.g., Battistuzzi, Feijão, and Hedges 2004; Charlebois and Doolittle 2004; Ciccarelli et al. 2006). However, phylogenetic studies using core proteins often have differed in major ways from analyses of ribosomal RNA (rRNA) genes, leading to an overall uncertainty in prokaryote phylogeny. Here, we conducted sequence analyses of both types of genes to search for common patterns and reconcile the differences.

For our primary analysis we constructed a core protein tree with 25 protein-coding genes from 218 species. For comparison with the protein tree we also built an rRNA tree, from 189 species, that combined sequences of the small subunit (SSU), the gene traditionally used for analyses, and the rarely used large subunit (LSU). We subjected these data sets to a suite of sequence analyses and identified a sequence bias in the rRNA data that, when corrected, brings the rRNA and protein trees into closer agreement than in past studies. The trees reveal a large clade of phyla comprising two-thirds of the 9,740 recognized species of prokaryotes, including all Gram-positive species and most species that form spores. Together with environmental data from culture-independent studies, and molecular clock analyses, we show that this clade likely evolved on land early in the Precambrian, with some lineages later re-invading marine habitats. These results have implications for understanding the relations between the key adaptations of the terrestrial clade and the environment in which they evolved.

Materials and Methods

Data assembly and sequence analyses. For our primary analysis we constructed a protein tree with 25 protein-coding genes. These correspond to a subset of previously identified orthologous core proteins (Battistuzzi, Feijão, and Hedges 2004) that were used as queries for a similarity search (Altschul et al. 1997) against 311 fully sequenced genomes of Eubacteria and Archaebacteria (Table S1, Supplementary Material). Given the large number of species analyzed, a few species-specific gene losses are expected even in widely distributed genes. To maximize the number of protein-coding genes, 28 species showing such losses were omitted resulting in a data set of 283 species. In doing so, we created a complete matrix of genes and species and avoided any potential bias of missing data. We chose classes as our working taxonomic level because species of a same class are obtained in our and other phylogenies in highly supported monophyletic clusters (Ciccarelli et al. 2006; Pisani, Cotton, and McInerney 2007). The omitted species are members of monophyletic classes already represented. The retrieved sequences were aligned for each protein by ClustalX (Thompson, Higgins, and Gibson 1994). Distance and maximum likelihood (ML) single protein phylogenies were built in the program MEGA4 (Tamura et al. 2007) (Neighbor-Joining, model JTT +gamma = 0.5, 1, and 1.5, complete deletion of gaps) and the program RAxML (Stamatakis 2006) (maximum likelihood, model JTT+estimated gamma) respectively to check for orthology and possible HGT events. Genes with nested domains (Eubacteria and Archaebacteria) and/or highly supported (≥95 % bootstrap) nesting of one class within another were considered as candidates for non-vertical inheritance and deleted from the data set.

The remaining genes (25) were concatenated in a final alignment of 18,586 amino acid sites. From this alignment, site homology was further refined (Castresana 2000) using monophyly of classes as an approximation of the strength of the phylogenetic signal in progressively reduced data sets (i.e. a stronger signal results in more monophyletic classes). Based on this analysis, non-conserved sites were omitted, resulting in a final concatenated alignment of 6,884 amino acids and 218 non-redundant (i.e., one strain per species) species, which were used in non-partitioned and partitioned analyses. For comparison we built a phylogeny with all available non-redundant species (189 total; 19 eubacterial classes, 10 archaebacterial classes) from the European Ribosomal RNA Database. The initial rRNA alignment based on secondary structure (Wuyts, Perriere, and de Peer 2004) was modified to include only conserved sites using the same approach applied to proteins to select a threshold between number of sites and phylogenetic signal (Castresana 2000). The final alignment included a total of 3,786 conserved nucleotides (60% of the original alignment) from the concatenation of SSU and LSU rRNA genes. We made little modifications to the species composition of the rRNA alignments to preserve the original secondary structure alignment; only two species (*Methanopyrus kandleri* and *Nanoarchaeum equitans*) that were absent from the database were added because they represented additional classes.

Phylogenetic analyses of aligned sequences were conducted with ML and Bayesian methods (Ronquist and Huelsenbeck 2003; Stamatakis 2006) on partitioned data sets in order to allow the optimization of parameters for each gene. Phylogenetic confidence was estimated with 100 bootstrap replicates in the ML phylogeny and by posterior probability in the Bayesian approach. Additional analyses were carried out on the protein and rRNA data set with a method (Brinkmann and Philippe 1999) designed to identify slow-evolving sites. For the primary phylogenetic analyses, the root was set between eubacteria and archaeabacteria, which is the current consensus based on duplicate gene evidence (Zhaxybayeva, Lapierre, and Gogarten 2005). In the rRNA analyses we also used a modified version (Tamura and Kumar 2002) of the LogDet analysis (Lockhart et al. 1994) for modeling base compositional differences, as implemented in the program MEGA4 (Tamura and Kumar 2002); this was carried out on the complete data set with 100 bootstrap replicates.

Times of divergence were estimated using the protein and rRNA data sets separately, ML phylogenies, and three methods: nonparametric rate smoothing (Sanderson 1997), penalized likelihood (Sanderson 1997), and Bayesian analysis (partitioned and non-partitioned data sets) (Thorne and Kishino 2002). Separate analyses were carried out with eubacteria and archaebacteria using reciprocal rooting. Branch lengths were estimated with a JTT+gamma model for the protein data set and Felsenstein 84 (F84) model (Kishino and Hasegawa 1989; Felsenstein and Churchill 1996) with estimation of gamma distribution and transition/transversion ratio for the rRNA data set; this was accomplished with the programs Estbranches (Thorne and Kishino 2002), and PamL (Yang 1997). We used six calibration points from the geologic and biomarker records, including the earliest habitable time at 4.2 Ga based on ocean-boiling impact probabilities (such impacts also may have occurred as late as 3.8 Ga during the Late Heavy Bombardment) (Sleep et al. 1989; Zahnle et al. 2007), earliest continents at 4.0 Ga (Rosing et al. 2006), earliest methanogens at 3.46 Ga (Bapteste, Brochier, and Boucher 2005; Ueno et al. 2006), earliest oxygen at 2.3 Ga (Holland 2002), divergence of Chlorobia and Bacteroidetes at 1.64 Ga (Brocks et al. 2005), and of Gammaproteobacteria and Betaproteobacteria at 1.64 Ga (Brocks et al. 2005). Additional details on parameter specifications for each analysis are in the Supplementary Material.

Species counts. A list of validly published bacterial names was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), the German Collection of Microorganisms and Cell Culture (www2.dsmz.de). From this list all subspecies and synonymous names were removed to obtain a total count of prokaryote species. Cyanobacteria were not included in the DSMZ list because they have been historically associated with algae in taxonomic treatments. We retrieved information regarding this phylum from Algaebase (www.algaebase.org). Furthermore, we integrated the genera listed in DSMZ with those present in NCBI (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/) (e.g., Dehalococcoides). A breakdown of the number of species in each major category is given in Table S3 in the Supplementary Material.

Environmental evidence. Information on the natural habitat of families or single genera was retrieved from the literature. Lineages were categorized as terrestrial if their known habitat is strictly non-marine (e.g., soil or rock on continents), freshwater (e.g., lakes, rivers, springs) or if their host is a non-marine species. Marine lineages have their primary habitat in salt water environments (e.g., sea surface, water column, sea floor, deep see vent, etc.), or are associated with marine hosts. ML family-level phylogenies for each of the classes Actinobacteria, Cyanobacteria, and Deinococci were estimated from an SSU alignment (secondary structure) using one representative per family. One member of each of the other classes in the terrestrial clade (Group I) was used as outgroup. The class-level phylogeny of Firmicutes (Fig. 1B) and an existing phylogeny of Chloroflexi (Costello and Schmidt 2006) were used. The habitat assignments of the lineages and of the common ancestor were estimated with maximum parsimony (MP) and ML (Maddison and Maddison 1989; Maddison and Maddison 2008). Evidence supporting Group I and II was drawn from phylogenetic analyses (this study) and the literature for Gram staining and spore production (Holt 1984; Garrity 2001). For quantitative estimates of Group-I versus Group-II sequences from different environments (Table 1), only culture-independent studies were considered, to avoid biases introduced by culturing methods, although other biases may be present. Information for four diverse habitat classifications was retrieved from the literature: (i) deep sea (Tringe et al. 2005; Sogin et al. 2006; Huber et al. 2007; Lauro and Bartlett 2008), (ii) sea surface (DeLong 2005; Rusch et al. 2007), (iii) humid soils (Tringe et al. 2005; Roesch et al. 2007; Aislabie, Jordan, and Barker 2008), and (iv) arid (warm and dry) soils (Chanal et al. 2006; Connon et al. 2007). Additional details are available in the Supplementary Material.

Results and Discussion

Phylogenetic evidence. The maximum likelihood (ML) phylogeny obtained with the concatenated data set of SSU and LSU rRNA genes from 189 species (Fig. 1A) is similar to earlier SSU-only phylogenies in identifying a single large group of classes and phyla, supported here by 89% ML bootstrap probability (BP) and 100% Bayesian posterior probability (PP). The group contains Bacteroidetes, Chlamydiae, Chlorobia, Fibrobacteres, Planctomycetacia, Proteobacteria, and Spirochaetes. The tree was rooted with Archaebacteria and the remaining classes stem in a ladder-like fashion from the rooted tree (Fig. 1A, insets). The hyperthermophilic classes Aquificae and Thermotogae are the most basal branches, followed by *Deinococcus-Thermus* and Cyanobacteria. A ML phylogeny built from an alignment with only slow-evolving sites, and a Bayesian analysis of all sites, both formed the identical large group of classes and phyla and showed the same topology at the base of the tree. Furthermore, they differed only at nodes that were poorly supported in both trees (see Supplementary Material).

The protein tree (Fig. 1B) is similar to the rRNA tree in supporting the same cluster of classes and phyla, at 89% BP and 100% PP. It differs from the rRNA tree in placing all other eubacteria, except for the hyperthermophiles and Fusobacteria, in an even larger group (Group-I), supported by 53% BP and 100% PP, rather than in a step-wise branching order near the root. Members of Group-I include the phyla Actinobacteria, Chloroflexi, Cyanobacteria, *Deinococcus-Thermus*, and Firmicutes. A ML phylogeny built from an alignment with only slow-evolving sites was identical and showed increased support for Group-I (81% BP) (Fig. 1B). Trees showing similar major groupings of phyla have been found in the past (Gupta and Johari 1998; Brochier et al. 2002; Wolf et al. 2002; House, Runnegar, and Fitz-Gibbon 2003; Battistuzzi, Feijão, and Hedges 2004; Lienau et al. 2006) indicating stability with increased taxon sampling and application of diverse methods. Nonetheless, most relationships of the phyla within Group-I and the other, smaller group (Group-II) remain uncertain.

Although the rooted versions of the two trees (rRNA and protein tree) are different in the order of their earliest branches (Fig. 1, insets), the overall similarity of the unrooted trees suggested that a base compositional bias present in the rRNA sequences might explain the difference, especially given the high GC ratio of SSU and LSU in taxa near the root of the rRNA tree (Deinococci, Aquificae, and Thermotogae; Fig. 1A). When methods designed to compensate for such biases have been used on rRNA gene data in the past (Brochier et al. 2002) they did not fully reproduce Group-I but nonetheless supported major components of Group-I. For example, the high GC taxon of Group-I, *Deinococcus-Thermus*, that typically clusters with other high GC taxa (hyperthermophiles) near the root, instead clustered with the Group-I taxon Cyanobacteria (Brochier et al. 2002).

When we used a nucleotide substitution model (Tamura and Kumar 2002) to compensate for compositional biases in the combined SSU-LSU rRNA data set, all components of Group-I were obtained (69% BP) except *Deinococcus-Thermus*. Group-II was also obtained, albeit with a lower support (41% BP) (Fig. 1A and Supplementary Material). Nonetheless, the deep position of the high-GC *Deinococcus-Thermus* lineage probably reflects the susceptibility of rRNA data sets to compositional biases even when ameliorating methods are applied. As is typical of most sequence analyses of these deeply divergent groups (Brochier et al. 2002), none of these trees are strongly supported, except with Bayesian posterior probabilities. While further resolution and support of the GC-bias hypothesis may not be possible, this evidence suggests that it has affected several key nodes in the prokaryote rRNA phylogeny, placing greater emphasis on the protein phylogeny (Fig. 1B). Despite the small number of nodes affected in the rRNA phylogeny, it appears to have delayed general recognition of a major evolutionary clade, Group-I.

The deepest (most basal) nodes in the protein and rRNA trees are occupied by the hyperthermophiles, Groups IV and V (Aquificae and Thermotogae), a position that has been criticized based mostly on compositional biases dictated by their lifestyle (Brochier and Philippe 2002). However, contrary to previous phylogenies (Brochier and Philippe 2002; Ciccarelli et al. 2006; Pisani, Cotton, and McInerney 2007), the use of multiple methods to compensate for this and other biases (e.g., analysis of only slow evolving sites) did not change the phylogenetic position of these two lineages in either the protein or rRNA trees, increasing the confidence in an early origin of the hyperthermophiles. The phylum Fusobacteria (Group-I/III) appears in the protein tree of eubacteria basal to Groups I and II and above the hyperthermophiles. Although this lineage has generally been considered a close relative of Firmicutes (Mira et al. 2004), alternative positions have been found, often associated with hyperthermophiles, in large phylogenetic studies (Gupta 2003; Ciccarelli et al. 2006; Pisani, Cotton, and McInerney 2007). Furthermore, in a Bayesian analysis of the protein data set, Fusobacteria is placed within Group-I with 100% PP. Based on this phylogenetic evidence and on the extensive HGT history of this lineage (Mira et al. 2004), the position of Fusobacteria remains uncertain.

Organismal evidence. The cytological and physiological characteristics of eubacteria (Table 1) lend support to the recognition of these two major groups. Group-I phyla Actinobacteria and Firmicutes (including the classes Bacilli, Clostridia, and Mollicutes) are Gram-positive and as such have a thick peptidoglycan layer; they also include mostly terrestrial taxa (see below). Group-II (ancestrally marine, see below) includes most of the Gram-negative taxa, many of which are also terrestrial. These include members of Proteobacteria, Acidobacteria, and the Cytophaga-Flavobacteria-Bacteroidetes (CFB) group (Connon et al. 2007). However, experiments have shown that Gram-negative species that are terrestrial decrease in abundance after soil drying while Gram-positives (Actinobacteria and Firmicutes) increase (Rokitko et al. 2001), suggesting an ancestral function (desiccation resistance) of the peptidoglycan layer. Furthermore, the Gram-positive taxa and Cyanobacteria produce resting stages (e.g., spores), albeit not evolutionarily related, which confer resistance to multiple stresses typical of terrestrial habitats such as desiccation, ultraviolet radiation, and high salt concentration (Potts 1994; Nicholson et al. 2000). Only one other type of spore is known in prokaryotes and it is constrained to one order (i.e., derived) within the Group-II Class Deltaproteobacteria (Myxococcales) (Nicholson et al. 2000).

There is confusion in the literature over the number of described species of prokaryotes. Often, the number reported is approximately 6,000 (Oren 2004) but our preliminary survey showed this number to be an underestimate by as much as 30–40%. We found that there are 9,740 recognized species of prokaryotes, of which Group-I comprises 63% and Group-II comprises 33%. The most species-rich lineages are Actinobacteria and Cyanobacteria (Group-I) and Gammaproteobacteria (Group-II), with more than 1,000 known species in each taxon (Supplementary Material). Many pathogens of humans and other terrestrial eukaryotes are Gram-positive and therefore are members of Group-I (Holt 1984; Fischetti et al. 2006). The structural characteristics of Gram-positive prokaryotes, such as the lack of an outer membrane and presence of a thick peptidoglycan layer, have led to novel adaptations for pathogenicity including unique surface proteins, toxins, and enzymes (Fischetti et al. 2006). Thus, aspects of their pathogenicity are probably related to a terrestrial ancestry, either directly or indirectly. Similarly, radiation tolerance of *Deinococcus* is likely related to selection for desiccation tolerance (Mattimore and Battista 1996).

Environmental evidence. The environment occupied by species in these two groups is consistent with the evolution of desiccation-resistant traits in Group-I. Culture-independent sampling of prokaryotes, including metagenomic studies, show that marine samples have the lowest fraction of Group-I taxa and continental (terrestrial) samples have the highest fraction (Table 1). At the extremes of the marine and terrestrial environments, some deep sea sampling (Tringe et al. 2005) reveals a virtual absence (0–1%) of Group-I sequences whereas hyperarid desert samples are comprised almost exclusively (99%) of Group-I sequences (Connon et al. 2007). Near-surface marine samples (Rusch et al. 2007) have on average a higher fraction (14%) of Group-I sequences than those from the deep sea, and samples of arid soils (Chanal et al. 2006) usually have a higher fraction than those of humid soils (Tringe et al. 2005). Viral communities also parallel this pattern, with viruses of Group-I species dominating terrestrial samples and those of Group-II dominating marine samples (Fierer et al. 2007). Despite these general trends, the composition of soil communities is phylogenetically and structurally complex, with different phyla dominating based on the location, type, and structure of the soil (Mummey et al. 2006).

Ancestor-analysis provides additional support by showing that the earliest-branching lineages of each phylum in Group-I are terrestrial (Fig. S6, S7 in Supplementary Material). In agreement with previous studies, these include Gloebacteria (Cyanobacteria) and Rubrobacteriales (Actinobacteria) which are found exclusively in terrestrial environments (Stackebrandt, Rainey, and WardRainey 1997; Ludwig and Klenk 2001; Seo and Yokota 2003; Gao, Paramanathan, and Gupta 2006; Tomitani et al. 2006; Kunisawa 2007), and most of Clostridia (Firmicutes) which inhabit soil or are parasites of terrestrial hosts. There are only three known families in *Deinoccocus-Thermus*; two of them (Deinococcaceae, and Trueperaceae) are terrestrial and the third contains both marine and terrestrial species. Finally, terrestriality is widespread in the Phylum Chloroflexi with evidence of the earliest branches living in terrestrial habitats (Costello and Schmidt 2006). Parsimony and ML ancestral state reconstructions show support (MP: 100%, ML: 73%) for a terrestrial habitat preference in the ancestor of Group-I. Although the natural habitat and distribution of most species of prokaryotes is not well-known, the combined evidence from phylogenetic, organismal and environmental analyses supports a terrestrial origin of Group-I (Table 1).

For Group-I, the appropriate name Terrabacteria is available, previously applied to a subset of phyla (Actinobacteria, Cyanobacteria, and *Deinococcus-Thermus*) in a study involving fewer sequences (Battistuzzi, Feijão, and Hedges 2004). The current analysis differs in defining a larger land clade (expanded to include Bacilli, Chloroflexi, Clostridia, and Mollicutes), reconciling rRNA and protein tree differences, and integrating cytological and environmental data. Fusobacteria may be an additional member of Terrabacteria because its position varied from below the major Group-I/Group-II split in the ML protein tree (weakly supported) to within Group-I in the Bayesian tree (strongly supported). Members of Group-II occupy diverse environments from marine to terrestrial (Madigan, Martinko, and Parker 2003). However, the limited ecological information indicates that terrestrial adaptations of Group II are mostly restricted to low taxonomic levels (species and genera) rather than higher (derived) levels. This would suggest an aquatic ancestor for this group as a whole and thus we propose the name Hydrobacteria (from the Greek, *hydro*, water) in allusion to the moist environment inferred for the common ancestor of these species. Although specific environments appear to have influenced the early evolutionary history of each of the two major groups, many descendant species living today are adapted to other environments.

Early evolution. The earliest evidence of life in the fossil record is from marine environments, 3.5 billion years ago (Ga) (Schopf et al. 2007) whereas ancient soils from South Africa (2.6 Ga) record the earliest terrestrial ecosystems (Watanabe, Martini, and Ohmoto 2000). Later in the Precambrian, there is abundant evidence of terrestrial life (Horodyski and Knauth 1994; Schwartzman 1999). To better constrain the timing of the colonization of land, we estimated divergence times among lineages using Bayesian and maximum likelihood methods. The divergence of Terrabacteria and Hydrobacteria was estimated to have occurred in the mid-Archean, 3.18 Ga (2.83–3.54 Ga) (Fig. 2), which is consistent with both the origin of continents that occurred earlier (4.0–3.8 Ga) (Hawkesworth and Kemp 2006; Rosing et al. 2006) and the first evidence of terrestrial ecosystems that occurred later (2.6 Ga). Alternatively, assuming that the Earth's surface was not habitable until as late as 3.8 Ga (instead of 4.2 Ga), the resulting estimates are ~4–5% younger. A recent study on the effects of UV fluxes for terrestrial life (Cockell and Raven 2007) suggests that colonization of land was possible even before the establishment of a protective ozone layer. This scenario agrees with our evolutionary hypothesis of a land clade (Terrabacteria) in which Cyanobacteria and, thus, oxygenic photosynthesis (Raymond and Blankenship 2008), evolved after the colonization of land (3.54–2.66 Ga). While it is too soon to conclude that all of the major adaptations of Terrabacteria—including oxygenic photosynthesis and resistance to environmental hazards—necessarily evolved on land, these results now leave open the possibility that terrestrial adaptations may have played a larger role in prokaryote evolution than currently understood.

Supplementary Material

Additional methodological information as well as a list of the species used in each data set is available in the Supplementary Material.

Acknowledgments

We thank J. G. Ferry, J. F. Kasting, S. Kumar, J. Macalady, and H. Ohmoto for discussion. This work was supported by grants from the National Science Foundation and National Aeronautics and Space Administration (to S.B.H.).

Literature Cited

Aislabie JM, Jordan S, Barker GM. 2008. Relation between soil classification and bacterial diversity in soils of the Ross Sea region, Antarctica. Geoderma 144:9-20.

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402.

Bapteste E, Brochier C, Boucher Y. 2005. Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. Archaea 1:353-363.

Battistuzzi FU, Feijão A, Hedges SB. 2004. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. BMC Evol Biol 4:44.

Brinkmann H, Philippe H. 1999. Archaea sister group of Bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. Mol Biol Evol 16:817-825.

Brochier C, Babteste E, Moreira D, Philippe H. 2002. Eubacterial phylogeny based on translational apparatus proteins. Trends Genet 18:1-5.

Brochier C, Philippe H. 2002. Phylogeny - A non-hyperthermophilic ancestor for bacteria. Nature 417:244-244.

Brocks JJ, Love GD, Summons RE, Knoll AH, Logan GA, Bowden SA. 2005. Biomarker evidence for green and purple sulphur bacteria in a stratified Palaeoproterozoic sea. Nature 437:866-870.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540-552.

Chanal A, Chapon V, Benzerara K, Barakat M, Christen R, Achouak W, Barras F, Heulin T. 2006. The desert of tataouine: an extreme environment that hosts a wide diversity of microorganisms and radiotolerant bacteria. Environ Microbiol 8:514-525.

Charlebois RL, Doolittle WF. 2004. Computing prokaryotic gene ubiquity: rescuing the core from extinction. Genome Res 14:2469-2477.

Choi IG, Kim SH. 2007. Global extent of horizontal gene transfer. Proc Natl Acad Sci U S A 104:4489-4494.

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. 2006. Toward automatic reconstruction of a highly resolved tree of life. Science 311:1283-1287.

Cockell CS, Raven JA. 2007. Ozone and life on the Archaean Earth. Philos Transact A Math Phys Eng Sci 365:1889-1901.

Connon SA, Lester ED, Shafaat HS, Obenhuber DC, Ponce A. 2007. Bacterial diversity in hyperarid Atacama Desert soils. J Geophys Res 112: Article G04S17.

Costello EK, Schmidt SK. 2006. Microbial diversity in alpine tundra wet meadow soil: novel Chloroflexi from a cold, water-saturated environment. Environ Microbiol 8:1471-1486.

DeLong EE. 2005. Microbial community genomics in the ocean. Nature Rev Microbiol 3:459-469.

Doolittle WF, Bapteste E. 2007. Pattern pluralism and the Tree of Life hypothesis. Proc Natl Acad Sci U S A 104:2043-2049.

Felsenstein J, Churchill GA. 1996. A Hidden Markov Model approach to variation among sites in rate of evolution. Mol Biol Evol 13:93-104.

Fierer N, Breitbart M, Nulton J, et al. 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. Appl Environ Microbiol 73:7059-7066.

Fischetti VA, Novick RP, Ferretti JJ, Portnoy DA, Rood JI. 2006. Gram-positive pathogens. Virginia: ASM Press.

Gao B, Paramanathan R, Gupta RS. 2006. Signature proteins that are distinctive characteristics of Actinobacteria and their subgroups. Antonie Van Leeuwenhoek 90:69-91.

Garrity GM. 2001. Bergey's manual of systematic bacteriology, 2nd ed. New York: Springer.

Gupta RS. 2003. Evolutionary relationships among photosynthetic bacteria. Photosynth Res 76:173-183.

Gupta RS. 1998. Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. Microbiol Mol Biol Rev 62:1435-1491.

Gupta RS, Johari V. 1998. Signature sequences in diverse proteins provide evidence of a close evolutionary relationship between the *Deinococcus-Thermus* group and cyanobacteria. J Mol Evol 46:716-720.

Hawkesworth CJ, Kemp AIS. 2006. Evolution of the continental crust. Nature 443:811-817.

Holland HD. 2002. Volcanic gases, black smokers, and the Great Oxidation Event. Geochim Cosmochim Acta 21:3811-3826.

Holt JG. 1984. Bergey's manual of systematic bacteriology, 1st ed. Baltimore: Williams & Wilkins.

Horodyski RJ, Knauth LP. 1994. Life on Land in the Precambrian. Science 263:494-498.

House CH, Runnegar B, Fitz-Gibbon ST. 2003. Geobiological analysis using whole genome-based tree building applied to the Bacteria, Archaea and Eukarya. Geobiology 1:15-26.

Huber JA, Mark Welch D, Morrison HG, Huse SM, Neal PR, Butterfield DA, Sogin ML. 2007. Microbial population structures in the deep marine biosphere. Science 318:97-100.

Kishino H, Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J Mol Evol 29:170-179.

Kunisawa T. 2007. Gene arrangements characteristic of the phylum Actinobacteria. Antonie Van Leeuwenhoek 92:359-365.

Lauro FM, Bartlett DH. 2008. Prokaryotic lifestyles in deep sea habitats. Extremophiles 12:15-25.

Lienau EK, DeSalle R, Rosenfeld JA, Planet PJ. 2006. Reciprocal illumination in the gene content tree of life. Syst Biol 55:441-453.

Lockhart PJ, Steel MA, Hendy MD, Penny D. 1994. Recovering Evolutionary Trees under a More Realistic Model of Sequence Evolution. Mol Biol Evol 11:605-612.

Ludwig W, Klenk H-P. 2001. Overview: a phylogenetic backbone and taxonomic framework for prokaryotic systematics. In Boone DR and Castenholtz RW, editors. Bergey's manual of systematic bacteriology. Berlin: Springer-Verlag.

Maddison WP, Maddison DR. 2008. Mesquite: a modular system for evolutionary analysis. Version 2.5 http://mesquiteproject.org.

Maddison WP, Maddison DR. 1989. Interactive analysis of phylogeny and character evolution using the computer program MacClade. Folia Primatol (Basel) 53:190-202.

Madigan MT, Martinko JM, Parker J. 2003. Brock Biology of microorganisms. New Jersey: Prentice-Hall Inc.

Mattimore V, Battista JR. 1996. Radioresistance of *Deinococcus radiodurans*: Functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation. J Bacteriol 178:633-637.

Mira A, Pushker R, Legault BA, Moreira D, Rodriguez-Valera F. 2004. Evolutionary relationships of Fusobacterium nucleatum based on phylogenetic analysis and comparative genomics. BMC Evol Biol 4:50.

Mummey D, Holben W, Six J, Stahl P. 2006. Spatial stratification of soil bacterial populations in aggregates of diverse soils. Microbial Ecology 51:404-411.

Nicholson WL, Munakata N, Horneck G, Melosh HJ, Setlow P. 2000. Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments. Microbiol Mol Biol Rev 64:548-572.

Oren A. 2004. Prokaryote diversity and taxonomy: current status and future challenges. Philos T Roy Soc B 359:623-638.

Pisani D, Cotton JA, McInerney JO. 2007. Supertrees disentangle the chimerical origin of eukaryotic genomes. Mol Biol Evol 24:1752-1760.

Potts M. 1994. Desiccation tolerance of prokaryotes. Microbiol Rev 58:755-805.

Raymond J, Blankenship RE. 2008. The origin of the oxygen-evolving complex. Coord Chem Rev 252:377-383.

Roesch LF, Fulthorpe RR, Riva A, Casella G, Hadwin AKM, Kent AD, Daroub SH, Camargo FAO, Farmerie WG, Triplett EW. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. ISME Journal 1:283-290.

Rokitko PV, Romanovskaya VA, Malashenko YR, Chernaya NA, Gushcha NI, Mikheev AN. 2001. Soil drying as a model for the action of stress factors on natural bacterial populations. Microbiology 72:756-761.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Rosing MT, Bird DK, Sleep NH, Glassley W, Albarede F. 2006. The rise of continents - An essay on the geologic consequences of photosynthesis. Palaeogeogr Palaeoclimatol Palaeoecol 232:99-113.

Rusch DB, Halpern AL, Sutton G, et al. 2007. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biol 5:e77.

Sanderson M. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol Biol Evol 14:1218-1231.

Schopf JW, Kudryavtsev AB, Czaja AD, Tripathi AB. 2007. Evidence of archean life: Stromatolites and microfossils. Precambrian Res 158:141-155.

Schwartzman D. 1999. Life, temperature, and the Earth. New York: Columbia University Press.

Seo PS, Yokota A. 2003. The phylogenetic relationships of cyanobacteria inferred from 16S rRNA, gyrB, rpoC1 and rpoD1 gene sequences. J Gen Appl Microbiol 49:191-203.

Sleep NH, Zahnle KJ, Kasting JF, Morowitz HJ. 1989. Annihilation of ecosystems by large asteroid impacts on the early Earth. Nature 342:139-142.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci USA 103:12115-12120.

Stackebrandt E, Rainey FA, WardRainey NL. 1997. Proposal for a new hierarchic classification system, Actinobacteria classis nov. Int J Syst Bacteriol 47:479-491.

Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596-1599.

Tamura K, Kumar S. 2002. Evolutionary distance estimation under heterogeneous substitution pattern among lineages. Molecular Biology and Evolution 19:1727-1736.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673-4680.

Thorne JL, Kishino H. 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst Biol 51:689-702.

Tomitani A, Knoll AH, Cavanaugh CM, Ohno T. 2006. The evolutionary diversification of cyanobacteria: molecular-phylogenetic and paleontological perspectives. Proc Natl Acad Sci U S A 103:5442-5447.

Tringe SG, von Mering C, Kobayashi A, et al. 2005. Comparative metagenomics of microbial communities. Science 308:554-557.

Ueno Y, Yamada K, Yoshida N, Maruyama S, Isozaki Y. 2006. Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. Nature 440:516-519.

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

Wolf YI, Rogozin IB, Grishin NV, Koonin EV. 2002. Genome trees and the tree of life. Trends Genet 18:472-479.

Wolf YI, Rogozin IB, Grishin NV, Tatusov RL, Koonin EV. 2001. Genome trees constructed using five different approaches suggest new major bacterial clades. BMC Evol Biol 1:8.

Wuyts J, Perriere G, de Peer YV. 2004. The European ribosomal RNA database. Nucleic Acids Res 32:D101-D103.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS 13:555-556.

Zahnle K, Arndt N, Cockell C, Halliday A, Nisbet E, Selsis F, Sleep NH. 2007. Emergence of a habitable planet. Space Sci Rev 129:35-78.

Zhaxybayeva O, Lapierre P, Gogarten JP. 2005. Ancient gene duplications and the root(s) of the tree of life. Protoplasma 227:53-64.

Table 1. Multiple evidence supporting two major groups of eubacteria (Groups I and II).

	Phylogeny				Environmental surveys[b]			
Phylum or lineage	Protein	rRNA	Gram stain[a]	Spores	Deep-sea	Sea surface	Humid soils	Arid Soils
Actinobacteria	I	I	P	Yes	5%	1%	13%	64%
Chloroflexi	I	-	P/N	No	4%	1%	5%	1%
Cyanobacteria	I	I	N	Yes	<1%	6%	4%	-
Deinococcus-Thermus	I	III	P	No	-	<1%	<1%	1%
Firmicutes	I	I	P	Yes	2%	6%	6%	1%
Group I, total					12%	14%	28%	67%
(min-max)					(0–23%)	(7–20%)	(7–41%)	(32–99%)
Acidobacteria	II	-	N	No	<1%	-	13%	1%
Bacteroidetes	II	II	N	No	8%	9%	19%	2%
Chlamydiae	II	II	N	No	-	-	-	-
Chlorobi	II	II	N	No	-	-	-	-
Fibrobacteres	-	II	N	No	-	-	-	-
Planctomycetes	II	II	N	No	1%	13%	<1%	1%

Proteobacteria	II	II	N	(No)[c]	79%	64%	40%	22%
Spirochaetes	II	II	N	No	-	<1%	-	-
Group II, averages					88% (77–100%)	86% (80–93%)	72% (59–93%)	33% (1–68%)
Fusobacteria	I/III	-	N	No	-	-	-	-
Aquificae	IV	V	N	No	-	-	-	-
Thermotogae	V	IV	N	No	-	-	-	-

[a] P: Gram-positive stain; N: Gram-negative stain; *Deinococcus*-*Thermus* stains P but has a cell wall structurally similar to that of Gram-negative taxa

[b] Percentages refer to average taxonomic composition of sequences across multiple geographic sites; see Supplementary Material for references

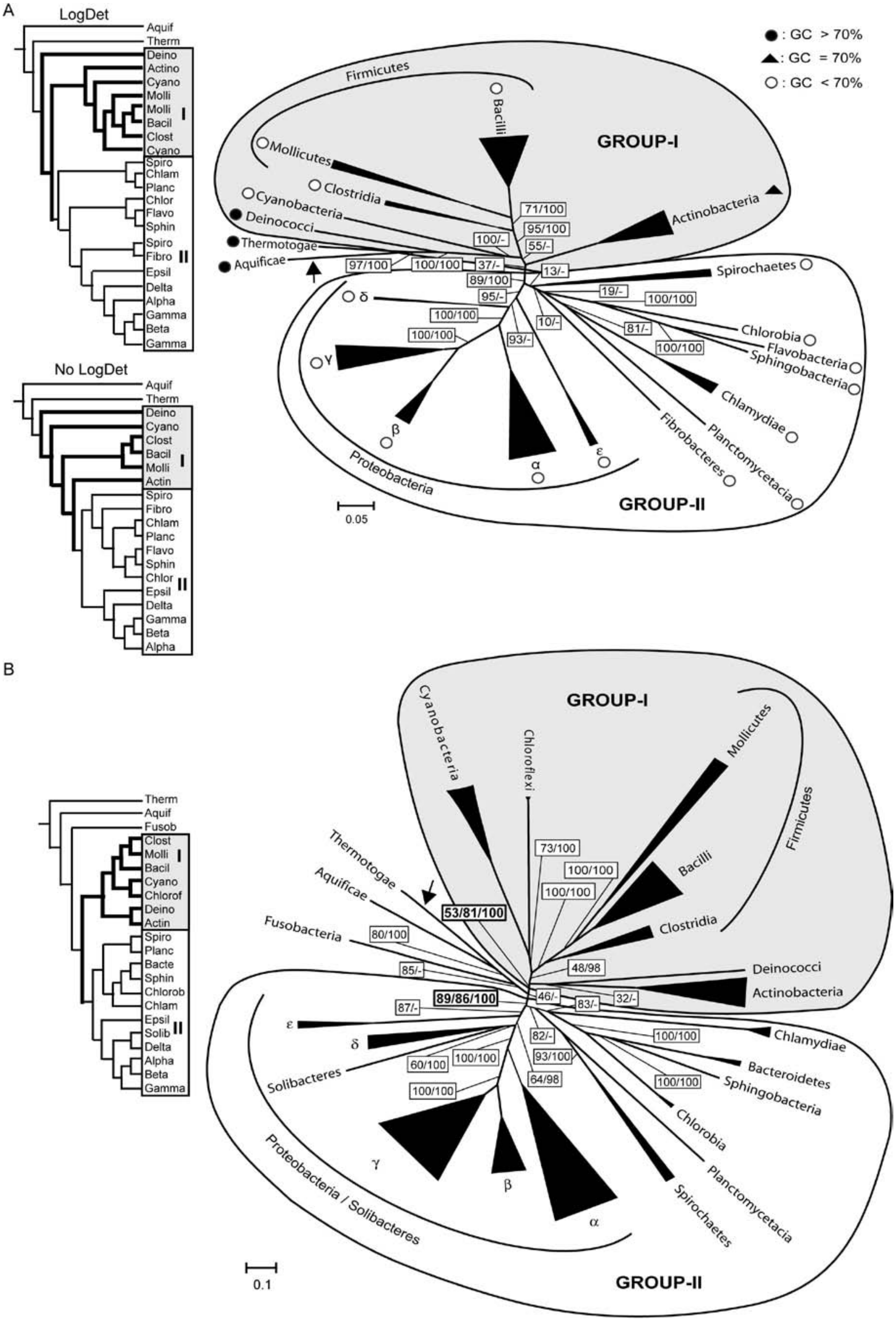
[c] Spores in Proteobacteria are confined to one order in the Deltaproteobacteria

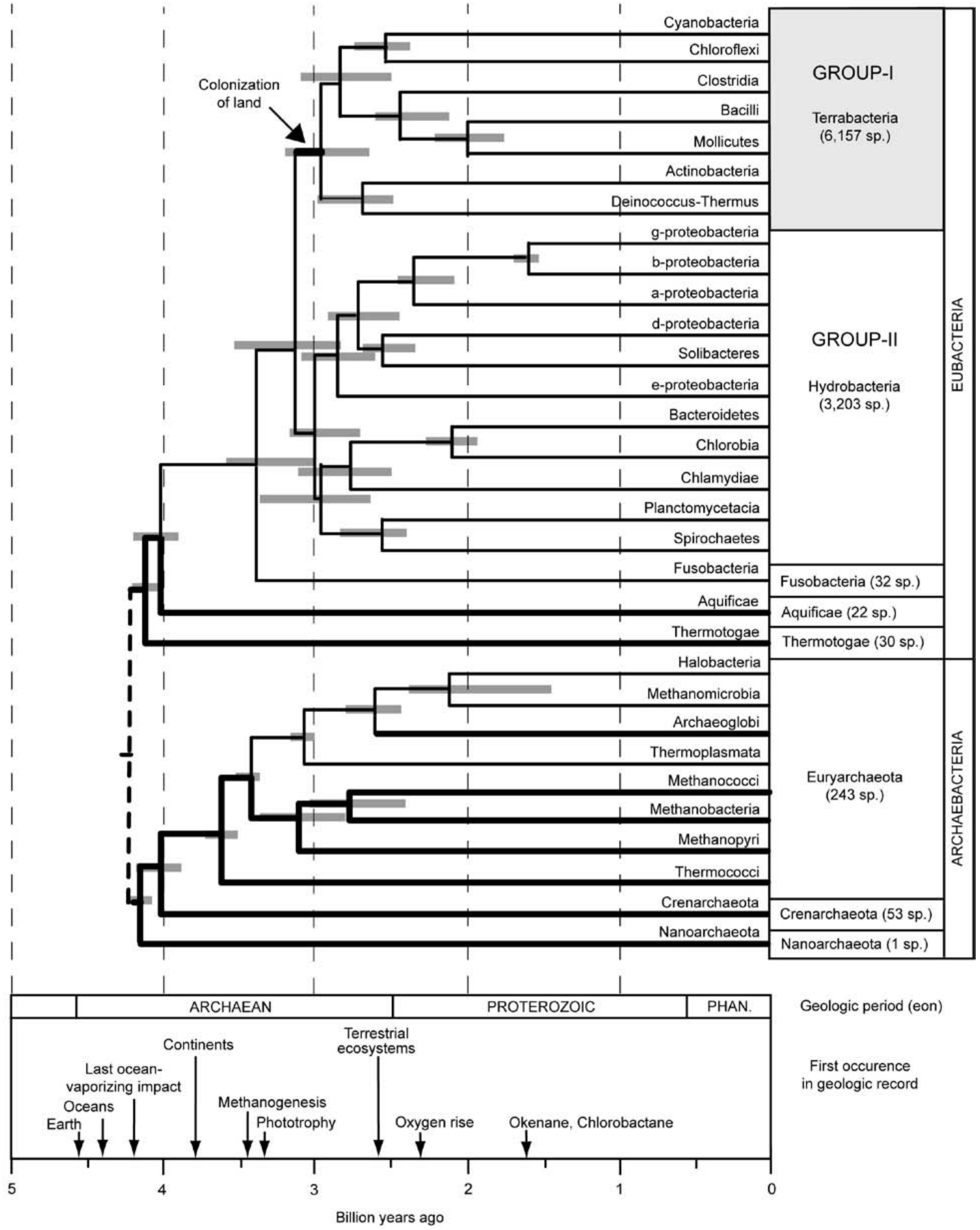
Dashes indicate that no data were available.

Figure captions

Fig. 1. Unrooted ML phylogenies of the ribosomal RNA tree (**A**) and protein tree (**B**) for Eubacteria. Each panel has an inset showing the relationship of the trees rooted with Archaebacteria. Insets in panel A show phylogenies before (No LogDet) and after (LogDet) the correction for compositional biases. Triangles on branches are proportional to the number of sequences analyzed within each lineage (total = 189 and 218, respectively). ML confidence values (left of slash) and Bayesian posterior probabilities are shown at each node; nodes supporting the two major groups in (B) are bold, with middle support value from ML analysis of slow-evolving sites. Filled circles next to clade name in (A) indicate >70% GC content of the conserved sites for each lineage; filled triangle indicates 70%; open circles indicate < 70%. Dashes represent groups not present in the Bayesian phylogeny. The Greek letters indicate the five classes of the Phylum Proteobacteria. Lineages in insets are abbreviated. Actino: Actinobacteria, Alpha: Alphaproteobacteria , Aquif: Aquificae, Bacil: Bacilli, Bacte: Bacteroidetes, Beta: Betaproteobacteria , Chlam: Chlamydiae, Chlor: Chlorobia , Chlorof: Chloroflexi, Clost: Clostridia , Cyano: Cyanobacteria, Deino: *Deinococcus-Thermus*, Delta: Deltaproteobacteria, Epsilon: Epsilonproteobacteria, Fibro: Fibrobacteres, Flavo: Flavobacteria, Fusob: Fusobacteria, Gamma: Gammaproteobacteria, Molli: Mollicutes, Planc: Planctomycetacia, Solib: Solibacteres, Spiro: Spirochaetes, Sphin: Sphingobacteria, and Therm: Thermotogae. Some classes appear multiple times in the tree because their representative species are non-monophyletic. The arrow points to the root.

Fig. 2. Timescale of prokaryote evolutionary history. The timetree shows divergences for Eubacteria and Archaebacteria (ML, protein data set) with particular attention to major groups: Hydrobacteria and Terrabacteria (Eubacteria) and Euryarchaeota and Crenarchaeota (Archaebacteria). First occurrences of major events in the geologic record are represented by arrows on the timescale. The timescale is in billion years ago (Ga). Each horizontal line represents a class; exceptions are the phyla Bacteroidetes (which includes two classes), Cyanobacteria, and Nanoarchaeota. Thicker lines are lineages that include hyperthermophilic species. Gray bars show the range of time estimates for each node, from each of the four estimation methods. For source of species counts and methods, see Supplementary Material.


A
LogDet
Aquif
Therm
Deino
Actino
Cyano
Molli
Molli
Bacil
Clost
Cyano
I
Spiro
Chlam
Planc
Chlor
Flavo
Sphin
Spiro
Fibro
II
Epsil
Delta
Alpha
Gamma
Beta
Gamma
No LogDet
Aquif
Therm
Deino
Cyano
Clost
Bacil
Molli
Actin
I
Spiro
Fibro
Chlam
Planc
Flavo
Sphin
Chlor
Epsil
II
Delta
Gamma
Beta
Alpha
● : GC > 70%
▲: GC = 70%
○ : GC < 70%
Firmicutes
Bacilli
GROUP-I
Mollicutes
Clostridia
Cyanobacteria
Deinococci
Thermotogae
Aquificae
Actinobacteria
71/100
95/100
55/-
100/-
97/100
100/100
37/-
13/-
89/100
95/-
19/-
100/100
100/100
100/100
10/-
93/-
81/-
100/100
Spirochaetes
Chlorobia
Flavobacteria
Sphingobacteria
Chlamydiae
Planctomycetacia
Fibrobacteres
δ
γ
β
α
ε
Proteobacteria
GROUP-II
0.05
B
Therm
Aquif
Fusob
Clost
Molli
I
Bacil
Cyano
Chlorof
Deino
Actin
Spiro
Planc
Bacte
Sphin
Chlorob
Chlam
Epsil
Solib
II
Delta
Alpha
Beta
Gamma
GROUP-I
Cyanobacteria
Chloroflexi
Mollicutes
Firmicutes
Bacilli
Clostridia
Thermotogae
Aquificae
Fusobacteria
73/100
100/100
100/100
53/81/100
80/100
85/-
48/98
89/86/100
46/-
83/-
32/-
87/-
Deinococci
Actinobacteria
ε
δ
Solibacteres
82/-
93/100
60/100
100/100
100/100
64/98
100/100
100/100
Chlamydiae
Bacteroidetes
Sphingobacteria
Chlorobia
Planctomycetacia
Spirochaetes
γ
β
α
Proteobacteria / Solibacteres
GROUP-II
0.1


Colonization of land
Cyanobacteria
Chloroflexi
Clostridia
Bacilli
Mollicutes
Actinobacteria
Deinococcus-Thermus
g-proteobacteria
b-proteobacteria
a-proteobacteria
d-proteobacteria
Solibacteres
e-proteobacteria
Bacteroidetes
Chlorobia
Chlamydiae
Planctomycetacia
Spirochaetes
Fusobacteria
Aquificae
Thermotogae
Halobacteria
Methanomicrobia
Archaeoglobi
Thermoplasmata
Methanococci
Methanobacteria
Methanopyri
Thermococci
Crenarchaeota
Nanoarchaeota
GROUP-I
Terrabacteria (6,157 sp.)
GROUP-II
Hydrobacteria (3,203 sp.)
Fusobacteria (32 sp.)
Aquificae (22 sp.)
Thermotogae (30 sp.)
Euryarchaeota (243 sp.)
Crenarchaeota (53 sp.)
Nanoarchaeota (1 sp.)
EUBACTERIA
ARCHAEBACTERIA
ARCHAEAN
PROTEROZOIC
PHAN.
Geologic period (eon)
Continents
Terrestrial ecosystems
Last ocean-vaporizing impact
Oceans
Earth
Methanogenesis
Phototrophy
Oxygen rise
Okenane, Chlorobactane
First occurence in geologic record
5
4
3
2
1
0
Billion years ago

Supplementary Material
A major clade of prokaryotes with ancient adaptations to life on land

Fabia U. Battistuzzi and S. Blair Hedges

Data assembly and phylogenetic analyses

Protein data set: Amino acid sequences of 25 protein-coding genes ("proteins") were concatenated in an alignment of 18,586 amino acid sites and 283 species. These proteins included: 15 ribosomal proteins (RPL1, 2, 3, 5, 6, 11, 13, 16; RPS2, 3, 4, 5, 7, 9, 11), four genes (RNA polymerase alpha, beta, and gamma subunits, Transcription antitermination factor NusG) from the functional category of Transcription, three proteins (Elongation factor G, Elongation factor Tu, Translation initiation factor IF2) of the Translation, Ribosomal Structure and Biogenesis functional category, one protein (DNA polymerase III, beta subunit) of the DNA Replication, Recombination and repair category, one protein (Preprotein translocase SecY) of the Cell Motility and Secretion category, and one protein (O-sialoglycoprotein endopeptidase) of the Posttranslational Modification, Protein Turnover, Chaperones category, as annotated in the Cluster of Orthologous Groups (COG) (Tatusov et al. 2001).

After removal of multiple strains of the same species, GBlocks 0.91b (Castresana 2000) was applied to each protein in the concatenation to delete poorly aligned sites (i.e., sites with gaps in more than 50% of the species and conserved in less than 50% of the species) with the following parameters: minimum number of sequences for a conserved position: 110, minimum number of sequences for a flank position: 110, maximum number of contiguous non-conserved positions: 32000, allowed gap positions: with half. The signal-to-noise ratio was determined by altering the "minimum length of a block" parameter. This was increased, starting from a minimum of two to a maximum of 80, in order to obtain different data sets retaining approximately 40% (the longest alignment obtainable with the parameters chosen), 30%, 20%, 10%, 5%, and 2% of the original alignment. A phylogeny was built with MEGA4 (NJ, JTT+gamma, with the alpha parameter estimated by the program RAxML (Stamatakis 2006) and the number of monophyletic classes, their bootstrap support, and the monophyly of the phyla Proteobacteria (excluding the position of Solibacteres) and Firmicutes were compared. Solibacteres (Phylum Acidobacteria) was not considered in assessing Proteobacteria monophyly because its taxonomic position as an independent phylum has been questioned in light of recent phylogenetic results (Ciccarelli et al. 2006). In the evaluation of Firmicutes monophyly the position of *Symbiobacterium thermophilum* was not considered (see below). An increase in stringency levels caused a decrease in bootstrap support for the monophyly of classes (used as an approximation of the strength of the phylogenetic signal) because fewer sites were available, yet there was no apparent effect on the recovery of monophyletic classes. For this reason, we selected the 40% stringency level because it maximized the length of the alignment and the number of monophyletic eubacterial classes (Fig. S1).

Preliminary phylogenetic analyses showed a potential bias caused by the presence in the data set of the thermophile *Thermus thermophilus* (Phylum *Deinococcus-Thermus*), most likely caused by its thermophilic adaptations (Omelchenko et al. 2005). In the final data set, we decided to remove this species so that the final composition included 218 species and 6,884 sites (37% of the original alignment). This data set was analyzed with ML (RAxML v. 2.2.1, PROTMIXJTT+gamma) and bayesian methods (MrBayes3, partitioned data set, 2 independent runs of 20 million generations each, sample frequency=1000, model=jones, rates=gamma)

(Ronquist and Huelsenbeck 2003). One representative per class and one for the Phylum Bacteroidetes were chosen in the Bayesian analysis for a total of 31 species. Support for the use of a concatenation of genes came from a consensus analysis of the 25 ML protein trees. This was built using the program Consense of the Phylip package (Felsenstein 1989). This tree showed a generally poor phylogenetic signal in single phylogenies for relationships among classes and phyla and supported the use of a concatenation of these genes to increase the signal to noise ratio (Fig. S2).

Additional analyses were carried out on a data set created by applying the Slow-Fast (SF) method (Brinkmann and Philippe 1999; Philippe et al. 2000) to the original concatenation and building the phylogeny as described above (Fig. S3). This method progressively eliminates from the data set variable sites (i.e., sites with a number of changes above a threshold) leaving only slow evolving positions to estimate the phylogeny. PAUP* v.4 beta10 (Swofford 1998) was used to calculate the number of changes per site in each class represented by multiple species (a maximum of six species representing different genera was used when available). Archaebacteria were analyzed at the domain level because only one class was represented by more than three species. The threshold between slow and fast evolving sites was based on the sum of changes across all phylogenetic categories for a given site: any site showing fewer changes than the selected threshold was considered slow evolving and retained in the alignment. Distance trees (NJ, JTT+gamma, with the alpha parameter estimated by the program RAxML) were built for each data set with threshold of 45, 30, 15, ten, five, and two changes per site. Increase threshold stringency resulted in paraphyly of classes and phyla, and loss of phylogenetic signal. We selected a threshold of 45 changes because it maximized the number of monophyletic classes and phyla (Fig. S1).

Rooting of phylogenetic trees: For the primary phylogenetic analyses, Eubacteria were rooted with Archaebacteria, as has been the consensus in the field based on analyses of duplicated genes (Zhaxybayeva, Lapierre, and Gogarten 2005). However, this is an active area of research and other positions for the root have been suggested.

Symbiobacterium thermophilum: This species is a thermophilic bacterium dependent on microbial commensalism for growth (Ohno et al. 2000). It was classified as an actinobacterium based on its high GC content (Ueda et al. 2001) but recent studies have shown its affiliation with Firmicutes based on genome characteristics, indels, and the absence of proteins uniquely shared with Actinobacteria (Ueda et al. 2004; Gao and Gupta 2005; Gao, Paramanathan, and Gupta 2006). A recent supertree analysis also showed *S. thermophilum* clustering with Clostridia (Pisani, Cotton, and McInerney 2007) as in our phylogeny (both ML and NJ, BP 68% and 58% respectively). Given the amount of evidence, we consider this species as a misclassified actinobacterium and the first high GC member of the Class Clostridia.

Ribosomal RNA (rRNA) data set: small subunit (SSU) and large subunit (LSU) sequences available at the European Ribosomal RNA Database (Van de Peer et al. 2000; Wuyts, Perriere, and de Peer 2004) were used in their aligned form. The alignment was based on the secondary structure of rRNA using *Methanococcus jannachii* and *Sulfolobus acidocaldarius* as models (Van de Peer et al. 2000). A few classes present in the protein data set were absent from the rRNA data set (Bacteroidetes, Chloroflexi, Fusobacteria, and Solibacteres in the eubacteria, and Methanopyri and Nanoarchaeota in the archaebacteria). Two sequences for archaebacteria, *Methanopyrus kandleri* and *Nanoarchaeum equitans*, were added and manually aligned. The missing eubacterial classes were not added because of the ambiguities in manually aligning a few species of uncertain phylogenetic position with hundreds of highly divergent sequences.The sequences for the two subunits were concatenated. As for the protein data set, GBlocks was applied to remove non-conserved sites and the stringency level was chosen using a criterion based on monophyly of eubacterial classes. The parameters used were: minimum number of sequences for a conserved position: 95, minimum number of sequences for a flank position: 95, maximum number of contiguous non-conserved positions: 32000, allowed gap positions: with half. The “minimum length of a block” parameter was progressively increased to obtain different data sets retaining approximately 60%, 50%, 40%, 30%, 20%, and 10% of the original alignment (columns with only gaps are deleted at the beginning of the analysis). A phylogeny was built with MEGA4 (NJ, TamuraNei+gamma, with the alpha parameter estimated by the program RAxML) and the number of monophyletic classes, their bootstrap support and the monophyly of Proteobacteria and Firmicutes were calculated. In the evaluation of Proteobacteria monophyly the position of *Zoogloea ramigera* was not considered (see below). Higher stringency levels caused a decrease in number of monophyletic classes (paraphyly of Gamma and Deltaproteobacteria, Spirochaetes, and Bacilli) as well as a decrease in bootstrap support of the remaining monophyletic ones. Monophyly of the two phyla is unaffected. We selected a stringency of 60% to maximize the number of sites (Fig. S1). The final data set was composed of 189 species for 3,786 sites (approximately 60% of the original alignment) (Table S2). ML and Bayesian trees were built with RAxML and MrBayes3 using GTRMIX+gamma and GTR+gamma, respectively, and partitioning the two subunits. One representative per class was chosen in the Bayesian analysis run with the following parameters: 2 independent runs of 20 million generations each, sample frequency=1000, model=GTR, rates=gamma.

An additional data set was created using the SF method and analyzed as explained above (Fig. S5). The number of changes per site in each eubacterial class represented by multiple species was calculated using the program PAUP* v.4 beta10. Archaebacteria were treated at the domain level because only two classes were represented by more than three species. A maximum of six species was used in each class, spanning different genera when available. As for the protein data set, the number of changes within each class was summed across the two domains to obtain an estimate of variability of each site. Based on this, four threshold levels were tested: 15, 10, 5, and 3 changes per site. Distance trees (NJ, JTT+gamma, with the alpha parameter estimated by the program RAxML) were built for each one of these levels and monophyly of classes and phyla, and bootstrap supports were calculated. Increasing stringency (i.e., lower threshold) resulted in paraphyly of many classes and phyla, and lower bootstrap supports. We selected a threshold of 15 changes because it maximized the number of monophyletic classes, phyla, and their bootstrap values. This new data set includes approximately 60% of the variable sites present in the original data set (Fig. S1).

Zoogloea ramigera: The original classification of this species had placed it within the Gammaproteobacteria (Shin, Hiraishi, and Sugiyama 1993). A more detailed analysis of various strains revealed that this was a misclassification and placed the type strain within the Betaproteobacteria. Nonetheless, some strains did not cluster with the type strain in an SSU phylogenetic tree and were also found missing a particular rhodoquinone-8 (RQ-8) synthesized by the type strain. The putatively misclassified strains were shown to cluster within the Alphaproteobacteria close to *Agrobacterium tumefaciens* (Shin, Hiraishi, and Sugiyama 1993). This position is the same found in our phylogenetic tree of rRNA subunits (BP 100%) and suggests that the sequence named *Z. ramigera* X88894 in the European Ribosomal Database belongs to one of the misclassified strains. We thus consider it an alphaproteobacterium.

Time estimation

Protein data set: One representative per class in Eubacteria and Archaebacteria was chosen for a total of 21 ingroup eubacterial species and ten ingroup archaebacterial species. Five additional data sets were created using randomly chosen eubacterial species to test for sampling bias. Divergence times were estimated with a Bayesian method, Multidivtime T3 (Thorne and Kishino 2002), both with partitioned (T3p) and non partitioned (T3np) genes, and rate smoothing methods: nonparametric rate smoothing (NPRS) and penalized likelihood (PL) (Sanderson 1997). The Bayesian method and NPRS performed as expected but PL showed inconsistent results. The monotonic decrease in square-errors with increasing smoothing factor obtained under this method suggests either a constant rate throughout the tree or rate variations that do not follow a specific pattern (Sanderson 2002). When this case occurs, use of the constant rate molecular clock (LF) is favored, although the reliability of these time estimates remains unclear under the circumstances of uncorrelated rate variations. However, in the absence of other evidence, neither of the methods can be excluded.

Multiple calibration points were used in both the eubacterial and archaebacterial data sets. We used three calibrations within Eubacteria. The first was a maximum boundary for the ingroup root node at 4.2 Ga, which is the mid-point of the time range estimated for the last ocean-vaporizing event (Sleep et al. 1989), while acknowledging a late heavy bombardment at 3.9 Ga (Zahnle et al. 2007) may have included an ocean-boiling impact, and that life may have survived such an event (Wells, Armstrong, and Gonzalez 2003; Zahnle et al. 2007). The second is a minimum time for the divergence of Chlorobia and Bacteroidetes at 1.64 Ga, based on biomarker evidence for chlorobactane in the Barney Creek Formation of the MacArthur Group, Northern Australia (Brocks et al. 2005). The third is a minimum time for the divergence of Gamma and Betaproteobacteria at 1.64 Ga, which comes from biomarker evidence of okenane in the Barney Creek Formation of the MacArthur Group, Northern Australia (Brocks et al. 2005).

For the primary time estimation analyses, we avoided additional calibrations that included Cyanobacteria or involved oxygen metabolism so that we could draw inferences about those organisms and metabolisms. However, two additional calibrations were used to test the robustness of the time estimates. One was a minimum at 2.3 Ga for the divergence of Cyanobacteria and Dehalococcoidetes (Phylum Chloroflexi), corresponding to the presence of oxygen in the atmosphere (Holland 2002). The other was a maximum of 4.0 Ga for the earliest land-dwelling taxa (Group-I), corresponding to the presence of continents (Rosing et al. 2006). The small number of calibration points available for Archaebacteria is a reflection of the poor geologic record of these organisms. Fluid inclusions in dykes of the Dresser Formation (North Pole area, Pilbara craton, Western Australia) have a content of methane highly depleted in the heavy carbon isotope ^{13}C. This depletion is comparable to that produced by methanogenic prokaryotes, offering a calibration point for the origin of these organisms at a minimum of 3.46 Ga (Bapteste, Brochier, and Boucher 2005; Ueno et al. 2006). A second calibration point is determined by the time of the last ocean-vaporizing event, inferred to have happened at 4.2 (maximum boundary) Ga (Sleep et al. 1989) on the ingroup root node.

Ribosomal RNA (rRNA) data set: The same methods used in the analysis of the protein data set were applied to the ML phylogeny of the combined SSU and LSU rRNA data set.

Habitat

We categorized the different lineages of Terrabacteria (Group-I) based on the ecological habitat of terminal taxa to infer the habitat of the common ancestor of this group (Table S4). Information for families, when available, or single genera was retrieved from the literature (Jackson, Ramaley, and Meinsch 1973; Holt 1984; Mohagheghi et al. 1986; Rao and Kumar 1989; Jensen, Dwight, and Fenical 1991; Takizawa, Colwell, and Hill 1993; Fletchner, Johansen, and Clarck 1998; Silva and Pienaar 1999; Wade et al. 1999; Loffler et al. 2000; Gich, Garcia-Gil, and Overmann 2001; Webster et al. 2001; Fletchner et al. 2002; Hanada et al. 2002; Hentschel et al. 2002; Nakamura et al. 2003; Hugenholtz and Stackebrandt 2004; Leiva et al. 2004; Albuquerque et al. 2005; Cox and Battista 2005; Jimenez, Magos, and Collado-Vides 2005; Montalvo et al. 2005; Pires et al. 2005; Thomas 2005; Beleneva and Zhukova 2006; Costello and Schmidt 2006; Hunter, Mills, and Kostka 2006; Miller et al. 2006; Miroshnichenko and Bonch-Osmolovskaya 2006; Rivera-Aguilar et al. 2006; Taddei et al. 2006; Yamada et al. 2006; Anderson and Haygood 2007; Fermani, Mataloni, and Van de Vijver 2007; Garrity et al. 2007; Gorbushina 2007; Jiang et al. 2007; Jumas-Bilak et al. 2007; Li and Brand 2007; Liang et al. 2007; Moore et al. 2007; Rusch et al. 2007; Zhou et al. 2007; Zvyagintsev et al. 2007). A ML family-level phylogeny for each of the classes Actinobacteria, Cyanobacteria, and *Deinococcus-Thermus* was estimated from an SSU alignment (secondary structure) using one representative per family, when available. One member of each of the other classes in Terrabacteria was used as outgroup. The phylogeny of Chloroflexi used was after Costello and Schmidt (Costello and Schmidt 2006), while Firmicutes were considered at the class level. The habitat assignments of the lineages and of the common ancestor was estimated using MacClade (Maddison and Maddison 1989) (maximum parsimony reconstruction of an unordered character) and Mesquite (Maddison and Maddison 2008) (ML reconstruction, Mk1 model) (Figs. S6 and S7). The ancestral states reconstruction shown by the ML method reflects the uncertainty in reconstructing characters for deep phylogenetic nodes. However, the high probability of a terrestrial ancestry for the last common ancestor of the clade (73% terrestrial, 3% marine) is in agreement with the maximum parsimony analysis.

Environmental distribution of eubacterial species was obtained from culture-independent studies, which were considered to avoid biases introduced by culturing methods. However, these studies present biases as well. In deep sea studies, for example, because it is not possible to identify those species that are metabolically active, it is possible that a fraction of the sampled species is, in reality, surface derived (Lauro and Bartlett 2008). Ranges shown in Table 1 in the main text are the lowest and highest fractions for each group found among all studies and sites for each habitat; only Group-I and Group-II taxa are considered.

Literature Cited

Albuquerque L, Simoes C, Nobre MF, Pino NM, Battista JR, Silva MT, Rainey FA, da Costa MS. 2005. Truepera radiovictrix gen. nov., sp. nov., a new radiation resistant species and the proposal of Trueperaceae fam. nov. FEMS Microbiol Lett 247:161-169.

Anderson CM, Haygood MG. 2007. Alpha-proteobacterial symbionts of marine bryozoans in the genus Watersipora. Appl Environ Microbiol 73:303-311.

Bapteste E, Brochier C, Boucher Y. 2005. Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. Archaea 1:353-363.

Beleneva IA, Zhukova NV. 2006. [Bacterial communities of brown and red algae from Peter the Great Bay, the Sea of Japan]. Mikrobiologiia 75:410-419.

Brinkmann H, Philippe H. 1999. Archaea sister group of Bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. Mol Biol Evol 16:817-825.

Brocks JJ, Love GD, Summons RE, Knoll AH, Logan GA, Bowden SA. 2005. Biomarker evidence for green and purple sulphur bacteria in a stratified Palaeoproterozoic sea. Nature 437:866-870.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540-552.

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. 2006. Toward automatic reconstruction of a highly resolved tree of life. Science 311:1283-1287.

Costello EK, Schmidt SK. 2006. Microbial diversity in alpine tundra wet meadow soil: novel Chloroflexi from a cold, water-saturated environment. Environ Microbiol 8:1471-1486.

Cox MM, Battista JR. 2005. Deinococcus radiodurans - the consummate survivor. Nat Rev Microbiol 3:882-892.

Felsenstein J. 1989. Phylogeny Inference Package (Version 3.2). Cladistics 5:164-166.

Fermani P, Mataloni G, Van de Vijver B. 2007. Soil microalgal communities on an antarctic active volcano (Deception island, South Shetlands). Polar Biol 30:1381-1393.

Fletchner VR, Boyer SL, Johansen JR, DeNoble ML. 2002. Spirirestis rafalensis gen. et sp nov (Cyanophyceae), a new cyanobacterial genus from arid soils. Nova Hedwigia 74:1-24.

Fletchner VR, Johansen JR, Clarck WH. 1998. Algal composition of microbiotic crusts from the central desert of Baja California, Mexico. Great Basin Naturalist 58:295-311.

Gao B, Gupta RS. 2005. Conserved indels in protein sequences that are characteristic of the phylum Actinobacteria. Int J Syst Evol Microbiol 55:2401-2412.

Gao B, Paramanathan R, Gupta RS. 2006. Signature proteins that are distinctive characteristics of Actinobacteria and their subgroups. Antonie Van Leeuwenhoek 90:69-91.

Garrity GM, Lilburn TG, Cole JR, Harrison SH, Euzeby J, Tindall BJ. 2007. Taxonomic outline of the Bacteria and Archaea Release 7.7 [Online]. Michigan State University, East Lansing, Michigan.

Gich F, Garcia-Gil J, Overmann J. 2001. Previously unknown and phylogenetically diverse members of the green nonsulfur bacteria are indigenous to freshwater lakes. Arch Microbiol 177:1-10.

Gorbushina AA. 2007. Life on the rocks. Environ Microbiol 9:1613-1631.

Hanada S, Takaichi S, Matsuura K, Nakamura K. 2002. Roseiflexus castenholzii gen. nov., sp. nov., a thermophilic, filamentous, photosynthetic bacterium that lacks chlorosomes. Int J Syst Evol Microbiol 52:187-193.

Hentschel U, Hopke J, Horn M, Friedrich AB, Wagner M, Hacker J, Moore BS. 2002. Molecular evidence for a uniform microbial community in sponges from different oceans. Appl Environ Microbiol 68:4431-4440.

Holland HD. 2002. Volcanic gases, black smokers, and the Great Oxidation Event. Geochimic Cosmochim Acta 21:3811-3826.

Holt JG. 1984. Bergey's manual of systematic bacteriology, 1st ed. Baltimore: Williams & Wilkins.

Hugenholtz P, Stackebrandt E. 2004. Reclassification of Sphaerobacter thermophilus from the subclass Sphaerobacteridae in the phylum Actinobacteria to the class Thermomicrobia (emended description) in the phylum Chloroflexi (emended description). Int J Syst Evol Microbiol 54:2049-2051.

Hunter EM, Mills HJ, Kostka JE. 2006. Microbial community diversity associated with carbon and nitrogen cycling in permeable shelf sediments. Appl Environ Microbiol 72:5689-5701.

Jackson TJ, Ramaley RF, Meinsch WG. 1973. Thermomicrobium, a new genus of extremely thermophilic bacteria. Int J Syst Bacteriol 23:28-36.

Jensen PR, Dwight R, Fenical W. 1991. Distribution of actinomycetes in near-shore tropical marine sediments. Appl Environ Microbiol 57:1102-1108.

Jiang HC, Dong HL, Ji SS, Ye Y, Wu NY. 2007. Microbial diversity in the deep marine sediments from the Qiongdongnan Basin in South China Sea. Geomicrobiol J 24:505-517.

Jimenez JC, Magos YB, Collado-Vides L. 2005. Taxonomy and distribution of freshwater Blennothrix ganeshii Watanabe et KomArek (Oscillatoriaceae, cyanophyceae) from central Mexico. Nova Hedwigia 80:323-333.

Jumas-Bilak E, Carlier JP, Jean-Pierre H, Mory F, Teyssier C, Gay B, Campos J, Marchandin H. 2007. Acidaminococcus intestini sp. nov., isolated from human clinical samples. Int J Syst Evol Microbiol 57:2314-2319.

Lauro FM, Bartlett DH. 2008. Prokaryotic lifestyles in deep sea habitats. Extremophiles 12:15-25.

Leiva S, Yanez M, Zaror L, Rodriguez H, Garcia-Quintana H. 2004. [Antimicrobial activity of actinomycetes isolated from aquatic environments in southern Chile]. Rev Med Chil 132:151-159.

Li ZK, Brand J. 2007. Leptolyngbya nodulosa sp nov (Oscillatoriaceae, a subtropical marine cyanobacterium that produces a unique multicellular structure. Phycologia 46:396-401.

Liang J-B, Chen Y-Q, Lan C-Y, Tam NFY, Zan Q-J, Huang L-N. 2007. Recovery of novel bacterial diversity from mangrove sediment. Mar Biol 150:739-747.

Loffler FE, Sun Q, Li J, Tiedje JM. 2000. 16S rRNA gene-based detection of tetrachloroethene-dechlorinating Desulfuromonas and Dehalococcoides species. Appl Environ Microbiol 66:1369-1374.

Maddison WP, Maddison DR. 1989. Interactive analysis of phylogeny and character evolution using the computer program MacClade. Folia Primatol (Basel) 53:190-202.

Maddison WP, Maddison DR. 2008. Mesquite: a modular system for evolutionary analysis. Version 2.5 http://mesquiteproject.org.

Miller WA, Miller MA, Gardner IA, et al. 2006. Salmonella spp., Vibrio spp., Clostridium perfringens, and Plesiomonas shigelloides in marine and freshwater invertebrates from coastal California ecosystems. Microb Ecol 52:198-206.

Miroshnichenko ML, Bonch-Osmolovskaya EA. 2006. Recent developments in the thermophilic microbiology of deep-sea hydrothermal vents. Extremophiles 10:85-96.

Mohagheghi A, Grohmann K, Himmel M, Leighton L, Updegraff DM. 1986. Isolation and characterization of acidothermus-cellulolyticus gen-nov, sp-nov, a new genus of thermophilic, acidophilic, cellulolytic bacteria. Int J Syst Bacteriol 36:435-443.

Montalvo NF, Mohamed NM, Enticknap JJ, Hill RT. 2005. Novel actinobacteria from marine sponges. Antonie Van Leeuwenhoek 87:29-36.

Moore LR, Coe A, Zinser ER, Saito AM, Sullivan MB, Lindell D, Frois-Moniz K, Waterbury J, Chisholm SW. 2007. Culturing the marine cyanobacterium Prochlorococcus. Limnol Oceanog-Meth 5:353-362.

Nakamura Y, Kaneko T, Sato S, et al. 2003. Complete genome structure of Gloeobacter violaceus PCC 7421, a cyanobacterium that lacks thylakoids (supplement). DNA Res 10:181-201.

Ohno M, Shiratori H, Park MJ, Saitoh Y, Kumon Y, Yamashita N, Hirata A, Nishida H, Ueda K, Beppu T. 2000. Symbiobacterium thermophilum gen. nov., sp. nov., a symbiotic thermophile that depends on co-culture with a Bacillus strain for growth. Int J Syst Evol Microbiol 50 Pt 5:1829-1832.

Omelchenko MV, Wolf YI, Gaidamakova EK, Matrosova VY, Vasilenko A, Zhai M, Daly MJ, Koonin EV, Makarova KS. 2005. Comparative genomics of Thermus thermophilus and Deinococcus radiodurans: divergent routes of adaptation to thermophily and radiation resistance. BMC Evol Biol 5:57.

Philippe H, Lopez P, Brinkmann H, Budin K, Germot A, Laurent J, Moreira D, Muller M, Le Guyader H. 2000. Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. P Roy Soc Lond B Bio 267:1213-1221.

Pires AL, Albuquerque L, Tiago I, Nobre MF, Empadinhas N, Verissimo A, da Costa MS. 2005. Meiothermus timidus sp. nov., a new slightly thermophilic yellow-pigmented species. FEMS Microbiol Lett 245:39-45.

Pisani D, Cotton JA, McInerney JO. 2007. Supertrees disentangle the chimerical origin of eukaryotic genomes. Mol Biol Evol 24:1752-1760.

Rao R, Kumar HD. 1989. Isolation and characterization of the heterocystous blue-green alga Chlorogloeopsis-frtschii from a eutrophic pond. Microbios 57:7-13.

Rivera-Aguilar V, Montejano G, Rodriguez-Zaragoza S, Duran-Diaz A. 2006. Distribution and composition of cyanobacteria, mosses and lichens of the biological soil crusts of the Tehuacan Valley, Puebla, Mexico. Journal of Arid Environments 67:208-225.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Rosing MT, Bird DK, Sleep NH, Glassley W, Albarede F. 2006. The rise of continents - An essay on the geologic consequences of photosynthesis. Palaeogeogr Palaeoclimatol Palaeoecol 232:99-113.

Rusch DB, Halpern AL, Sutton G, et al. 2007. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biol 5:e77.

Sanderson M. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Molecular Biology and Evolution 14:1218-1231.

Sanderson M. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Molecular Biology and Evolution 19:101-109.

Shin YK, Hiraishi A, Sugiyama J. 1993. Molecular systematics of the genus Zoogloea and emendation of the genus. Int J Syst Bacteriol 43:826-831.

Silva SMF, Pienaar RN. 1999. Marine cyanophytes from the Western Cape, south Africa: Chroococcales. S Afr J Bot 65:32-49.

Sleep NH, Zahnle KJ, Kasting JF, Morowitz HJ. 1989. Annihilation of ecosystems by large asteroid impacts on the early Earth. Nature 342:139-142.

Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.

Swofford DL. 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland, MA: Sinauer Associates.

Taddei A, Rodriguez MJ, Marquez-Vilchez E, Castelli C. 2006. Isolation and identification of Streptomyces spp. from Venezuelan soils: morphological and biochemical studies. I. Microbiol Res 161:222-231.

Takizawa M, Colwell RR, Hill RT. 1993. Isolation and Diversity of Actinomycetes in the Chesapeake Bay. Appl Environ Microbiol 59:997-1002.

Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV. 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res 29:22-28.

Thomas DN. 2005. Photosynthetic microbes in freezing deserts. Trends Microbiol 13:87-88.

Thorne JL, Kishino H. 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst Biol 51:689-702.

Ueda K, Ohno M, Yamamoto K, et al. 2001. Distribution and diversity of symbiotic thermophiles, Symbiobacterium thermophilum and related bacteria, in natural environments. Appl Environ Microbiol 67:3779-3784.

Ueda K, Yamashita A, Ishikawa J, Shimada M, Watsuji TO, Morimura K, Ikeda H, Hattori M, Beppu T. 2004. Genome sequence of Symbiobacterium thermophilum, an uncultivable bacterium that depends on microbial commensalism. Nucleic Acids Res 32:4937-4944.

Ueno Y, Yamada K, Yoshida N, Maruyama S, Isozaki Y. 2006. Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. Nature 440:516-519.

Van de Peer Y, De Rijk P, Wuyts J, Winkelmans T, De wachter R. 2000. The European small subunit ribosomal RNA database. Nucleic Acids Res 28:175-176.

Wade WG, Downes J, Dymock D, Hiom SJ, Weightman AJ, Dewhirst FE, Paster BJ, Tzellas N, Coleman B. 1999. The family Coriobacteriaceae: reclassification of Eubacterium exiguum (Poco et al. 1996) and Peptostreptococcus heliotrinreducens (Lanigan 1976) as Slackia exigua gen. nov., comb. nov. and Slackia heliotrinireducens gen. nov., comb. nov., and Eubacterium lentum (Prevot 1938) as Eggerthella lenta gen. nov., comb. nov. Int J Syst Bacteriol 49 Pt 2:595-600.

Webster NS, Wilson KJ, Blackall LL, Hill RT. 2001. Phylogenetic diversity of bacteria associated with the marine sponge Rhopaloeides odorabile. Appl Environ Microbiol 67:434-444.

Wells LE, Armstrong JC, Gonzalez G. 2003. Reseeding of early Earth by impacts of returning ejecta during the late heavy bombardment. Icarus 162:38-46.

Wuyts J, Perriere G, de Peer YV. 2004. The European ribosomal RNA database. Nucleic Acids Res 32:D101-D103.

Yamada T, Sekiguchi Y, Hanada S, Imachi H, Ohashi A, Harada H, Kamagata Y. 2006. Anaerolinea thermolimosa sp. nov., Levilinea saccharolytica gen. nov., sp. nov. and Leptolinea tardivitalis gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes Anaerolineae classis nov. and Caldilineae classis nov. in the bacterial phylum Chloroflexi. Int J Syst Evol Microbiol 56:1331-1340.

Zahnle K, Arndt N, Cockell C, Halliday A, Nisbet E, Selsis F, Sleep NH. 2007. Emergence of a habitable planet. Space Sci Rev 129:35-78.

Zhaxybayeva O, Lapierre P, Gogarten JP. 2005. Ancient gene duplications and the root(s) of the tree of life. Protoplasma 227:53-64.

Zhou JP, Gu YQ, Zou CS, Mo MH. 2007. Phylogenetic diversity of bacteria in an earth-cave in Ghizhou Province, Southwest of China. Journal of microbiology 45:105-112.

Zvyagintsev DG, Zenova GM, Doroshenko EA, Gryadunova AA, Gracheva TA, Sudnitsyn II. 2007. Actynomycete growth in conditions of low moisture. Biology Bull 34:242-247.

Table S1 List of species of Eubacteria and Archaebacteria used in the protein data set and their classification (genome accession numbers can be found at http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi). Species in bold are the ones used in the final ML data set (218 species). Asterisks denote species used in the Bayesian phylogenetic analysis.

Species name	**Classification**
EUBACTERIA	
Acinetobacter sp. ADP1 *	Gammaproteobacteria
Agrobacterium tumefaciens str. C58 *	Alphaproteobacteria
Anabaena variabilis ATCC 29413 *	Cyanobacteria
Anaeromyxobacter dehalogenans 2CP-C *	Deltaproteobacteria
Anaplasma marginale str. St. Maries	Alphaproteobacteria
Anaplasma phagocytophilum HZ	Alphaproteobacteria
Aquifex aeolicus VF5 *	Aquificae
Aster yellows witches'-broom phytoplasma AYWB *	Firmicutes/Mollicutes
Azoarcus sp. EbN1 *	Betaproteobacteria
Bacillus anthracis str. 'Ames Ancestor' *	Firmicutes/Bacilli
Bacillus anthracis str. Ames	Firmicutes/Bacilli
Bacillus anthracis str. Sterne	Firmicutes/Bacilli
Bacillus cereus ATCC 10987	Firmicutes/Bacilli
Bacillus cereus ATCC 14579	Firmicutes/Bacilli
Bacillus cereus E33L	Firmicutes/Bacilli
Bacillus clausii KSM-K16	Firmicutes/Bacilli
Bacillus halodurans C-125	Firmicutes/Bacilli
Bacillus licheniformis ATCC 14580	Firmicutes/Bacilli
Bacillus subtilis subsp. subtilis str. 168	Firmicutes/Bacilli

Bacillus thuringiensis serovar konkukian str. 97-27	Firmicutes/Bacilli
Bacteroides fragilis NCTC 9343 *	Bacteroidetes
Bacteroides fragilis YCH46	Bacteroidetes
Bacteroides thetaiotaomicron VPI-5482	Bacteroidetes
Bartonella henselae str Houston-1	Alphaproteobacteria
Bartonella quintana str. Toulouse	Alphaproteobacteria
Bdellovibrio bacteriovorus HD100	Deltaproteobacteria
Bifidobacterium longum NCC2705 *	Actinobacteria
Bordetella bronchiseptica RB50	Betaproteobacteria
Bordetella parapertussis 12822	Betaproteobacteria
Bordetella pertussis Tomaha I	Betaproteobacteria
Borrelia burgdorferi B31 *	Spirochaetes
Borrelia garinii Pbi	Spirochaetes
Bradyrhizobium japonicum USDA 110	Alphaproteobacteria
Brucella abortus biovar 1 str. 9-941	Alphaproteobacteria
Brucella melitensis 16M	Alphaproteobacteria
Brucella melitensis biovar Abortus 2308	Alphaproteobacteria
Brucella suis 1330	Alphaproteobacteria
Buchnera aphidicola str. APS	Gammaproteobacteria
Buchnera aphidicola str. Bp	Gammaproteobacteria
Buchnera aphidicola str. Sg	Gammaproteobacteria
Burkholderia mallei ATCC 23344	Betaproteobacteria
Burkholderia pseudomallei 1710b	Betaproteobacteria
Burkholderia pseudomallei K96243	Betaproteobacteria
Burkholderia sp. 383	Betaproteobacteria
Burkholderia thailandensis E264	Betaproteobacteria

Campylobacter jejuni RM1221 *	Epsilonproteobacteria
Campylobacter jejuni subsp. Jejuni NCTC 11168	Epsilonproteobacteria
Candidatus Blochmannia floridanus	Gammaproteobacteria
Candidatus Blochmannia pennsylvanicus str. BPEN	Gammaproteobacteria
Candidatus Pelagibacter ubique HTCC1062	Alphaproteobacteria
Candidatus Protochlamydia amoebophila UWE25	Chlamydiae
Carboxydothermus hydrogenoformans Z-2901	Firmicutes/Clostridia
Caulobacter crescentus CB15	Alphaproteobacteria
Chlamydia muridarum Nigg *	Chlamydiae
Chlamydia trachomatis A/HAR-13	Chlamydiae
Chlamydia trachomatis D/UW-3/CX	Chlamydiae
Chlamydophila abortus S26/3	Chlamydiae
Chlamydophila caviae GPIC	Chlamydiae
Chlamydophila felis Fe/C-56	Chlamydiae
Chlamydophila pneumoniae AR39	Chlamydiae
Chlamydophila pneumoniae CWL029	Chlamydiae
Chlamydophila pneumoniae J138	Chlamydiae
Chlamydophila pneumoniae TW-183	Chlamydiae
Chlorobium chlorochromatii CaD3 *	Chlorobia
Chlorobium tepidum TLS	Chlorobia
Chromobacterium violaceum ATCC 12472	Betaproteobacteria
Clostridium acetobutylicum ATCC 824 *	Firmicutes/Clostridia
Clostridium perfringens str. 13	Firmicutes/Clostridia
Clostridium tetani E88	Firmicutes/Clostridia
Colwellia psychrerythraea 34H	Gammaproteobacteria
Corynebacterium diphtheriae NCTC 13129	Actinobacteria

Corynebacterium efficiens YS-314	Actinobacteria
Corynebacterium glutamicum ATCC 13032	Actinobacteria
Corynebacterium jeikeium K411	Actinobacteria
Coxiella burnetii RSA 493	Gammaproteobacteria
Dechloromonas aromatica RCB	Betaproteobacteria
Dehalococcoides ethenogenes 195 *	Chloroflexi/Dehalococcoidetes
Dehalococcoides sp. CBDB1	Chloroflexi/Dehalococcoidetes
Deinococcus radiodurans R1 *	Deinococci
Desulfitobacterium hafniense Y51	Firmicutes/Clostridia
Desulfotalea psychrophila LSv54	Deltaproteobacteria
Desulfovibrio desulfuricans G20	Deltaproteobacteria
Desulfovibrio vulgaris subsp.vulgaris str. Hildenborough	Deltaproteobacteria
Ehrlichia canis str. Jake	Alphaproteobacteria
Ehrlichia chaffeensis str. Arkansas	Alphaproteobacteria
Ehrlichia ruminantium str. Gardel	Alphaproteobacteria
Ehrlichia ruminantium str. Welgevonden	Alphaproteobacteria
Enterococcus faecalis V583	Firmicutes/Bacilli
Erwinia carotovora subsp. atroseptica SCRI1043	Gammaproteobacteria
Erythrobacter litoralis HTCC2594	Alphaproteobacteria
Escherichia coli CFT073	Gammaproteobacteria
Escherichia coli K12	Gammaproteobacteria
Escherichia coli O157:H7	Gammaproteobacteria
Escherichia coli O157:H7 EDL933	Gammaproteobacteria
Escherichia coli W3110	Gammaproteobacteria
Francisella tularensis subsp. holarctica	Gammaproteobacteria
Francisella tularensis subsp. tularensis SCHU S4	Gammaproteobacteria

Frankia sp. CcI3	Actinobacteria
Fusobacterium nucleatum subsp. nucleatum ATCC 25586 *	Fusobacteria
Geobacillus kaustophilus HTA426	Firmicutes/Bacilli
Geobacter metallireducens GS-15	Deltaproteobacteria
Geobacter sulfurreducens PCA	Deltaproteobacteria
Gloeobacter violaceus PCC 7421	Cyanobacteria
Gluconobacter oxydans 621H	Alphaproteobacteria
Haemophilus ducreyi 35000HP	Gammaproteobacteria
Haemophilus influenzae 86-028NP	Gammaproteobacteria
Haemophilus influenzae Rd KW20	Gammaproteobacteria
Hahella chejuensis KCTC 2396	Gammaproteobacteria
Helicobacter hepaticus ATCC 51449	Epsilonproteobacteria
Helicobacter pylori 26695	Epsilonproteobacteria
Helicobacter pylori J99	Epsilonproteobacteria
Idiomarina loihiensis L2TR	Gammaproteobacteria
Jannaschia sp. CCS1	Alphaproteobacteria
Lactobacillus acidophilus NCFM	Firmicutes/Bacilli
Lactobacillus johnsonii NCC 533	Firmicutes/Bacilli
Lactobacillus plantarum WCFS1	Firmicutes/Bacilli
Lactobacillus sakei subsp. sakei 23K	Firmicutes/Bacilli
Lactococcus lactis subsp. Lactis Il1403	Firmicutes/Bacilli
Legionella pneumophila str.Lens	Gammaproteobacteria
Legionella pneumophila str.Paris	Gammaproteobacteria
Legionella pneumophila subsp. pneumophila str. Philadelphia 1	Gammaproteobacteria
Leifsonia xyli subsp. xyli str. CTCB07	Actinobacteria
Leptospira interrogans serovar Copenhageni str. Fiocruz L1-130	Spirochaetes

Leptospira interrogans serovar Lai str. 56601	Spirochaetes
Listeria innocua Clip11262	Firmicutes/Bacilli
Listeria monocytogenes EGD-e	Firmicutes/Bacilli
Listeria monocytogenes str. 4b F2365	Firmicutes/Bacilli
Magnetospirillum magneticum AMB-1	Alphaproteobacteria
Mannheimia succiniciproducens MBEL55E	Gammaproteobacteria
Mesoplasma florum L1	Firmicutes/Mollicutes
Mesorhizobium loti MAFF303099	Alphaproteobacteria
Methylococcus capsulatus str. Bath	Gammaproteobacteria
Moorella thermoacetica ATCC 39073	Firmicutes/Clostridia
Mycobacterium avium subsp. paratubercolosis K-10	Actinobacteria
Mycobacterium bovis AF2122/97	Actinobacteria
Mycobacterium leprae TN	Actinobacteria
Mycobacterium tuberculosis CDC1551	Actinobacteria
Mycobacterium tuberculosis H37Rv	Actinobacteria
Mycoplasma capricolum subsp. capricolum ATCC 27343	Firmicutes/Mollicutes
Mycoplasma gallisepticum R	Firmicutes/Mollicutes
Mycoplasma genitalium G37	Firmicutes/Mollicutes
Mycoplasma hyopneumoniae 232	Firmicutes/Mollicutes
Mycoplasma hyopneumoniae 7448	Firmicutes/Mollicutes
Mycoplasma hyopneumoniae J	Firmicutes/Mollicutes
Mycoplasma mobile 163K	Firmicutes/Mollicutes
Mycoplasma mycoides subsp. Mycoides SC str. PG1	Firmicutes/Mollicutes
Mycoplasma penetrans HF-2	Firmicutes/Mollicutes
Mycoplasma pneumoniae M129	Firmicutes/Mollicutes
Mycoplasma pulmonis UAB CTIP	Firmicutes/Mollicutes

Mycoplasma synoviae 53	Firmicutes/Mollicutes
Neisseria gonorrhoeae FA 1090	Betaproteobacteria
Neisseria meningitidis MC58	Betaproteobacteria
Neisseria meningitidis Z2491	Betaproteobacteria
Neorickettsia sennetsu str. Miyayama	Alphaproteobacteria
Nitrobacter winogradskyi Nb-255	Alphaproteobacteria
Nitrosococcus oceani ATCC 19707	Gammaproteobacteria
Nitrosomonas europaea ATCC 19718	Betaproteobacteria
Nitrosospira multiformis ATCC 25196	Betaproteobacteria
Nocardia farcinica IFM 10152	Actinobacteria
Nostoc sp. PCC 7120	Cyanobacteria
Novosphingobium aromaticivorans DSM 12444	Alphaproteobacteria
Oceanobacillus iheyensis HTE831	Firmicutes/Bacilli
Onion yellows phytoplasma OY-M	Firmicutes/Mollicutes
Pasteurella multocida subsp. multocida str. Pm70	Gammaproteobacteria
Pelobacter carbinolicus DSM 2380	Deltaproteobacteria
Pelodictyon luteolum DSM 273	Chlorobia
Photobacterium profundum SS9	Gammaproteobacteria
Photorhabdus luminescens subsp. laumondii TTO1	Gammaproteobacteria
Porphyromonas gingivalis W83	Bacteroidetes
Prochlorococcus marinus str. MIT 9312	Cyanobacteria
Prochlorococcus marinus str. MIT 9313	Cyanobacteria
Prochlorococcus marinus str. NATL2A	Cyanobacteria
Prochlorococcus marinus subsp. marinus str CCMP1375	Cyanobacteria
Prochlorococcus marinus subsp. pastoris str. CCMP1986	Cyanobacteria
Propionibacterium acnes KPA171202	Actinobacteria

Pseudoalteromonas haloplanktis TAC125	Gammaproteobacteria
Pseudomonas aeruginosa PAO1	Gammaproteobacteria
Pseudomonas fluorescens Pf-5	Gammaproteobacteria
Pseudomonas fluorescens PfO-1	Gammaproteobacteria
Pseudomonas putida KT2440	Gammaproteobacteria
Pseudomonas syringae pv. phaseolicola 1448A	Gammaproteobacteria
Pseudomonas syringae pv. syringae B728a	Gammaproteobacteria
Pseudomonas syringae pv. tomato str. DC3000	Gammaproteobacteria
Psychrobacter arcticus 273-4	Gammaproteobacteria
Ralstonia eutropha JMP134	Betaproteobacteria
Ralstonia solanacearum GMI1000	Betaproteobacteria
Rhizobium etli CFN 42	Alphaproteobacteria
Rhodobacter sphaeroides 2.4.1	Alphaproteobacteria
Rhodoferax ferrireducens DSM 15236	Betaproteobacteria
Rhodopirellula baltica SH1 *	Planctomycetacia
Rhodopseudomonas palustris CGA009	Alphaproteobacteria
Rhodopseudomonas palustris HaA2	Alphaproteobacteria
Rhodospirillum rubrum ATCC 11170	Alphaproteobacteria
Rickettsia conorii str. Malish 7	Alphaproteobacteria
Rickettsia felis URRWXCa12	Alphaproteobacteria
Rickettsia prowazekii str. Madrid E	Alphaproteobacteria
Rickettsia typhi str. Wilmington	Alphaproteobacteria
Salinibacter ruber DSM 13855	Bacteroidetes
Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67	Gammaproteobacteria
Salmonella enterica subsp. enterica serovar Paratyphi A str. ATCC 9150	Gammaproteobacteria
Salmonella enterica subsp. enterica serovar Typhi Ty2	Gammaproteobacteria

Salmonella enterica subsp. enterica serovar Typhi str. CT18	Gammaproteobacteria
Salmonella typhimurium LT2	Gammaproteobacteria
Shewanella oneidensis MR-1	Gammaproteobacteria
Shigella boydii Sb227	Gammaproteobacteria
Shigella dysenteriae Sd197	Gammaproteobacteria
Shigella flexneri 2a str. 2457T	Gammaproteobacteria
Shigella flexneri 2a str. 301	Gammaproteobacteria
Shigella sonnei Ss046	Gammaproteobacteria
Silicibacter pomeroyi DSS-3	Alphaproteobacteria
Sinorhizobium meliloti 1021	Alphaproteobacteria
Sodalis glossinidius str. 'morsitans'	Gammaproteobacteria
Solibacter usitatus Ellin6076 *	Acidobacteria/Solibacteres
Staphylococcus aureus RF122	Firmicutes/Bacilli
Staphylococcus aureus subsp. aureus COL	Firmicutes/Bacilli
Staphylococcus aureus subsp. aureus MRSA252	Firmicutes/Bacilli
Staphylococcus aureus subsp. aureus MSSA476	Firmicutes/Bacilli
Staphylococcus aureus subsp. aureus MW2	Firmicutes/Bacilli
Staphylococcus aureus subsp. aureus Mu50	Firmicutes/Bacilli
Staphylococcus aureus subsp. aureus N315	Firmicutes/Bacilli
Staphylococcus aureus subsp. aureus NCTC 8325	Firmicutes/Bacilli
Staphylococcus aureus subsp. aureus USA300	Firmicutes/Bacilli
Staphylococcus epidermidis ATCC 12228	Firmicutes/Bacilli
Staphylococcus epidermidis RP62A	Firmicutes/Bacilli
Staphylococcus haemolyticus JCSC1435	Firmicutes/Bacilli
Staphylococcus saprophyticus subsp. saprophyticus ATCC 15305	Firmicutes/Bacilli
Streptococcus agalactiae 2603V/R	Firmicutes/Bacilli

Streptococcus agalactiae A909	Firmicutes/Bacilli
Streptococcus agalactiae NEM316	Firmicutes/Bacilli
Streptococcus mutans UA159	Firmicutes/Bacilli
Streptococcus pneumoniae R6	Firmicutes/Bacilli
Streptococcus pneumoniae TIGR4	Firmicutes/Bacilli
Streptococcus pyogenes M1 GAS	Firmicutes/Bacilli
Streptococcus pyogenes MGAS10394	Firmicutes/Bacilli
Streptococcus pyogenes MGAS315	Firmicutes/Bacilli
Streptococcus pyogenes MGAS5005	Firmicutes/Bacilli
Streptococcus pyogenes MGAS6180	Firmicutes/Bacilli
Streptococcus pyogenes MGAS8232	Firmicutes/Bacilli
Streptococcus pyogenes SSI-1	Firmicutes/Bacilli
Streptococcus thermophilus CNRZ1066	Firmicutes/Bacilli
Streptococcus thermophilus LMG 18311	Firmicutes/Bacilli
Streptomyces avermitilis MA-4680	Actinobacteria
Streptomyces coelicolor A3 (2)	Actinobacteria
Symbiobacterium thermophilum IAM 14863	Actinobacteria
Synechococcus elongatus PCC 6301	Cyanobacteria
Synechococcus elongatus PCC 7942	Cyanobacteria
Synechococcus sp. CC9605	Cyanobacteria
Synechococcus sp. CC9902	Cyanobacteria
Synechococcus sp. JA-2-3B'a (2-13)	Cyanobacteria
Synechococcus sp. JA-3-3Ab	Cyanobacteria
Synechococcus sp. WH 8102	Cyanobacteria
Synechocystis sp. PCC 6803	Cyanobacteria
Thermoanaerobacter tengcongensis MB4	Firmicutes/Clostridia

Thermobifida fusca YX	Actinobacteria
Thermosynechococcus elongatus BP-1	Cyanobacteria
Thermotoga maritima MSB8 *	Thermotogae
Thermus thermophilus HB27	Deinococci
Thermus thermophilus HB8	Deinococci
Thiobacillus denitrificans ATCC 25259	Betaproteobacteria
Thiomicrospira crunogena XCL-2	Gammaproteobacteria
Thiomicrospira denitrificans ATCC 33889	Espilonproteobacteria
Treponema denticola ATCC 35405	Spirochaetes
Treponema pallidum subsp. pallidum str. Nichols	Spirochaetes
Tropheryma whipplei TW08/27	Actinobacteria
Ureaplasma parvum serovar 3 str. ATCC 700970	Firmicutes/Mollicutes
Vibrio cholerae O1 biovar eltor str. N16961	Gammaproteobacteria
Vibrio fischeri ES114	Gammaproteobacteria
Vibrio parahaemolyticus RIMD 2210633	Gammaproteobacteria
Vibrio vulnificus CMCP6	Gammaproteobacteria
Vibrio vulnificus YJ016	Gammaproteobacteria
Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis	Gammaproteobacteria
Wolbachia	Alphaproteobacteria
Wolinella succinogenes DSM 1740	Epsilonproteobacteria
Xanthomonas axonopodis pv. citri str. 306	Gammaproteobacteria
Xanthomonas campestris pv. campestris str. 8004	Gammaproteobacteria
Xanthomonas campestris pv. campestris str. ATCC 33913	Gammaproteobacteria
Xanthomonas campestris pv. vesicatoria str. 85-10	Gammaproteobacteria
Xanthomonas oryzae pv. oryzae KACC10331	Gammaproteobacteria
Xylella fastidiosa 9a5c	Gammaproteobacteria

Xylella fastidiosa Temecula1	Gammaproteobacteria
Yersinia pestis CO92	Gammaproteobacteria
Yersinia pestis KIM	Gammaproteobacteria
Yersinia pestis biovar Medievalis str. 91001	Gammaproteobacteria
Yersinia pseudotuberculosis IP 32953	Gammaproteobacteria
Zymomonas mobilis subsp. Mobilis ZM4	Alphaproteobacteria
ARCHAEBACTERIA	
Aeropyrum pernix K1	Crenarchaeota/Thermoprotei
Archaeoglobus fulgidus DSM 4304 *	Euryarchaeota/Archaeoglobi
Haloarcula marismortui ATCC 43049 *	Euryarchaeota/Halobacteria
Halobacterium sp. NRC-1	Euryarchaeota/Halobacteria
Methanocaldococcus jannaschii DSM 2661 *	Euryarchaeota/Methanococci
Methanococcus maripaludis S2	Euryarchaeota/Methanococci
Methanopyrus kandleri AV19 *	Euryarchaeota/Methanopyri
Methanosarcina acetivorans C2A	Euryarchaeota/Methanomicrobia
Methanosarcina barkeri str. Fusaro	Euryarchaeota/Methanomicrobia
Methanosarcina mazei Go1 *	Euryarchaeota/Methanomicrobia
Methanosphaera stadmanae DSM 3091 *	Euryarchaeota/Methanobacteria
Methanospirillum hungatei JF-1	Euryarchaeota/Methanomicrobia
Methanothermobacter thermoautotrophicus str. Delta H	Euryarchaeota/Methanobacteria
Nanoarchaeum equitans Kin4-M *	Nanoarchaeota
Natronomonas pharaonis DSM 2160	Euryarchaeota/Halobacteria
Picrophilus torridus DSM 9790 *	Euryarchaeota/Thermoplasmata
Pyrobaculum aerophilum str. IM2	Crenarchaeota/Thermococci
Pyrococcus abyssi GE5 *	Euryarchaeota/Thermococci

Pyrococcus furiosus DSM 3638	Euryarchaeota/Thermococci
Pyrococcus horikoshii OT3	Euryarchaeota/Thermococci
Sulfolobus acidocaldarius DSM 639	Crenarchaeota/Thermoprotei
Sulfolobus solfataricus P2 *	Crenarchaeota/Thermoprotei
Sulfolobus tokodaii str. 7	Crenarchaeota/Thermoprotei
Thermococcus kodakarensis KOD1	Euryarchaeota/Thermococci
Thermoplasma acidophilum DSM 1728	Euryarchaeota/Thermoplasmata
Thermoplasma volcanium GSS1	Euryarchaeota/Thermoplasmata

Table S2 List of Eubacteria and Archaebacteria species used in the ribosomal RNA data set (shared by SSU and LSU) and their classification. Species used in the Bayesian analysis are marked with an asterisk.

Species	**Classification**
EUBACTERIA	
Acetobacter europaeus AJ012698 *	Alphaproteobacteria
Acetobacter intermedius AJ012697	Alphaproteobacteria
Acetobacter xylinum X75619	Alphaproteobacteria
Acinetobacter calcoaceticus M34139 *	Gammaproteobacteria
Aeromonas hydrophila AF099021	Gammaproteobacteria
Agrobacterium radiobacter AJ130719	Alphaproteobacteria
Agrobacterium rubi D12787	Alphaproteobacteria
Agrobacterium tumefaciens D12784	Alphaproteobacteria
Agrobacterium vitis D12795	Alphaproteobacteria
Alcaligenes faecalis AF155147 *	Betaproteobacteria
Aquifex aeolicus AE000751 *	Aquificae
Bacillus alcalophilus AF078812 *	Firmicutes/Bacilli
Bacillus anthracis AF155951	Firmicutes/Bacilli
Bacillus cereus AF155952	Firmicutes/Bacilli
Bacillus globisporus X68415	Firmicutes/Bacilli
Bacillus halodurans D AP001507	Firmicutes/Bacilli
Bacillus licheniformis AF234844	Firmicutes/Bacilli
Bacillus stearothermophilus AJ005760	Firmicutes/Bacilli
Bacillus subtilis B K00637	Firmicutes/Bacilli
Bacillus thuringiensis AF155954	Firmicutes/Bacilli

Bartonella bacilliformis M65249	Alphaproteobacteria
Bordetella avium AF177666	Betaproteobacteria
Bordetella bronchiseptica U04948	Betaproteobacteria
Bordetella parapertussis U04949	Betaproteobacteria
Bordetella pertussis AF142326	Betaproteobacteria
Borrelia burgdorferi X85202 *	Spirochaetes
Bradyrhizobium japonicum Z35330	Alphaproteobacteria
Bradyrhizobium lupini U69636	Alphaproteobacteria
Brevundimonas diminuta AB021415	Alphaproteobacteria
Brucella melitensis AF220148	Alphaproteobacteria
Buchnera aphidicola L18927	Gammaproteobacteria
Burkholderia gladioli AB012916	Betaproteobactria
Burkholderia mallei AF110187	Betaproteobactria
Burkholderia pseudomallei	Betaproteobactria
Campylobacter coli L04312 *	Epsilonproteobacteria
Campylobacter hyoilei L19738	Epsilonproteobacteria
Campylobacter jejuni AL139074	Epsilonproteobacteria
Campylobacter lari L04316	Epsilonproteobacteria
Carsonella ruddii AF211123	Gammaproteobacteria
Chlamydia muridarum aA16S AE002280 *	Chlamydiae
Chlamydia trachomatis AE001347	Chlamydiae
Chlamydophila abortus U76710	Chlamydiae
Chlamydophila felis U68457	Chlamydiae
Chlamydophila pecorum U68434	Chlamydiae
Chlamydophila pneumoniae aA16S AE002256	Chlamydiae
Chlamydophila psittaci U68447	Chlamydiae
Chlorobium limicola Y10640 *	Chlorobia

Citrobacter freundii AJ233408	Gammaproteobacteria
Clostridium botulinum A L37586 *	Firmicutes/Clostridia
Clostridium histolyticum M59094	Firmicutes/Clostridia
Clostridium tyrobutyricum L08062	Firmicutes/Clostridia
Coxiella burnetii D89791	Gammaproteobacteria
Enterococcus faecalis AB012212	Firmicutes/Bacilli
Erysipelothrix rhusiopathiae AB034200 *	Firmicutes/Mollicutes
Erysipelothrix tonsillarum AB034201	Firmicutes/Mollicutes
Escherichia coli B AE000471	Gammaproteobacteria
Fibrobacter succinogenes M62683 *	Fibrobacteres
Flavobacterium odoratum D14019 *	Bacteroidetes/Flavobacteria
Flexibacter flexilis M62794 *	Bacteroidetes/Sphingobacteria
Frankia sp. M55343 *	Actinobacteria
Haemophilus influenzae D U32847	Gammaproteobacteria
Helicobacter pylori A AE000620	Epsilonproteobacteria
Klebsiella pneumoniae AB004753	Gammaproteobacteria
Lactobacillus amylolyticus Y17361	Firmicutes/Bacilli
Lactobacillus confusus M23036	Firmicutes/Bacilli
Lactobacillus delbrueckii AB007908	Firmicutes/Bacilli
Lactococcus lactis X64887	Firmicutes/Bacilli
Leptospira interrogans M71241	Spirochaetes
Leuconostoc carnosum AB022925	Firmicutes/Bacilli
Leuconostoc lactis M23031	Firmicutes/Bacilli
Leuconostoc mesenteroides AB023243	Firmicutes/Bacilli
Leuconostoc oenos M35820	Firmicutes/Bacilli
Leuconostoc paramesenteroides M23033	Firmicutes/Bacilli
Leucothrix mucor X87277	Gammaproteobacteria

Listeria grayi X56150	Firmicutes/Bacilli
Listeria innocua S55473	Firmicutes/Bacilli
Listeria ivanovii X98529	Firmicutes/Bacilli
Listeria monocytogenes U84150	Firmicutes/Bacilli
Listeria murrayi X56154	Firmicutes/Bacilli
Listeria seeligeri X56148	Firmicutes/Bacilli
Listeria welshimeri X56149	Firmicutes/Bacilli
Microbispora bispora U58524	Actinobacteria
Micrococcus luteus AF234843	Actinobacteria
Mycobacterium avium M29573	Actinobacteria
Mycobacterium kansasii M29575	Actinobacteria
Mycobacterium leprae X55022	Actinobacteria
Mycobacterium paratuberculosis M61680	Actinobacteria
Mycobacterium phlei M29566	Actinobacteria
Mycobacterium smegmatis AJ131761	Actinobacteria
Mycobacterium tuberculosis X55588	Actinobacteria
Mycoplasma flocculare X63377	Firmicutes/Mollicutes
Mycoplasma gallisepticum L08897	Firmicutes/Mollicutes
Mycoplasma genitalium A16S U39694	Firmicutes/Mollicutes
Mycoplasma hyopneumoniae Y00149	Firmicutes/Mollicutes
Nannocystis exedens AJ233946*	Deltaproteobacteria
Neisseria gonorrhoeae AF146369	Betaproteobacteria
Neisseria meningitidis AF059671	Betaproteobacteria
Paracoccus denitrificans AJ288159	Alphaproteobacteria
Peptococcus niger X55797	Firmicutes/Clostridia
Pirellula marina X62912 *	Planctomycetacia
Plesiomonas shigelloides M59159	Gammaproteobacteria

Propionibacterium freudenreichi AJ009989	Actinobacteria
Pseudomonas aeruginosa AF023658	Gammaproteobacteria
Pseudomonas fluorescens AF068010	Gammaproteobacteria
Pseudomonas stutzeri AF038653	Gammaproteobacteria
Ralstonia pickettii AB004790	Betaproteobacteria
Ralstonia solanacearum AB024604	Betaproteobacteria
Renibacterium salmoninarum AB017538	Actinobacteria
Rhizobium galegae AF025853	Alphaproteobacteria
Rhizobium leguminosarum D12782	Alphaproteobacteria
Rhizobium tropici D11344	Alphaproteobacteria
Rhodobacter capsulatus D13474	Alphaproteobacteria
Rhodobacter sphaeroides B X53854	Alphaproteobacteria
Rhodococcus erythropolis AJ237967	Actinobacteria
Rhodococcus fascians X81932	Actinobacteria
Rhodopseudomonas palustris AB017261	Alphaproteobacteria
Rhodospirillum rubrum D30778	Alphaproteobacteria
Rickettsia akari L36099	Alphaproteobacteria
Rickettsia australis L36101	Alphaproteobacteria
Rickettsia bellii L36103	Alphaproteobacteria
Rickettsia canada L36104	Alphaproteobacteria
Rickettsia conorii L36105	Alphaproteobacteria
Rickettsia parkeri L36673	Alphaproteobacteria
Rickettsia prowazekii AJ235272	Alphaproteobacteria
Rickettsia rhipicephali L36216	Alphaproteobacteria
Rickettsia rickettsii U11021	Alphaproteobacteria
Rickettsia sibirica D38628	Alphaproteobacteria
Rickettsia typhi L36221	Alphaproteobacteria

Ruminobacter amylophilus AB004908	Gammaproteobacteria
Salmonella typhi U88545	Gammaproteobacteria
Serpulina hyodysenteriae U14931	Spirochaetes
Serpulina innocens U14924	Spirochaetes
Simkania negevensis U68460	Chlamydiae
Staphylococcus aureus AF076030	Firmicutes/Bacilli
Staphylococcus carnosus AB009934	Firmicutes/Bacilli
Staphylococcus condimenti Y15750	Firmicutes/Bacilli
Staphylococcus piscifermentans Y15754	Firmicutes/Bacilli
Stigmatella aurantiaca AJ233935	Deltaproteobacteria
Streptococcus macedonicus Z94012	Firmicutes/Bacilli
Streptococcus oralis S70359	Firmicutes/Bacilli
Streptococcus parauberis X89967	Firmicutes/Bacilli
Streptococcus thermophilus X59028	Firmicutes/Bacilli
Streptococcus uberis AB002527	Firmicutes/Bacilli
Streptomyces ambofaciens M27245	Actinobacteria
Streptomyces coelicolor A AL356612	Actinobacteria
Streptomyces griseus B AB030568	Actinobacteria
Streptomyces lividans AB037565	Actinobacteria
Streptomyces rimosus F X62884	Actinobacteria
Synechocystis sp. D64000 *	Cyanobacteria
Thermomonospora chromogena AF002261	Actinobacteria
Thermotoga maritima aA16S AE001703 *	Thermotogae
Thermus thermophilus L09659 *	Deinococcus-Thermus
Treponema pallidum AE001208	Spirochaetes
Tropheryma whippelii AF190688	Actinobacteria
Ureaplasma urealyticum AE002127	Firmicutes/Mollicutes

Vibrio cholerae AE004096	Gammaproteobacteria
Vibrio vulnificus X56582	Gammaproteobacteria
Waddlia chondrophila AF042496	Chlamydiae
Wolbachia pipientis AF179630	Alphaproteobacteria
Wolinella succinogenes M26636	Epsilonproteobacteria
Xylella fastidiosa aA16S AE003870	Gammaproteobacteria
Yersinia enterocolitica M59292	Gammaproteobacteria
Zoogloea ramigera D14254	Betaproteobacteria
Zymobacter palmae AF211871	Gammaproteobacteria
Zymomonas mobilis C AF117351	Alphaproteobacteria
ARCHAEBACTERIA	
Aeropyrum pernix AB019552 *	Crenarchaeota/Thermoprotei
Archaeoglobus fulgidus AE000965 *	Euryarchaeota/Archaeoglobi
Desulfurococcus mobilis M36474	Crenarchaeota/Thermoprotei
Haloarcula marismortui AF034620 *	Euryarchaeota/Halobacteria
Halobacterium halobium AJ002949	Euryarchaeota/Halobacteria
Halobacterium marismortui X61689	Euryarchaeota/Halobacteria
Halococcus morrhuae D11106	Euryarchaeota/Halobacteria
Haloferax mediterranei D11107	Euryarchaeota/Halobacteria
Methanobacterium thermoautotrop AE000940 *	Euryarchaeota/Methanobacteria
Methanococcus jannaschii B U67517 *	Euryarchaeota/Methanococci
Methanococcus vannielii M36507	Euryarchaeota/Methanococci
Methanopyrus kandleri *	Euryarchaeota/Methanopyri
Methanospirillum hungatei M60880 *	Euryarchaeota/Methanomicrobia
Nanoarchaeum equitans *	Nanoarchaeota
Natronobacterium magadii X72495	Euryarchaeota/Halobacteria

Pyrobaculum islandicum L07511	Crenarchaeota/Thermoprotei
Pyrococcus abyssi AJ248283 *	Euryarchaeota/Thermococci
Pyrococcus horikoshii AP000001	Euryarchaeota/Thermococci
Sulfolobus acidocaldarius U05018	Crenarchaeota/Thermoprotei
Sulfolobus shibatae M32504	Crenarchaeota/Thermoprotei
Sulfolobus solfataricus X90483	Crenarchaeota/Thermoprotei
Thermococcus celer M21529	Euryarchaeota/Thermococci
Thermofilum pendens X14835	Crenarchaeota/Thermoprotei
Thermoplasma acidophilum M38637 *	Euryarchaeota/Thermoplasmata

Table S3 Total number of species per group (source: DSMZ, NCBI, Algaebase). P: phylum; C. Class.

EUBACTERIA	Total number of species
Acidobacteria (p)	4
Actinobacteria (p, c)	1784
Alphaproteobacteria (c)	711
Aquificae (p, c)	22
Bacilli (c)	845
Bacteroidetes (p)	493
Betaproteobacteria (c)	373
Chlamydiae (p, c)	13
Chlorobia (p, c)	17
Chloroflexi (p)	45
Clostridia (c)	578
Cyanobacteria (p)	2654
Deinococci (c)	45
Deltaproteobacteria (c)	226
Epsilonproteobacteria (c)	77
Fibrobacteres (p, c)	2
Fusobacteria (p, c)	37
Gammaproteobacteria (c)	1177
Mollicutes (c)	204
Planctomycetes (p)	12
Spirochaetes (p, c)	98
Thermolithobacteria (c)	2
Thermotogae (p, c)	30

ARCHAEBACTERIA	
Archaeoglobi (c)	5
Halobacteria (c)	82
Methanobacteria (c)	37
Methanococci (c)	13
Methanomicrobia (c)	61
Methanopyri (c)	1
Nanoarchaeota (p)	1
Thermococci (c)	33
Thermoplasmata (c)	5
Thermoprotei (c)	53

Table S4 Habitat preference of families in Group-I phyla. Symbols: t, terrestrial; m, marine; m/t, marine and terrestrial. Bacilli, Clostridia, and Mollicutes are treated at the class level and have been conservatively coded as m/t (most classes within Clostridia and Mollicutes are strictly terrestrial while Bacilli colonize both habitats).

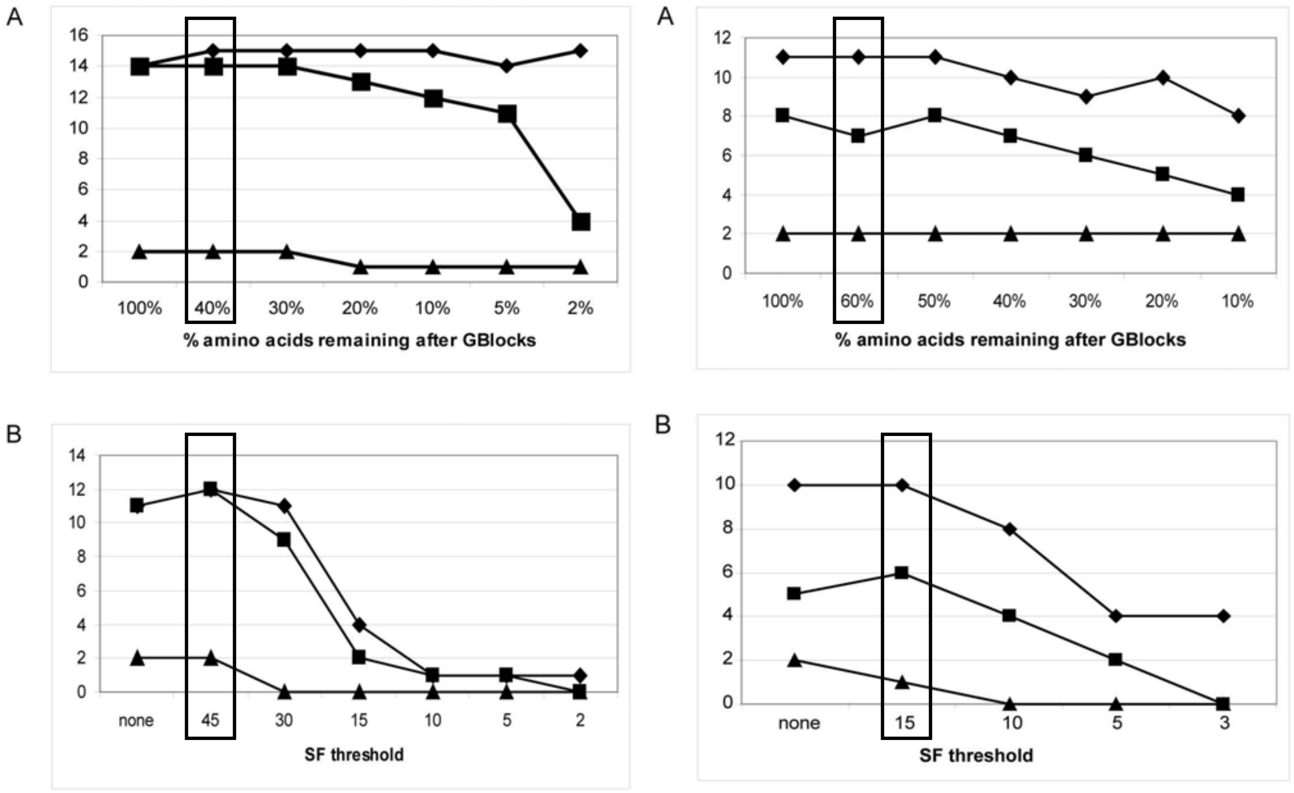
Phylum	Family	Habitat
Actinobacteria	Acidimicrobiaceae	m/t
	Acidothermaceae	t
	Actinomycetaceae	t
	Actinospicaceae	t
	Actinosynnemataceae	t
	Beutenbergiaceae	t
	Bogoriellaceae	t
	Brevibacteriaceae	t
	Catenulisporaceae	t
	Corynebacteriaceae	t
	Dermabacteraceae	t
	Dermacoccaceae	m/t
	Dermatophilaceae	t
	Dietziaceae	t
	Frankiaceae	t
	Geodermatophilaceae	t
	Glycomycetaceae	t
	Gordoniaceae	t
	Intrasporangiaceae	m/t
	Jonesiaceae	t
	Kineosporiaceae	m/t
	Microbacteriaceae	m/t

Actinobacteria	Micrococcaceae	m/t
	Micromonosporaceae	m/t
	Mycobacteriaceae	t
	Nakamurellaceae	t
	Nocardiaceae	m/t
	Nocardioidaceae	m/t
	Promicromonosporaceae	m/t
	Propionibacteriaceae	m/t
	Pseudonocardiaceae	t
	Rarobacteraceae	t
	Sanguibacteraceae	t
	Segniliparaceae	t
	Sporichthyaceae	t
	Streptomycetaceae	m/t
	Streptosporangiaceae	t
	Thermomonosporaceae	t
	Tsukamurellaceae	m/t
	Williamsiaceae	m/t
	Yaniaceae	t
	Bifidobacteriaceae	t
	Coriobacteriaceae	t
	Conexibacteraceae	t
	Patulibacteraceae	t
	Rubrobacteraceae	t
	Solirubrobacteraceae	t
	Thermoleophilaceae	t
Bacilli		m/t

Chloroflexi	Chloroflexaceae	m/t
	Herpetosiphonaceae	t
	Thermomicrobiaceae	t
	Sphaerobacteraceae	t
	Dehalococcoides	t
	Anaerolinaceae	t
	Caldilinaceae	t
Clostridia		m/t
Cyanobacteria	Chroococcaceae	m/t
	Cyanobacteriaceae	m/t
	Dermocarpellaceae	m
	Entophysalidaceae	m/t
	Gloeobacteraceae	t
	Hydrococcaceae	m
	Microcystaceae	m/t
	Prochloraceae	m
	Xenococcaceae	m/t
	Chlorogloeopsidaceae	t
	Hapalosiphonaceae	t
	Microchaetaceae	m/t
	Nostocaceae	m/t
	Rivulariaceae	m/t
	Scytonemataceae	m/t
	Stigonemataceae	t
	Symphyonemataceae	m/t
	Oscillatoriaceae	m/t
	Phormidiaceae	m/t

	Schizotrichaceae	t
	Pseudanabenaceae	m/t
	Mastigocladaceae	t
	Chamaesiphonaceae	m/t
	Merismopediaceae	m/t
	Synechococcaceae	m
	Deinococcaceae	t
Deinococcus-Thermus	Trueperaceae	t
	Thermaceae	m/t
Mollicutes		m/t

Fig. S1 Effects of increasing GBlocks (panels A) and SF (panels B) stringencies on the phylogeny of the protein and rRNA data set. Diamonds: number of monophyletic eubacterial classes; Squares: number of significantly supported monophyletic classes; Triangles: number of monophyletic eubacterial phyla. Black rectangles show the selected stringency level.



Protein data set

rRNA data set

Fig. S2 Consensus of 25 single ML gene trees from the protein data set. Triangles are proportional to the number of sequences analyzed in each class. Numbers represent the percentage of genes supporting the cluster.

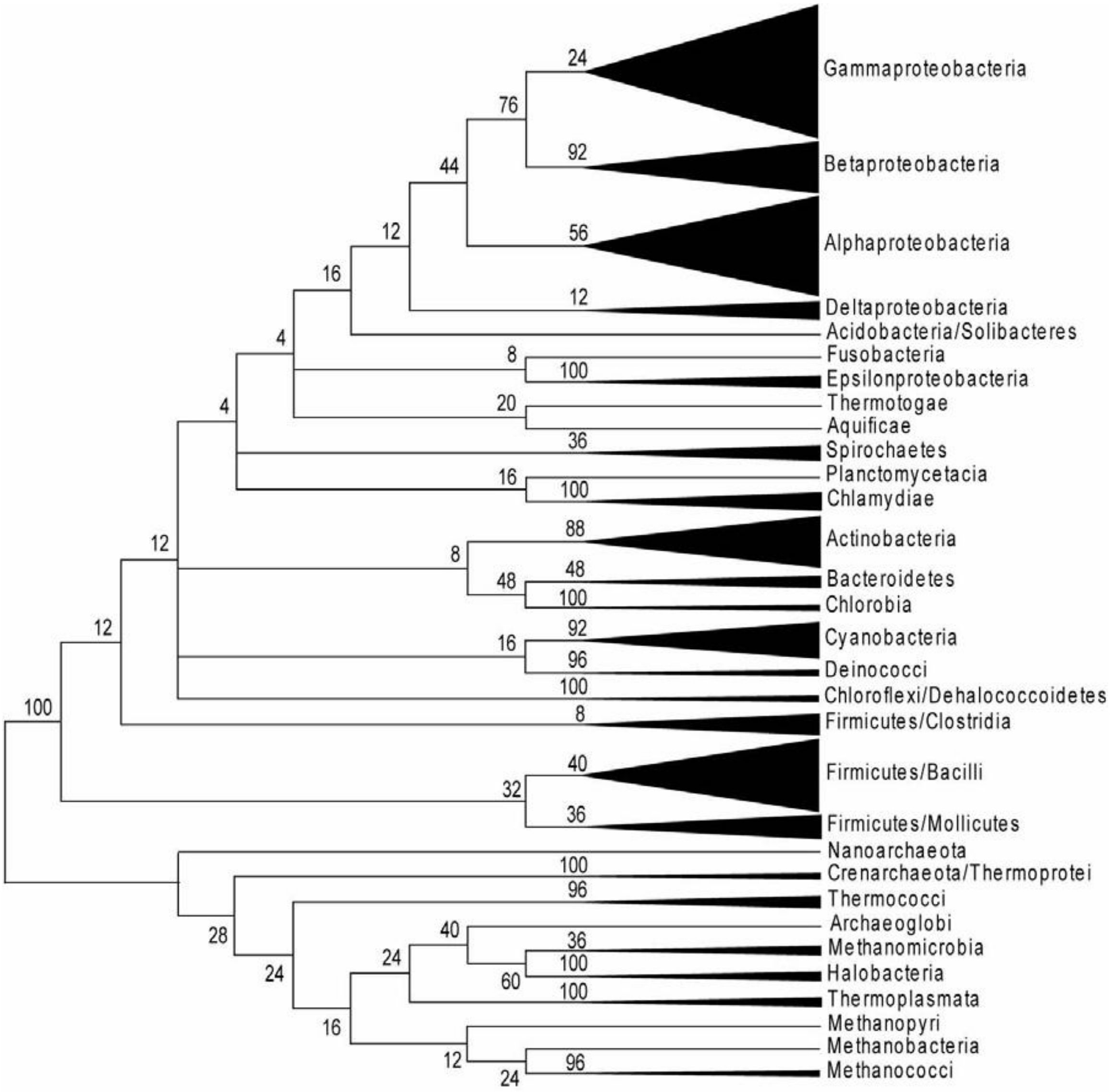


Fig. S3 Maximum likelihood phylogeny of slow evolving sites in the protein data set (Eubacteria and Archaebacteria). Asterisks: bootstrap values equal to or higher than 95%. Triangles are proportional to the number of sequences analyzed in each lineage. Values at each node are for 100 bootstrap replicates.

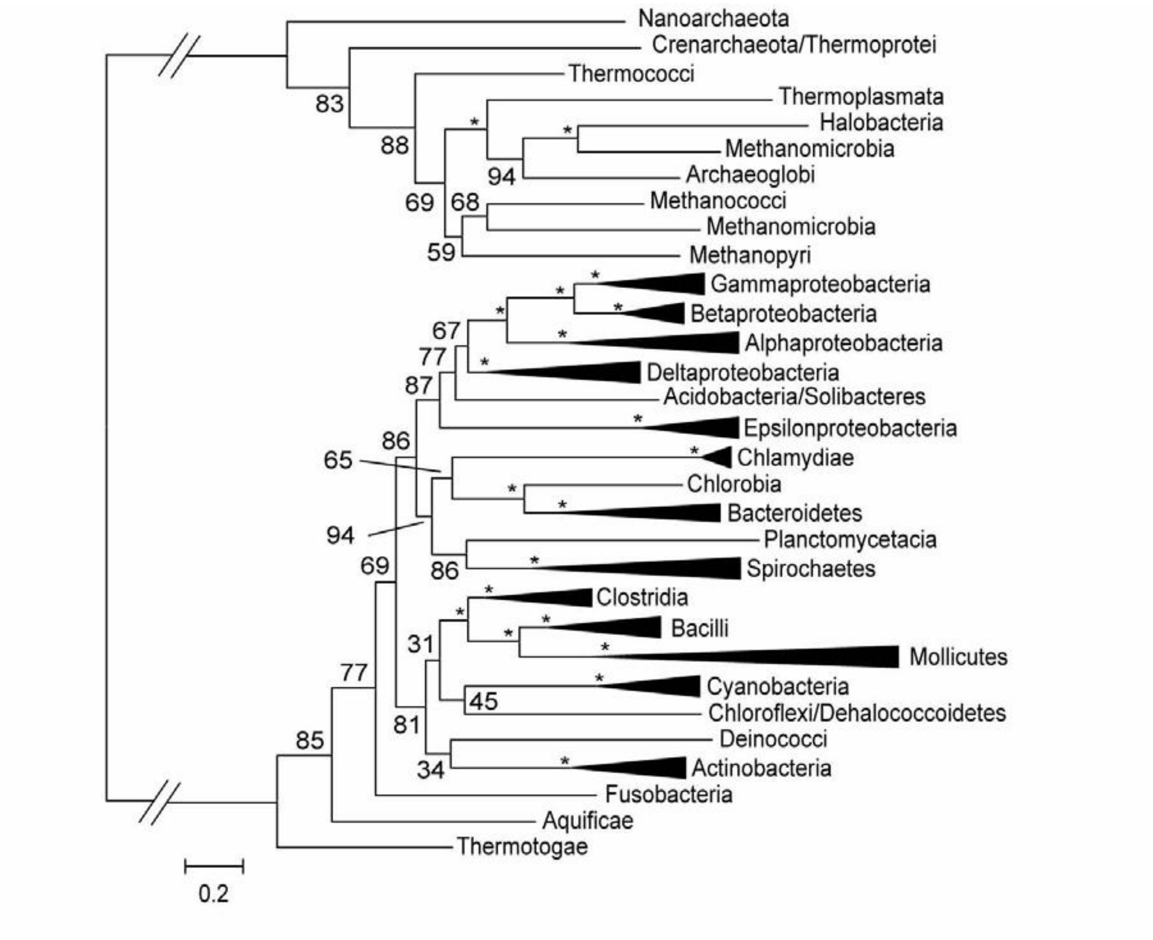


Fig. S4 LogDet phylogeny of rRNA (SSU+LSU) data set. Triangles are proportional to the number of sequences analyzed in each lineage. Values at each node are percentage support for 100 bootstrap replicates.

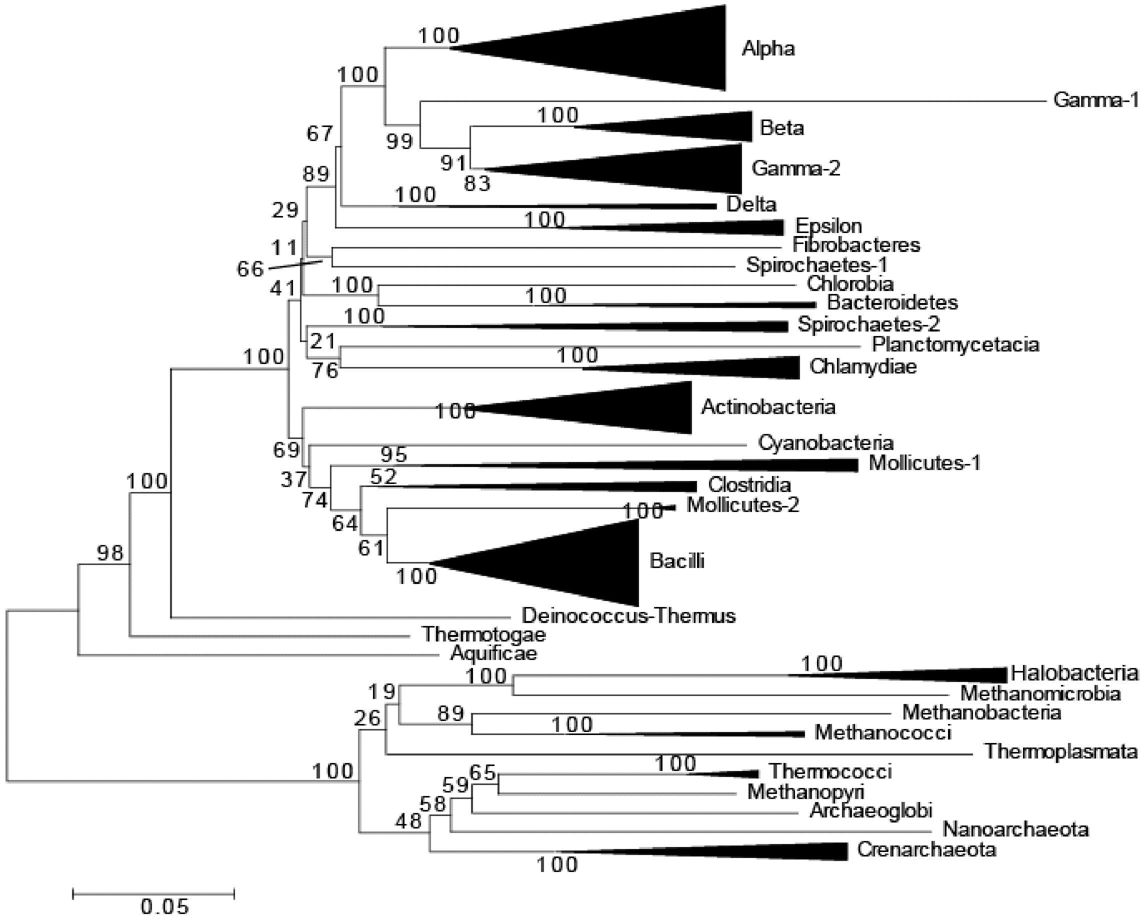


Fig. S5 Maximum likelihood phylogeny of slow evolving sites in the rRNA (SSU+LSU) data set (Eubacteria and Archaebacteria). Asterisks: bootstrap values equal to or higher than 95%. Triangles are proportional to the number of sequences analyzed in each lineage. Values at each node are for 100 bootstrap replicates.

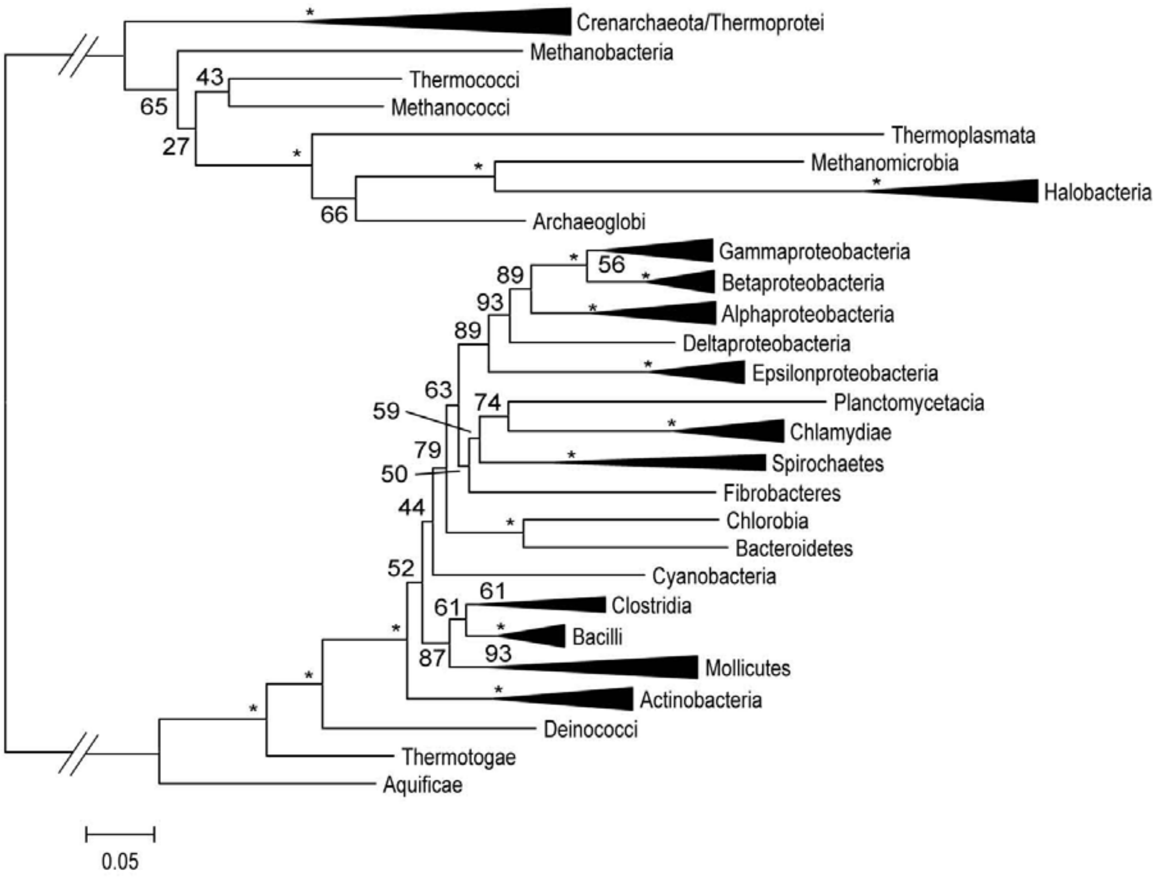


Fig. S6 Maximum parsimony ancestral states reconstruction in major lineages of Terrabacteria (Group-1). Terrestrial states (species) are shown in tan and marine states in blue; dashed lines indicate lineages in which there is at least one terrestrial and one marine species. The phylum-level topology of the tree and relationships within Firmicutes are from the ML protein analysis whereas the topology within other phyla (Actinobacteria, *Deinococcus-Thermus*, and Cyanobacteria) is from the ML SSU rRNA analysis. The phylogeny within Chloroflexi is from elsewhere (Costello and Schmidt 2006). The branch leading to Firmicutes is either terrestrial or mixed (assigned here conservatively as mixed). Each phylum is represented at the lowest determinable monophyletic taxonomic level beginning with family. Therefore, within a phylum if orders were not monophyletic then families were used; orders were used if they were monophyletic. Firmicutes are represented at the class level as in the protein data set.

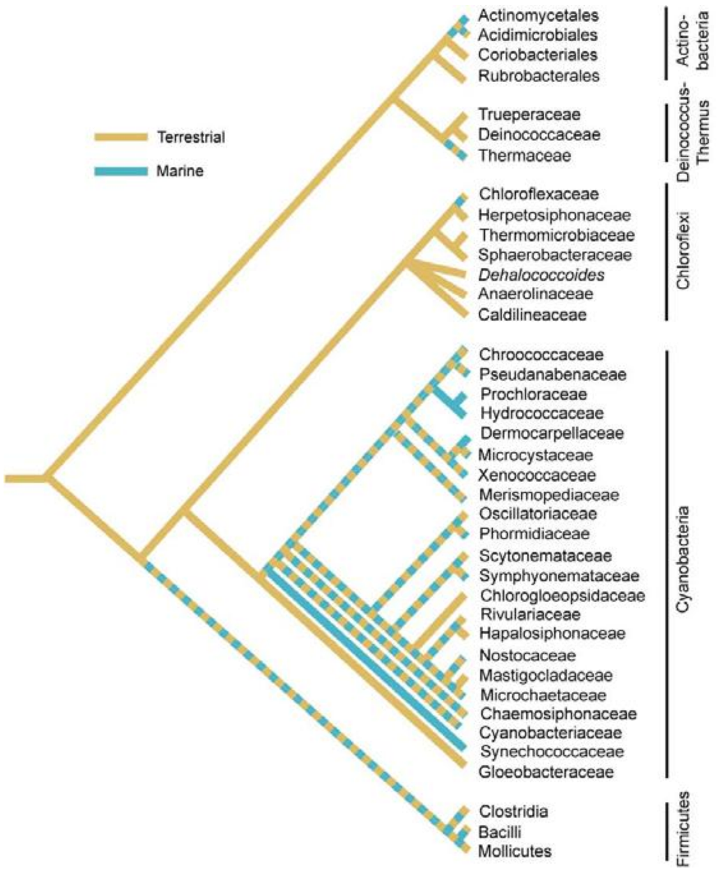
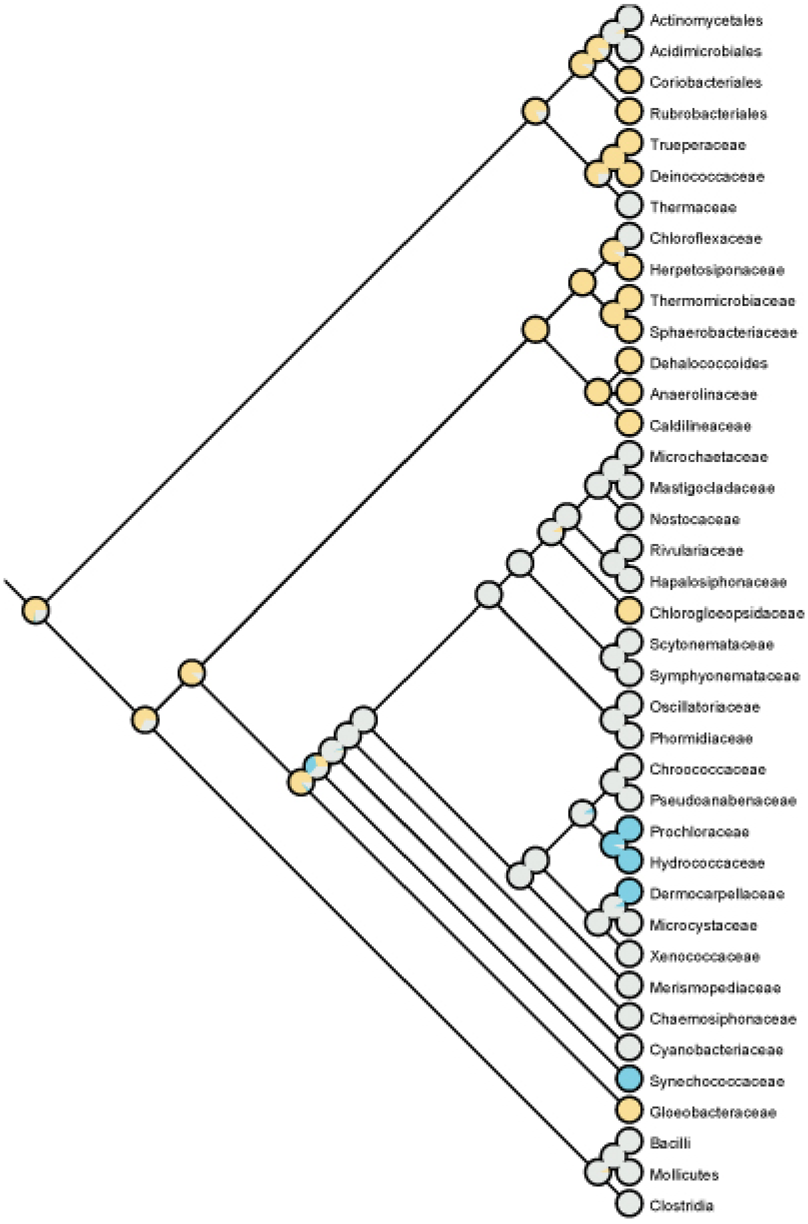


Fig. S7 Maximum likelihood ancestral states reconstruction of Terrabacteria (Group-I) lineages. Phylogenetic details are as in Fig. S6. Terrestrial state is shown in tan, marine state in blue, mixed state in gray. Probabilities of each state in the last common ancestor of the group are: 73% terrestrial, 24% mixed, and 3% marine.


Actinomycetales
Acidimicrobiales
Coriobacteriales
Rubrobacteriales
Trueperaceae
Deinococcaceae
Thermaceae
Chloroflexaceae
Herpetosiponaceae
Thermomicrobiaceae
Sphaerobacteriaceae
Dehalococcoides
Anaerolinaceae
Caldilineaceae
Microchaetaceae
Mastigocladaceae
Nostocaceae
Rivulariaceae
Hapalosiphonaceae
Chlorogloeopsidaceae
Scytonemataceae
Symphyonemataceae
Oscillatoriaceae
Phormidiaceae
Chroococcaceae
Pseudoanabenaceae
Prochloraceae
Hydrococcaceae
Dermocarpellaceae
Microcystaceae
Xenococcaceae
Merismopediaceae
Chaemosiphonaceae
Cyanobacteriaceae
Synechococcaceae
Gloeobacteraceae
Bacilli
Mollicutes
Clostridia