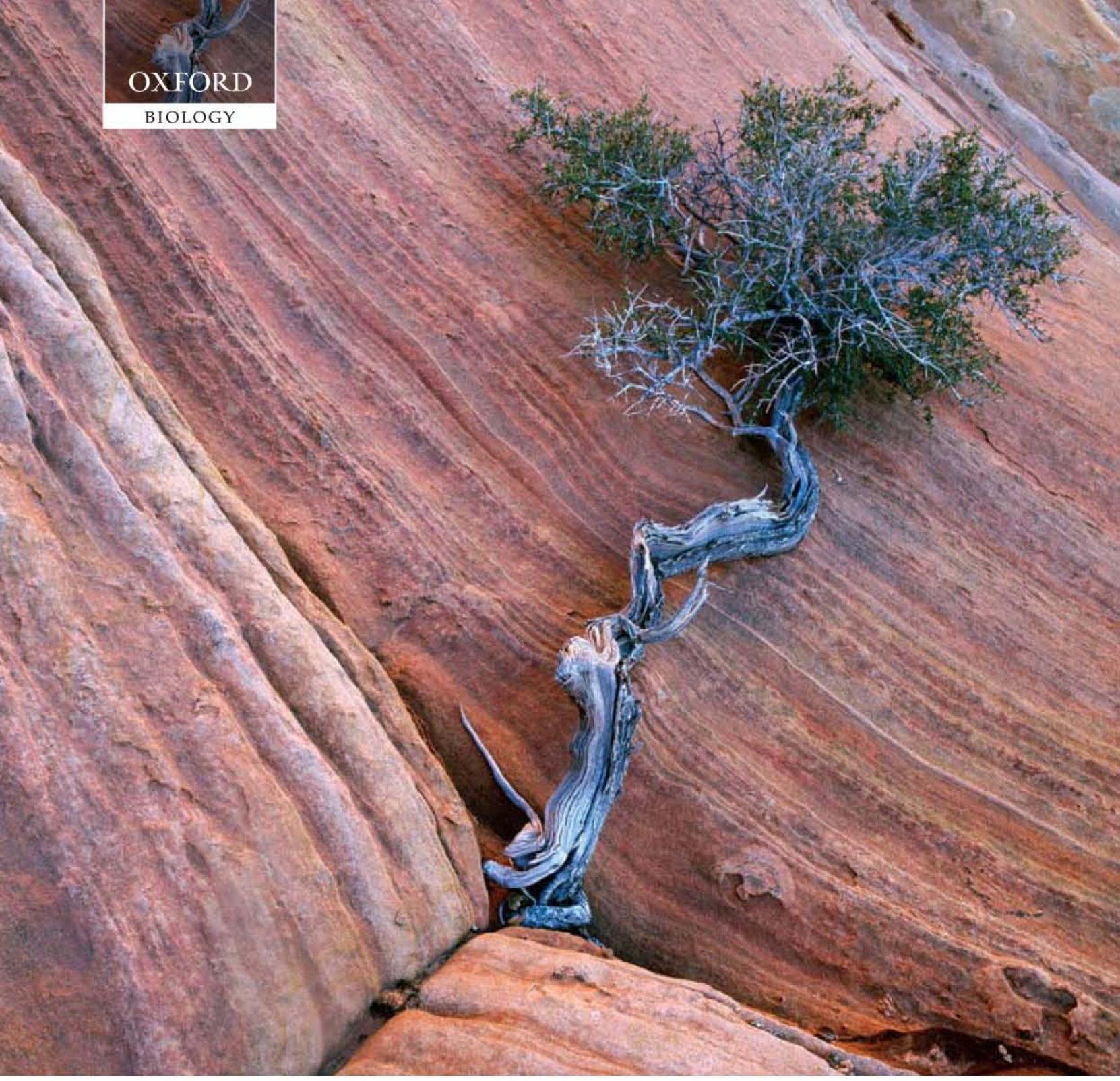

OXFORD
BIOLOGY

the TIMETREE *of* LIFE

edited by **S. BLAIR HEDGES** *and* **SUDHIR KUMAR**

foreword by James D. Watson

Eubacteria

Fabia U. Battistuzzi[a,b,*] *and S. Blair Hedges*[a]
Department of Biology, 208 Mueller Laboratory, The Pennsylvania State University, University Park, PA 16802-5301, USA; [b]Current address: Center for Evolutionary Functional Genomics, The Biodesign Institute, Arizona State University, Tempe, AZ 85287-5301, USA
*To whom correspondence should be addressed (Fabia.Battistuzzi@asu.edu)

Abstract

The ~9400 recognized species of prokaryotes in the Superkingdom Eubacteria are placed in 25 phyla. Their relationships have been difficult to establish, although some major groups are emerging from genome analyses. A molecular timetree, estimated here, indicates that most (85%) of the phyla and classes arose in the Archean Eon (4000–2500 million years ago, Ma) whereas most (95%) of the families arose in the Proterozoic Eon (2500–542 Ma). It also points to an early origin of phototrophy (3400–3200 Ma) followed by colonization of land (3041–2755 Ma), origin of oxygenic photosynthesis (2920–2592 Ma) and of aerobic methanotrophy (2630–2371 Ma).

The Superkingdom Eubacteria (also called "Bacteria," Fig. 1) is subdivided into 25 recognized phyla (*1*), although other classifications have been proposed (*2*) and putative lineages await taxonomic assignment (e.g., *3, 4*). In contrast to most eukaryotic groups, eubacterial phylogeny has been difficult to resolve at the highest level (i.e., phyla and class relationships) despite the availability of many complete genome sequences. The small subunit (SSU) ribosomal RNA (rRNA) gene has been widely used to study prokaryote phylogeny and first revealed the distinction of eubacteria from archaebacteria (Archaea) (*5, 6*). However, because of its limited length and the large genetic distances among prokaryote species, it is also subject to biases such as long-branch attraction, base composition differences, and a generally poor resolving power. These issues mostly have been addressed in more recent years by using genome-scale data sets (e.g., multiple genes, gene content, and insertion–deletion events) and increased taxonomic sampling (*7–19, 88*). While these studies have shown increasing support for lower-level phylogenetic clusters (e.g., classes and below), they have also shown the susceptibility of eubacterial phylogeny to biases such as horizontal gene transfer (HGT) (*20, 21*).

In recent years, three major approaches have been used for studying prokaryote phylogeny with data from complete genomes: (i) combining gene sequences in a single analysis of multiple genes (e.g., *7, 9, 10*), (ii) combining trees from individual gene analyses into a single "supertree" (e.g., *22, 23*), and (iii) using the presence or absence of genes ("gene content") as the raw data to investigate relationships (e.g., *17, 18*). While the results of these different approaches have not agreed on many details of relationships, there have been some points of agreement, such as support for the monophyly of all major classes and some phyla (e.g., Proteobacteria and Firmicutes). These findings, although criticized by some (e.g., *24, 25*), suggest the presence of a detectable evolutionary signal for prokaryotes when using carefully selected genes (e.g., vertically inherited) and appropriate methodologies (e.g., genome-scale data sets).

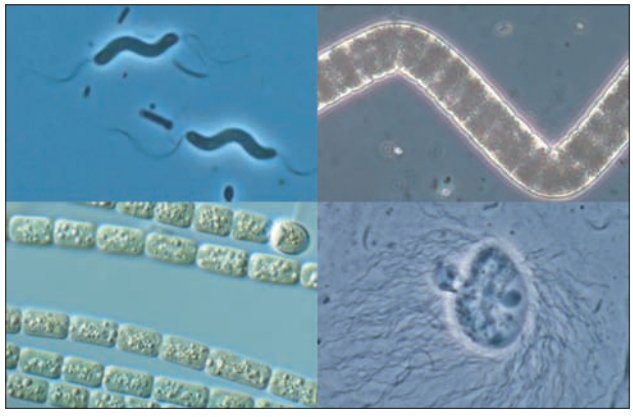

Fig. 1 *Spirillum*, a betaproteobacterium (upper left); *Beggiatoa*, a gammaproteobacterium (upper right); an unidentified cyanobacterium (lower left); and an unidentified spirochete (lower right). Credits: D. J. Patterson, L. Amaral-Zettler, and V. Edgcomb (upper left and right); D. J. Patterson (lower left); and L. Amaral-Zettler, L. Olendzenski, and D. J. Patterson (lower right). Images provided by micro*scope (http://microscope.mbl.edu) under creative commons license.

F. U. Battistuzzi and S. B. Hedges. Eubacteria. Pp. 106–115 in *The Timetree of Life*, S. B. Hedges and S. Kumar, Eds. (Oxford University Press, New York, 2009).

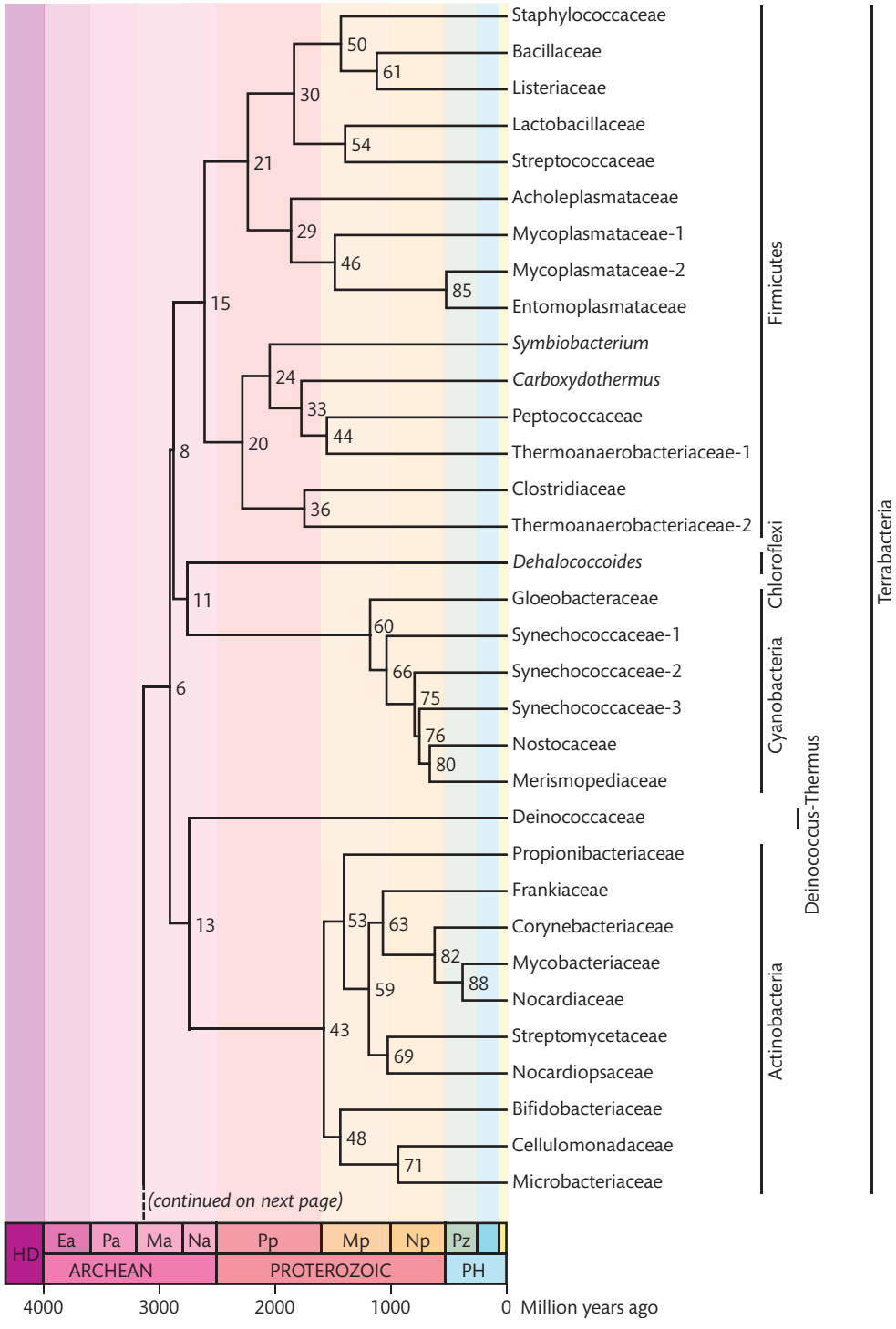

Fig. 2 *Continues*

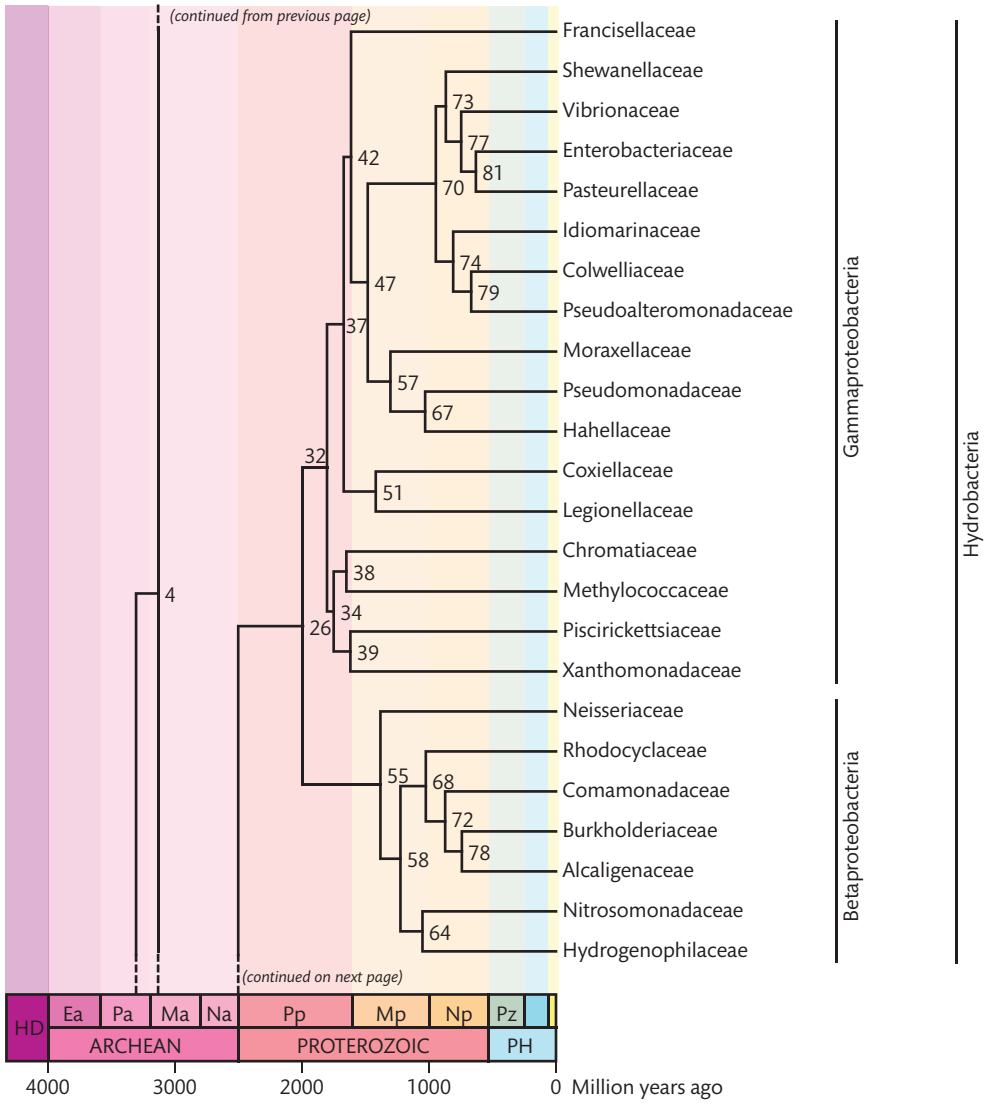

Fig. 2 *Continues*

Phylogenetic analyses that have focused on subgroups of prokaryotes (e.g., classes) also have supported, consistently, particular groupings. For example, cyanobacteria and low GC gram positives (Firmicutes) have been united based on maximum likelihood (ML) mapping of 21 orthologous genes (*26*); Fusobacteria is the most closely related group to Firmicutes based on combined analyses of SSU and large subunit (LSU) rRNA sequences and ribosomal proteins (*27*); Bacteroidetes, Chlorobi, and Fibrobacteres form a group based on insertion–deletion analysis (*28, 29*); and Planctomycetacia, Chlamydiae, and Spirochaetes form a group based on concatenated ribosomal proteins, DNA-directed RNA polymerase subunits (*30*), genome trees (*31*), and gene content analysis (*18*). Other taxa, such as Aquificae and Thermotogae, have been more difficult to place, appearing at the base of the tree in some analyses (*4, 7, 9, 32*) and in a higher, nested position in other analyses (*10, 33, 34*).

Generally concordant relationships were found between past studies and a recent ML and Bayesian analysis of multiple core gene sequences (88). In that study, a complete matrix of 25 vertically inherited orthologous genes shared by 197 fully sequenced eubacteria and 21 fully sequenced species of archaebacteria was built.

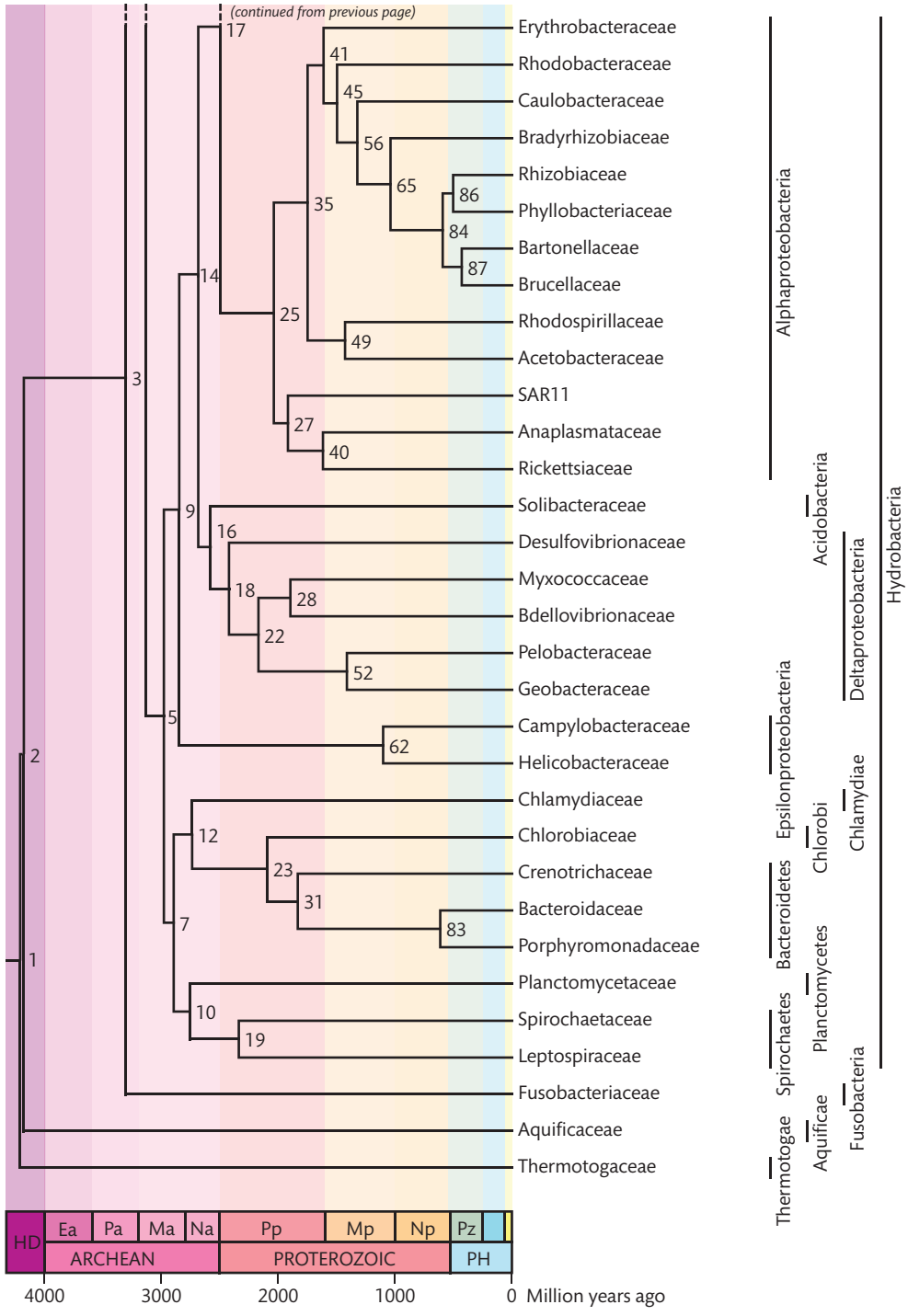

Fig. 2 A timetree of eubacteria. Divergence times are shown in Table 1. Codes for paraphyletic and/or polyphyletic groups are as follows: Mycoplasmataceae-1 (*Mycoplasma genitalium*), Mycoplasmataceae-2 (*Mycoplasma capricolum*), Synechococcaceae-1 (*Synechococcus* JA-2-3Ba), Synechococcaceae-2 (*Thermosynechococcus elongatus*), Synechococcaceae-3 (*Synechococcus elongatus*), Thermoanaerobacteriaceae-1 (Moorella thermoacetica), Thermoanaerobacteriaceae-2 (Thermoanaerobacter tengcongensis). *Abbreviations*: Ea (Eoarchean), HD (Hadean), Ma (Mesoarchean), Mp (Mesoproterozoic), Na (Neoarchean), Np (Neoproterozoic), Pa (Paleoarchean), PH (Phanerozoic), Pp (Paleoproterozoic), and Pz (Paleozoic).

Table 1. Divergence times (Ma) and their confidence/credibility intervals (CI) among eubacteria.

Timetree		Estimates				Timetree		Estimates			
Node	Time	Ref. (7)		This study		Node	Time	Ref. (7)		This study	
		Time	CI	Time	CI			Time	CI	Time	CI
1	4189	–	–	4189	4200–4159	45	1498	–	–	1498	1644–1351
2	4179	–	–	4179	4197–4141	46	1482	–	–	1482	1656–1313
3	3306	3186	3634–2801	3306	3447–3165	47	1481	1387	1763–1060	1481	1586–1372
4	3134	–	–	3134	3265–2987	48	1436	–	–	1436	1600–1270
5	2979	3096	3539–2723	2979	3116–2835	49	1432	–	–	1432	1580–1283
6	2908	–	–	2908	3041–2755	50	1429	1561	2012–1166	1429	1597–1269
7	2897	–	–	2897	3040–2747	51	1420	–	–	1420	1539–1297
8	2874	–	–	2874	3012–2716	52	1413	–	–	1413	1573–1257
9	2849	2800	3223–2452	2849	2983–2706	53	1402	–	–	1402	1557–1241
10	2762	–	–	2762	2926–2595	54	1392	–	–	1392	1573–1223
11	2761	–	–	2761	2920–2592	55	1386	1470	1873–1114	1386	1511–1262
12	2739	–	–	2739	2894–2575	56	1326	1537	1942–1151	1326	1471–1183
13	2739	–	–	2739	2897–2570	57	1306	–	–	1306	1420–1190
14	2687	–	–	2687	2815–2549	58	1224	–	–	1224	1349–1101
15	2607	2688	3108–2360	2607	2753–2448	59	1189	1380	1798–1038	1189	1332–1040
16	2579	–	–	2579	2715–2428	60	1180	–	–	1180	1334–1035
17	2504	2508	2928–2154	2504	2630–2371	61	1121	1282	1690–916	1121	1276–977
18	2421	–	–	2421	2563–2267	62	1104	1244	1694–856	1104	1271–943
19	2339	–	–	2339	2517–2154	63	1069	–	–	1069	1209–928
20	2281	–	–	2281	2439–2115	64	1055	–	–	1055	1184–932
21	2233	2305	2729–1944	2233	2401–2067	65	1042	–	–	1042	1180–912
22	2173	–	–	2173	2321–2018	66	1037	–	–	1037	1186–902
23	2099	–	–	2099	2261–1932	67	1030	–	–	1030	1145–917
24	2047	–	–	2047	2215–1873	68	1028	–	–	1028	1150–911
25	2042	2030	2449–1656	2042	2187–1887	69	1027	–	–	1027	1167–887
26	1993	1945	2368–1573	1993	2099–1894	70	950	–	–	950	1051–849
27	1919	–	–	1919	2083–1749	71	937	–	–	937	1084–794
28	1899	–	–	1899	2078–1719	72	872	–	–	872	993–759
29	1860	–	–	1860	2040–1680	73	871	–	–	871	972–772
30	1837	1816	2261–1415	1837	2011–1673	74	812	–	–	812	910–711
31	1834	–	–	1834	2005–1665	75	793	1039	1408–702	793	923–678
32	1806	1751	2163–1390	1806	1889–1731	76	751	–	–	751	880–639
33	1775	–	–	1775	1948–1602	77	751	–	–	751	854–653
34	1753	–	–	1753	1821–1690	78	744	–	–	744	859–640
35	1753	–	–	1753	1894–1605	79	668	–	–	668	768–573
36	1747	2132	2552–1760	1747	1931–1572	80	662	756	1070–484	662	783–553
37	1673	–	–	1673	1765–1581	81	634	–	–	634	739–538
38	1653	–	–	1653	1688–1640	82	621	928	1274–644	621	734–516
39	1621	–	–	1621	1722–1520	83	616	–	–	616	734–506
40	1620	–	–	1620	1792–1449	84	594	–	–	594	699–499

Table 1. *Continued*

Timetree		Estimates				Timetree		Estimates			
Node	Time	Ref. (*7*)		This study		Node	Time	Ref. (*7*)		This study	
		Time	CI	Time	CI			Time	CI	Time	CI
41	1613	–	–	1613	1761–1465	85	523	–	–	523	638–425
42	1612	–	–	1612	1712–1506	86	509	–	–	509	609–421
43	1579	–	–	1579	1740–1411	87	432	–	–	432	520–351
44	1554	–	–	1554	1729–1378	88	380	–	–	380	469–304

Note: Nodes times in the timetree are from this study.

Single-gene phylogenies were screened for orthology and vertical inheritance under the assumption that phylogenies showing mixing of the two superkingdoms or significantly supported deep nesting of one class within another were affected by HGT and deleted from the data set. Site homology of the multiple sequence alignment was established with GBlocks (*35*) and nonconserved sites were deleted. ML (*36*) and Bayesian (*37*) methods were then applied to the final alignment (6884 sites) to estimate phylogenetic relationships. A ML phylogeny was also constructed from an alternative alignment with only slow-evolving positions (16,344 sites) and showed a similar phylogeny (except for the position of Solibacteres). A topological feature common to all of these analyses is a major dichotomy in eubacteria. This is formed by the Terrabacteria (Actinobacteria, Chloroflexi, Cyanobacteria, *Deinococcus-Thermus*, and Firmicutes) and the Hydrobacteria (Acidobacteria, Bacteroidetes, Chlamydiae, Chlorobi, Planctomycetacia, Proteobacteria, and Spirochaetes) to the exclusion of the hyperthermophiles and, possibly, Fusobacteria (*88*). This definition of Terrabacteria is more inclusive than the initial definition provided in 2004, which did not include Chloroflexi and Firmicutes (*7*).

In addition, three possible high-level clusters within Hydrobacteria are highly supported by the ML phylogeny. Bacteroidetes, Chlorobi, Chlamydiae, Planctomycetes, and Spirochaetes are associated in a single cluster, with high bootstrap support (82%), for which we propose the name Spirochlamydiae. Within this infrakingdom, two other superphyla are strongly supported: (i) Planctomycetes and Spirochaetes (93%) and (ii) Bacteroidetes and Chlorobi (100%). For these we propose the names Spiroplancti and Bacterobi, respectively. These last two clusters are present and significantly supported in the Bayesian phylogeny as well. Multiple evidence from other analyses (*2, 11, 23, 28*) supports Bacterobi, while Spiroplancti is present in some but not all previous studies. However, currently no clear physiological or metabolic adaptations can be considered as unifying characteristics of these clusters.

Molecular clock methods have been used infrequently with eubacteria, probably because of difficulties in establishing a robust phylogeny and in selecting reliable calibration points. Calibrations have included the fossil record of prokaryotes and eukaryotes and biomarkers in the geologic record. A problem with most eukaryote fossils, otherwise useful for calibration, is that they are from the Phanerozoic eon (543–0 Ma), much younger than many key nodes of interest to time. This exacerbates the problems with different rates of evolution among the three superkingdoms (*38*), an issue diminished by using calibrations within each superkingdom. Although the prokaryote fossil record extends reliably back to ~2000 Ma (*39*), the taxonomic resolution is not sufficient in most cases to be broadly useful for calibration. Better resolution for a few groups is obtained from the geologic record with biomarkers produced specifically and exclusively by one lineage (*40*). Prokaryotes that are parasitic on eukaryotes provide other opportunities for calibration (*41, 42*), but they are almost entirely in the Phanerozoic, again much younger than many deep nodes of interest. One additional calibration that has been used is the maximum time for life on Earth (*7*), which can be set either to the origin of Earth at 4600 Ma as an absolute maximum or perhaps more realistically to the last ocean-vaporizing event, ~4200 Ga (*43*).

Data sets using different types of molecular information (i.e., rRNA, enzymes, core genes) have been used to estimate divergence times. A divergence time analysis of

prokaryotes in the mid-1990s used amino acid sequences of 64 proteins, calibrated within eukaryotes (*44*). Divergences among Cyanobacteria, Gram positives (Firmicutes and Actinobacteria), and Gram negatives (e.g., Proteobacteria) were estimated to be in the early Proterozoic, 2500–2100 Ma. This time interval includes what would be later identified as the major rise in atmospheric oxygen (*45*) caused by the evolution of oxygenic photosynthesis in Cyanobacteria. However, the times in that study are problematic because lineage-specific rate differences (*14*) were not taken into account. In 2001, another prokaryote timescale was constructed from the limited genome data available at that time and again calibrated within eukaryotes (*14*). Although the focus in that study was on the origin of eukaryotes and eukaryotic genes from among different groups of prokaryotes, a divergence time for the split of Cyanobacteria and its closest relative among eubacteria was estimated as 2560 (2810–2300) Ma. Thus two studies using different data and methods obtained a similar, "young" date for the origin of Cyanobacteria, contrasting with the ~3500 Ma fossils (*46*) widely believed to be of cyanobacteria at that time. Subsequently, those fossils were reexamined and shown to be either of some other organism (not cyanobacteria) or an artifact of preservation (*47–50*). Divergence times among 98 prokaryote species were also estimated with sequences of the SSU rRNA gene, a single calibration, and a global clock method (*51*). However, the results were problematic because uncorrected distances were used which greatly bias time estimates.

A genome-based molecular clock study appeared in 2004 which used sequences of 32 "core" proteins shared by 69 prokaryote species and three eukaryotes. A Bayesian clock method was used with multiple calibrations. The origin of Firmicutes was estimated at 2688 Ma (credibility interval, CI: 3108–2360 Ma) and the origins of alpha-, beta-, and gammaproteobacteria were estimated at 2508 Ma (CI: 2928–2154 Ma) (Table 1).

Here, a detailed timetree of 197 eubacterial species (81 families) was estimated from the ML phylogeny of our latest data set (*88*) using archaebacteria as an outgroup (Fig. 2). This rooting follows analyses of paralogous genes that found the root of the tree of life between archaebacteria and eubacteria (*52, 53*), a root position that has been generally accepted for the last two decades and supported in a recent, expanded analysis (*54*). Although analyses of insertions and deletions in paralogous genes have suggested that the root might be in some other location (*55, 56*), others have questioned the alignments used in those insertion–deletion analyses (*57, 58*). Additional evidence is needed before any root position can be considered well established.

Divergence times were obtained with a Bayesian method (*59*) using all available calibration points within eubacteria. These included: (i) a minimum of 1640 Ma for the origin of Chromatiaceae (gammaproteobacteria) based on biomarker evidence (i.e., okenane) (*60*); (ii) a minimum of 1640 Ma for the divergence of Chlorobi and Bacteroidetes based on biomarker evidence (i.e., chlorobactane) (*60*); (iii) a minimum at 2300 Ma for the divergence of Cyanobacteria and Chloroflexi corresponding to the first significant rise in oxygen concentration in the atmosphere (*45*); (iv) a maximum of 4000 Ma for the earliest land-dwelling taxa corresponding to the presence of continents (*61*); and (v) a maximum constraint at 4200 Ma for the first divergence within eubacteria from the midpoint of the time range estimated for the last ocean-vaporizing event (*43*). In the final time estimates, calibrations (iii) and (iv) were omitted because their absence did not significantly alter the estimates of the other nodes while it allowed inferences on these evolutionary adaptations. We used one species representative for monophyletic families and multiple species for paraphyletic families (Mycoplasmataceae, Thermoanaerobacteriaceae; Synechococcaceae); species lacking a family classification are also shown (*Candidatus Pelagibacter ubique*, *Symbiobacterium thermophilum*, *Carboxydothermus hydrogenoformans*).

Time estimates (Table 1, Fig. 2) for divergences among higher-level taxa – phyla and classes – were mostly in the Archean Eon (4500–2500 Ma), with only the phyla Chlorobi/Bacteroidetes, and the classes Beta-/Gammaproteobacteria, and Bacilli/Mollicutes diverging in the Proterozoic (2500–543 Ma). Family divergences, instead, were evenly distributed throughout the Proterozoic with only a few of them occurring more recently in the Phanerozoic (Table 1). Apart from the deep branches of the two hyperthermophilic lineages (Aquificae and Thermotogae), most divergences of phyla and classes were closely spaced in time, 3500–2500 Ma and especially 3000–2600, suggesting the colonization and adaptation of this superkingdom to new environments by the end of the Archean. This corresponds to the radiation of major eubacterial clades found in earlier time analyses (e.g., *7, 14*).

The short internal branches in portions of the timetree may explain why there has been difficulty in obtaining a robust phylogeny in past studies (*14, 62*). The position of hyperthermophiles at the base of the eubacterial tree

has been criticized on the grounds of potential artifacts caused by different nucleotide base compositions and outgroup rootings (*33*). However, in our most recent analysis (*88*) this topology was supported by multiple types of evidence. Regardless of the position of the hyperthermophiles, some recent support has been obtained for the adaptation of eubacterial proteins to thermophilic temperatures (i.e., above 50°C) (*63*). This suggests that the ancestor of eubacteria may have lived in a high-temperature environment.

The origin of major adaptations in eubacteria, such as phototrophy, agrees well with evidence from the geologic record. The patchy distribution of phototrophic and photosynthetic genes among eubacteria suggests a combination of vertical and horizontal gene transfer (*64–68*). Accordingly, the ancestor of all eubacteria, except the two hyperthermophiles and Fusobacteria, most likely was a phototroph, placing the origin of this lifestyle by at least the mid-Archean (3265–2987 Ma) (Table 1). We propose the name Selabacteria (from the Greek *selas*, light, and *bacteria*, rod, in allusion to the innovation of phototrophy) for this group. The exclusion of Fusobacteria from Selabacteria should be treated with caution because of its uncertain phylogenetic position. The first evidence for photosynthesis-mediated sediment deposition and anoxygenic photosynthetic ecosystems is present at 3400 Ma (*69–71*). Considering the timetree and geologic evidence, we infer an early evolution of phototrophy, ~3400–3200 Ma. It is likely that the origin of phototrophy allowed eubacteria to undergo further biochemical adaptation and evolutionary radiation, as seen in the relatively rapid branching of many phyla in the timetree 3000–2600 Ma (Fig. 2).

Shortly after the evolution of phototrophy, eubacteria diverged into two main groups of phyla, a gram-negative group, Hydrobacteria, and a primarily gram-positive group, Terrabacteria (Fig. 2). A previous analysis of aerobic methanotrophy showed that its origin within Hydrobacteria probably occurred by the end of the Archean, 2510 Ma (*7*). This metabolism was recently discovered in a species of *Verrucomicrobium* (a member of the Verrucomicrobia/Chlamydiae/Planctomycetes supergroup) (*72, 73*), suggesting its presence in the ancestor of Proteobacteria and the Verrucomicrobia. Although Verrucomicrobia is absent from our data set, evidence points to a relationship between it, Chlamydiae, and Planctomycetes (*74*), associating its ancestor with Proteobacteria, and thus the evolution of aerobic methanotrophy, in the late Archean (3116–2835 Ma) (Table 1). This time estimate postdates, as expected, the evolution of methanogenesis (minimum of 3460 Ma), which would have contributed to produce the substrate for methanotrophy. However, the requirement of oxygen for methanotrophy makes its evolution unlikely before the origin of the oxygen-producing cyanobacteria (2920–2592 Ma) (Table 1) and the rise in atmospheric oxygen recorded in the geologic record (*45*). Given this evidence we consider it more likely that the evolution of aerobic methanotrophy occurred in the ancestor of alpha- and gammaproteobacteria (2630–2371 Ma; Table 1) and then subsequently spread to other lineages by HGT.

Terrabacteria show multiple adaptations to terrestrial habitats such as the formation of resistant stages (e.g., endospores in Firmicutes, akinetes in Cyanobacteria), the production of photoprotective pigments (e.g., carotenoids), and resistance to desiccation (*75*). Although these traits are not uniquely distributed within Terrabacteria, the coexistence of more than one within a class and the presumed terrestrial niche of the ancestor (*88*) suggest that terrestrial environments were colonized by the late Archean (3041–2755 Ma) (Table 1). Evidence from the geologic record supports this result because the continents are thought to have formed by the early Archean, 4000–3800 Ma (*61, 76*), and the first terrestrial ecosystems were present by 2600 Ma (*77*). The early colonization of land opens the possibility that later evolutionary adaptations (e.g., oxygenic photosynthesis) were favored by conditions in this environment. Moreover, Terrabacteria includes approximately two-thirds of all prokaryote species (~9740) further suggesting a major influence of this environment on prokaryote speciation and adaptation (*88*).

The sparse geologic record of prokaryotes and their biomarkers of carbon assimilation and lipid biosynthesis offer few opportunities to compare or test the molecular timescale (Fig. 2), especially since the best of such records were used as calibration points. We did not use the hopane biomarker initially proposed for cyanobacteria (*78, 79*) because it has turned up in other groups recently (*80*), thus removing it as a unique biomarker for that group. Our time estimate for the origin of cyanobacteria (2760 Ma), although young by classical interpretations based on early evaluations of stromatolites and fossils, is nonetheless older than—and thus consistent with—the rise of oxygen in the geologic record at ~2300 Ma (*45*). This time also overlaps with the possible sterane evidence for eukaryote algae and hence aerobic life at 2700 Ma (*79, 81*). However, the sterane biomarker evidence is contradicted by molecular clock evidence

that plastids—necessary for algae—did not arise in eukaryotes until ~1500 Ma (*82, 83*), and by the fossil record of eukaryotes that only extends reliably back to about ~1500 Ma (*84–87*).

In summary, the sequence of biologic events inferred in the timetree is logical and consistent with the geologic record: the colonization of land after the appearance of continents and before the evidence of terrestrial ecosystems in the geologic record, methanogenesis followed by methanotrophy, and oxygenic photosynthesis followed by the major rise in atmospheric oxygen. With new genome sequences continually appearing, methods of time estimation constantly being refined, and new biomarkers discovered, the future appears bright for a well-defined timetree of eubacteria.

Acknowledgments

Support was provided by the U.S. National Science Foundation and National Aeronautics and Space Administration (NASA Astrobiology Institute) to S.B.H.

References

1. G. M. Garrity *et al.*, Taxonomic outline of the Bacteria and Archaea Release 7.7 (http://www.taxonomicoutline.org/) (Michigan State University, East Lansing, Michigan, 2007).
2. T. Cavalier-Smith, *Int. J. Syst. Evol. Microbiol.* **52**, 7 (2002).
3. G. Webster *et al.*, *FEMS Microbiol. Ecol.* **62**, 78 (2007).
4. N. R. Pace, *Science* **276**, 734 (1997).
5. G. E. Fox *et al.*, *Science* **209**, 457 (1980).
6. C. R. Woese, G. E. Fox, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5088 (1977).
7. F. U. Battistuzzi, A. Feijão, S. B. Hedges, *BMC Evol. Biol.* **4**, 44 (2004).
8. C. Brochier, S. Gribaldo, Y. Zivanovic, F. Confalonieri, P. Forterre, *Genome Biol.* **6**, R42 (2005).
9. J. R. Brown, C. J. Douady, M. J. Italia, W. E. Marshall, M. J. Stanhope, *Nat. Genet.* **28**, 281 (2001).
10. F. D. Ciccarelli *et al.*, *Science* **311**, 1283 (2006).
11. E. J. Deeds, H. Hennessey, E. I. Shakhnovich, *Genome Res.* **15**, 393 (2005).
12. V. Daubin, M. Gouy, G. Perriere, *Genome Res.* **12**, 1080 (2002).
13. U. Gophna, W. F. Doolittle, R. L. Charlebois, *J. Bacteriol.* **187**, 1305 (2005).
14. S. B. Hedges *et al.*, *BMC Evol. Biol.* **1**, 4 (2001).
15. S. R. Henz, D. H. Huson, A. F. Auch, K. Nieselt-Struwe, S. C. Schuster, *Bioinformatics* **21**, 2329 (2005).
16. C. H. House, B. Runnegar, S. T. Fitz-Gibbon, *Geobiology* **1**, 15 (2003).
17. E. K. Lienau, R. DeSalle, J. A. Rosenfeld, P. J. Planet, *Syst. Biol.* **55**, 441 (2006).
18. Y. I. Wolf, I. B. Rogozin, N. V. Grishin, E. V. Koonin, *Trends Genet.* **18**, 472 (2002).
19. R. S. Gupta, E. Griffiths, *Theor. Pop. Biol.* **61**, 423 (2002).
20. F. Delsuc, H. Brinkmann, H. Philippe, *Nat. Rev. Genet.* **6**, 361 (2005).
21. W. F. Doolittle, *Science* **284**, 2124 (1999).
22. V. Daubin, M. Gouy, G. Perrière, *Genome Informat.* **12**, 155 (2002).
23. D. Pisani, J. A. Cotton, J. O. McInerney, *Mol. Biol. Evol.* **24**, 1752 (2007).
24. T. Dagan, W. Martin, *Genome Biol.* **7**, 10 (2006).
25. E. Bapteste *et al.*, *BMC Evol. Biol.* **5**, 33 (2005).
26. O. Zhaxybayeva, J. P. Gogarten, *BMC Genomics* **3**, 4 (2002).
27. A. Mira, R. Pushker, B. A. Legault, D. Moreira, F. Rodriguez-Valera, *BMC Evol. Biol.* **4**, 50 (2004).
28. R. S. Gupta, *Crit. Rev. Microbiol.* **30**, 123 (2004).
29. R. S. Gupta, E. Lorenzini, *BMC Evol. Biol.* **7**, 71 (2007).
30. H. Teeling, T. Lombardot, M. Bauer, W. Ludwig, F. O. Glockner, *Int. J. Syst. Evol. Microbiol.* **54**, 791 (2004).
31. G. D. P. Clarke, R. G. Beiko, M. A. Ragan, R. L. Charlebois, *J. Bacteriol.* **184**, 2072 (2002).
32. T. Coenye, P. Vandamme, *Syst. Appl. Microbiol.* **27**, 313 (2004).
33. C. Brochier, H. Philippe, *Nature* **417**, 244 (2002).
34. E. Griffiths, R. S. Gupta, *Int. Microbiol.* **7**, 41 (2004).
35. J. Castresana, *Mol. Biol. Evol.* **17**, 540 (2000).
36. A. Stamatakis, *Bioinformatics* **22**, 2688 (2006).
37. F. Ronquist, J. P. Huelsenbeck, *Bioinformatics* **19**, 1572 (2003).
38. J. M. Kollman, R. F. Doolittle, *J. Mol. Evol.* **51**, 173 (2000).
39. A. H. Knoll, *Life on a Young Planet: The First Three Billion Years of Evolution on Earth* (Princeton University Press, Princeton, NJ, 2003).
40. J. J. Brocks, A. Pearson, *Mol. Geomicrobiol. Rev. Mineral. Geochem.* **59**, 233 (2005).
41. H. Ochman, A. C. Wilson, *J. Mol. Evol.* **26**, 74 (1987).
42. N. A. Moran, M. A. Munson, P. Baumann, H. Ishikawa, *Proc. Roy. Soc. Lond. B* **253**, 167 (1993).
43. N. H. Sleep, K. J. Zahnle, J. F. Kasting, H. J. Morowitz, *Nature* **342**, 139 (1989).
44. D.-F. Feng, G. Cho, R. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13028 (1997).
45. H. D. Holland, *Geochim. Cosmochim. Acta* **21**, 3811 (2002).
46. J. W. Schopf, *Science* **260**, 640 (1993).
47. M. Brasier, N. McLoughlin, O. Green, D. Wacey, *Philos. Trans. Roy. Soc. Lond. B* **361**, 887 (2006).
48. M. D. Brasier *et al.*, *Nature* **416**, 76 (2002).
49. J. W. Schopf, *Philos. Trans. Roy. Soc. Lond. B* **361**, 869 (2006).
50. J. W. Schopf, A. B. Kurdryavtsev, D. G. Agresti, T. J. Wdowiak, A. D. Czaja, *Nature* **416**, 73 (2002).

51. P. P. Sheridan, K. H. Freeman, J. E. Brenchley, *Geomicrobiol. J.* **20**, 1 (2003).
52. J. P. Gogarten *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6661 (1989).
53. N. Iwabe, K. Kuma, M. Hasegawa, S. Osawa, T. Miyata, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9355 (1989).
54. O. Zhaxybayeva, P. Lapierre, J. P. Gogarten, *Protoplasma* **227**, 53 (2005).
55. R. G. Skophammer, J. A. Servin, C. W. Herbold, J. A. Lake, *Mol. Biol. Evol.* **24**, 1761 (2007).
56. J. A. Servin, C. W. Herbold, R. G. Skophammer, J. A. Lake, *Mol. Biol. Evol.* **25**, 1 (2008).
57. J. T. F. Wong, J. Chen, W. K. Mat, S. K. Ng, H. Xue, *Gene* **403**, 39 (2007).
58. M. di Giulio, *Gene* **394**, 105 (2007).
59. J. L. Thorne, H. Kishino, *Syst. Biol.* **51**, 689 (2002).
60. J. J. Brocks, G. D. Love, R. E. Summons, A. H. Knoll, G. A. Logan, S. A. Bowden, *Nature* **437**, 866 (2005).
61. M. T. Rosing, D. K. Bird, N. H. Sleep, W. Glassley, F. Albarede, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **232**, 99 (2006).
62. S. Gribaldo, C. Brochier-Armanet, *Philos. Trans. Roy. Soc. Lond. B* **361**, 1007 (2006).
63. E. A. Gaucher, S. Govindarajan, O. K. Ganesh, *Nature* **451**, 704 (2008).
64. L. J. Mix, D. Haig, C. M. Cavanaugh, *J. Mol. Evol.* **60**, 153 (2005).
65. S. Sadekar, J. Raymond, R. E. Blankenship, *Mol. Biol. Evol.* **23**, 2001 (2006).
66. D. A. Bryant, N.-U. Frigaard, *Trends Microbiol.* **14**, 488 (2007).
67. J. Raymond, O. Zhaxybayeva, J. P. Gogarten, S. Y. Gerdes, R. E. Blankenship, *Science* **298**, 1616 (2002).
68. M. X. Ruiz-Gonzalez, I. Marin, *J. Mol. Evol.* **58**, 348 (2004).
69. J. F. Allen, W. Martin, *Nature* **445**, 610 (2007).
70. D. E. Canfield, M. T. Rosing, C. Bjerrum, *Philos. Trans. Roy. Soc. Lond. B* **361**, 1819 (2006).
71. M. M. Tice, D. R. Lowe, *Geology* **34**, 37 (2006).
72. P. F. Dunfield *et al.*, *Nature* **450**, 879 (2007).
73. A. Pol *et al.*, *Nature* **450**, 874 (2007).
74. M. Wagner, M. Horn, *Curr. Opin. Biotechnol.* **17**, 241 (2006).
75. P. V. Rokitko *et al.*, *Microbiology* **72**, 756 (2001).
76. C. J. Hawkesworth, A. I. S. Kemp, *Nature* **443**, 811 (2006).
77. Y. Watanabe, J. E. Martini, H. Ohmoto, *Nature* **408**, 574 (2000).
78. R. E. Summons, L. L. Jahnke, J. M. Hope, G. A. Logan, *Nature* **400**, 554 (1999).
79. J. J. Brocks, G. A. Logan, R. Buick, R. E. Summons, *Science* **285**, 1033 (1999).
80. S. E. Rashby, A. L. Sessions, R. E. Summons, D. K. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15099 (2007).
81. J. J. Brocks, R. Buick, R. E. Summons, G. A. Logan, *Geochim. Cosmochim. Acta* **67**, 4321 (2003).
82. H. S. Yoon, J. D. Hackett, C. Ciniglia, G. Pinto, D. Bhattacharya, *Mol. Biol. Evol.* **21**, 809 (2004).
83. S. B. Hedges, J. E. Blair, M. L. Venturi, J. L. Shoe, *BMC Evol. Biol.* **4**, 2 (2004).
84. N. J. Butterfield, *Paleobiology* **26**, 386 (2000).
85. J. Samuelsson, N. J. Butterfield, *Precambrian Res.* **107**, 235 (2001).
86. V. N. Sergeev, L. M. Gerasimenko, G. A. Zavarzin, *Microbiology* **71**, 623 (2002).
87. A. H. Knoll, E. J. Javaux, D. Hewitt, P. Cohen, *Philos. Trans. Roy. Soc. Lond. B* **361**, 1023 (2006).
88. F. U. Battistuzzi, S. B. Hedges, *Mol. Biol. Evol.* **26**, 335 (2009).