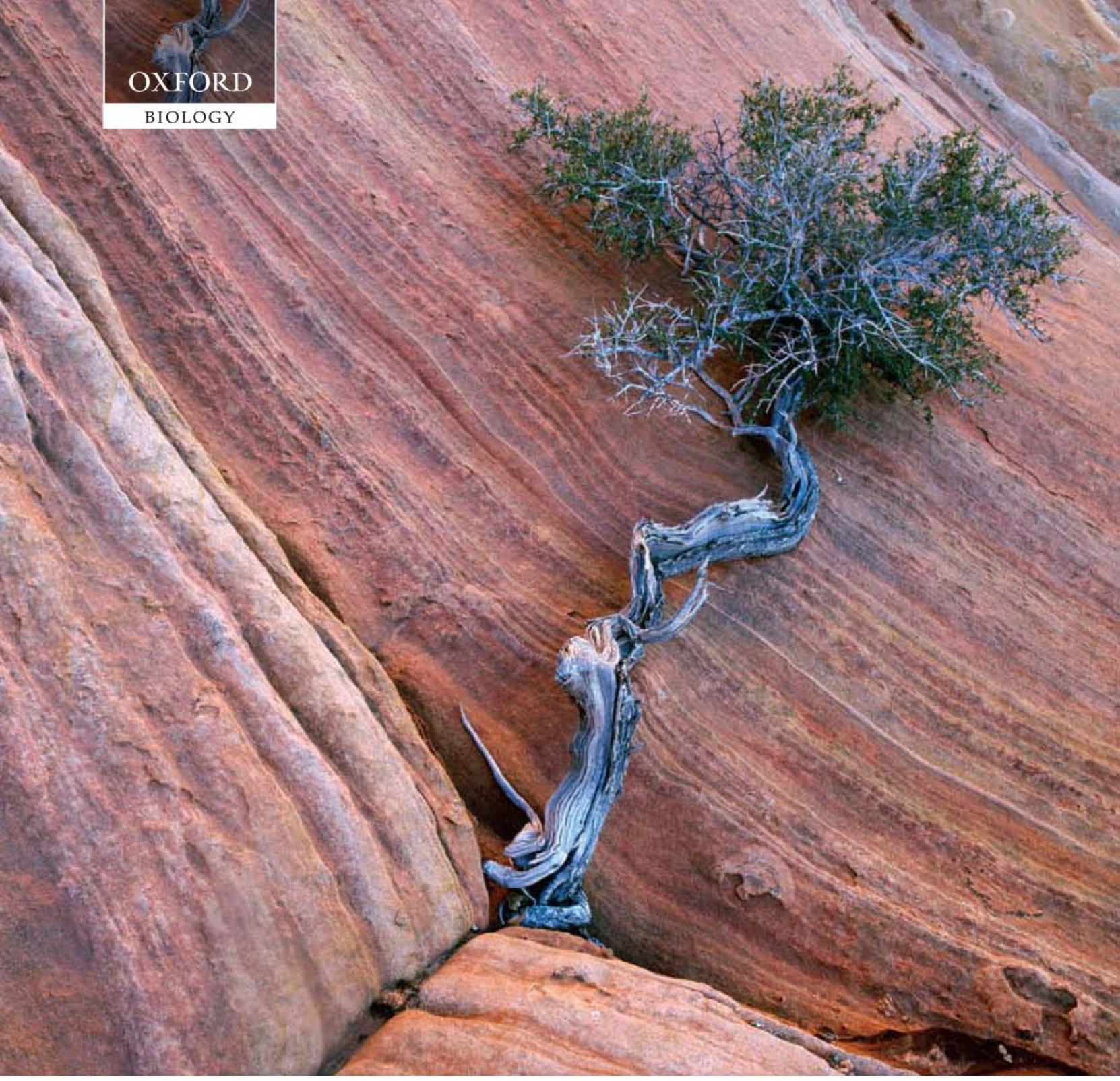

OXFORD
BIOLOGY

the TIMETREE *of* LIFE

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

foreword by James D. Watson

Archaebacteria

Fabia U. Battistuzzi[a,b,*] *and S. Blair Hedges*[a]
[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 Superkingdom Archaebacteria (~300 species) is divided into two phyla, Euryarchaeota and Crenarchaeota, with two other phyla (Korarchaeota and Nanoarchaeota) under consideration. Most large-scale phylogenetic analyses agree on a topology that clusters (i) Methanomicrobia, Halobacteria, Archaeoglobi, and Thermoplasmata and (ii) Methanobacteria, Methanococci, and Methanopyri. A molecular timetree estimated here shows divergences among classes in the Archean Eon, 3500–2500 million years ago (Ma), and family divergences in the Proterozoic Eon, 2394–829 Ma. The timetree also suggests that methanogenesis had arisen by the mid-Archean (>3500 Ma) and that adaptation to thermoacidophilic environments occurred before 1000 Ma.

Extremophiles are common among species in the Superkingdom Archaebacteria (also called "Archaea", Fig. 1) (*1*). For example, the species referred to as "Strain 121" can survive temperatures up to 121°C, higher than any other organism (*2*), and hyperacidophiles are found in the Family Thermoplasmataceae, where two species (*Picrophilus oshimae* and *Picrophilus torridus*) are the only known organisms capable of living at a pH as low as zero (*3, 4*). Archaebacteria show many other phenotypes including the unique ability to produce methane (methanogenesis). They have cell wall structures formed either by pseudopeptidoglycan (i.e., a material similar to the peptidoglycan of eubacteria), polysaccharides, or glycoproteins (S-layer) (*5*), which resemble the single-layer structure (i.e., cell membrane plus cell wall) present in gram-positive eubacteria. Furthermore, archaebacteria have a unique cell membrane structure composed of ether-linked glycerol diethers or tetraethers that confer a higher stability to extreme conditions (*5*). Chemotrophy is the most widely used metabolism, although phototrophic members of the Halobacteriaceae can use light to produce ATP (*6*). Six families also have the unique ability of obtaining energy by combining carbon dioxide (or other carbon compounds) and hydrogen into methane (*5*).

The Superkingdom Archaebacteria, comprising ~300 species, is subdivided into two recognized phyla, Euryarchaeota and Crenarchaeota (*7*). Two other phyla have been proposed based on environmental sequences only (Korarchaeota) and environmental sequences plus one fully sequenced genome (Nanoarchaeota) (*8–11*) but have not been officially recognized and their phylogenetic position is uncertain. The molecular information

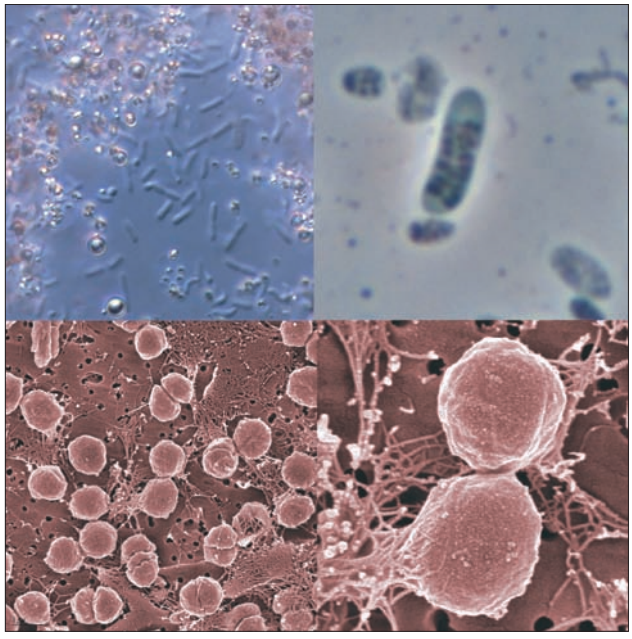

Fig. 1 Halobacteria (rod-shaped *Halobacterium*) from Mono Lake, California; mixed sample (upper left); and close-up of a cell (upper right). Methanococci (round-shaped *Methanococcus*); Cluster of cells (lower left) and close-up of two cells (lower right). Credits: D. J. Patterson, provided by micro*scope (http://microscope.mbl.edu) under creative commons license (upper images); and Electron Microscopy Laboratory, University of California, Berkeley (lower images).

F. U. Battistuzzi and S. B. Hedges. Archaebacteria. Pp. 101–105 in *The Timetree of Life*, S. B. Hedges and S. Kumar, Eds. (Oxford University Press, 2009).

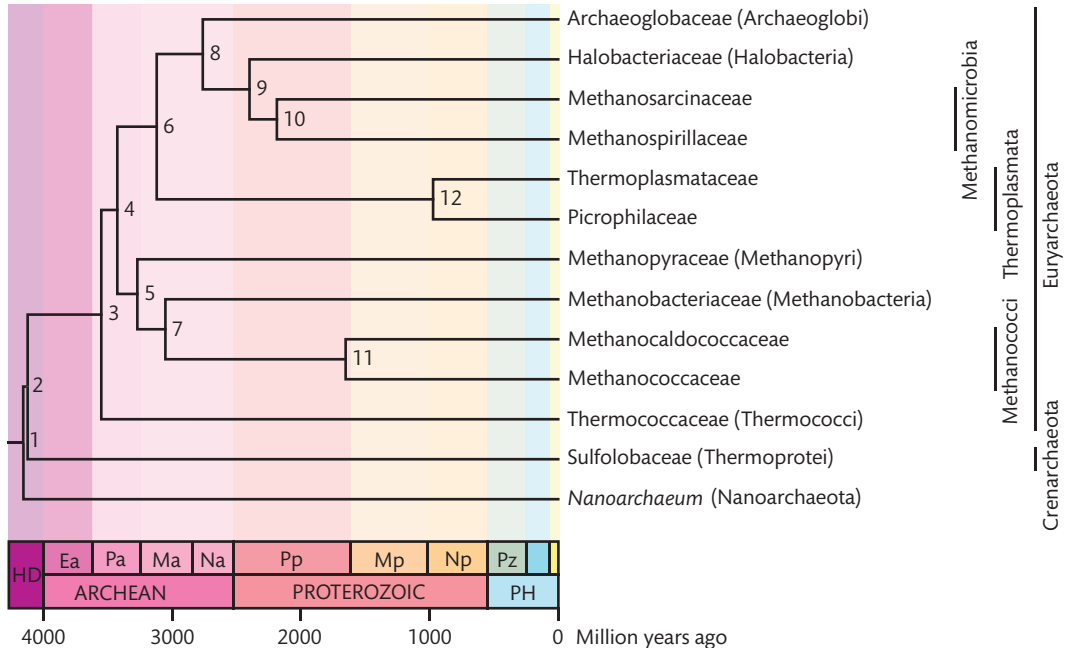

Fig. 2 A timetree of archaebacteria. Divergence times are shown in Table 1. *Abbreviations*: Ea (Eoarchean), HD (Hadean), Ma (Mesoarchean), Mp (Mesoproterozoic), Na (Neoarchean), Np (Neoproterozoic), Pa (Paleoarchean), PH (Phanerozoic), Pp (Paleoproterozoic), and Pz (Paleozoic).

available for Korarchaeota is based only on a few tens of environmental sequences (small subunit ribosomal RNA, SSU rRNA) as none of the Korarchaeota has been successfully cultivated. The sequences available place this group before the Euryarchaeota/Crenarchaeota divergence and show the presence of five major clusters within this putative phylum (*8, 12*). On the other hand, the genome of one species of Nanoarchaeota, *Nanoarchaeum equitans*, has been fully sequenced (*13*) and it is routinely used in multiple-gene phylogenies and indel analyses (*9, 14–18*). These studies show contrasting topologies for Nanoarchaeota, with some placing it basal to Euryarchaeota and Crenarchaeota (*13*), others clustering it with Crenarchaeota (*17, 18*), and other studies clustering it within Euryarchaeota (*9, 14, 16*). These different phylogenetic positions affect the classification of this species, which is either being considered a member of a new phylum (i.e., Nanoarchaeota) or a fast-evolving lineage of Euryarchaeota (*14*).

The phylogeny of taxa within the Phyla Euryarchaeota and Crenarchaeota is mostly stable. Three hypotheses have been proposed to explain this property of archaebacteria that sets them apart from eubacteria: (i) a younger origin of archaebacteria compared to eubacteria (i.e., weaker biases given by a shorter evolutionary history), (ii) a lower known taxonomic diversity (e.g., nine vs. 34 classes for archaebacteria and eubacteria, respectively) (*7*) which could mask contrasting phylogenetic signals, and (iii) speciation events more evenly distributed throughout time that did not result in rapid adaptive radiations (i.e., no or few short internal branches that are difficult to resolve phylogenetically) (*14*). While the geologic record and molecular clock studies do not support the first hypothesis (i.e., archaebacteria and their metabolisms appear early in Earth's history), it is not possible to discard either of the other two hypothesis as a possible cause for the apparently more stable phylogeny of archaebacteria.

Species of Crenarchaeota are placed in six families, three of which (Thermoproteaceae, Desulfurococcaceae, and Sulfolobaceae) have been used in multiple-gene analyses of relationships. These analyses consistently have found Sulfolobaceae and Desulfurococcaceae as closest relatives to the exclusion of Thermoproteaceae, with significant support (*14, 15, 17–19*). Within Euryarchaeota, topological differences are more common depending on the genes and methods used to build the phylogenetic tree. Gene content studies, for example, show the

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

Timetree		Estimates				
Node	Time	Ref. (*19*)		This study		Ref. (*32*)
		Time	CI	Time	CI	Time
1	4193	–	–	4193	4200–4176	–
2	4187	4112	4486–3314	4187	4199–4163	3460
3	3594	–	–	3594	3691–3503	–
4	3468	–	–	3468	3490–3460	–
5	3313	–	–	3313	3388–3232	–
6	3160	3085	3514–2469	3160	3257–3056	–
7	3093	3124	3520–2522	3093	3210–2968	–
8	2799	2625	3102–2037	2799	2936–2656	–
9	2430	–	–	2430	2596–2256	2740
10	2216	–	–	2216	2394–2034	–
11	1676	–	–	1676	1875–1475	–
12	992	–	–	992	1174–829	–

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

Class Halobacteria as a deep-diverging lineage at the base of Euryarchaeota and Crenarchaeota (*20, 21*) followed by the Class Thermoplasmata, either alone or clustering with Crenarchaeota. Because both Halobacteria and Thermoplasmata are classified as euryarchaeotes (*22*), their phylogenetic position determines the mono- or paraphyly of this phylum. Multiple-gene phylogenies show Halobacteria as closely related to the Class Methanomicrobia, a derived position that is highly supported by maximum likelihood (ML), Bayesian, and supertree phylogenies with different sets of genes (*14, 17, 18, 23, 53*). Thermoplasmata, instead, are alternatively supported at the base of Euryarchaeota (*17, 18*) or clustering with Archaeoglobi, Halobacteria, and Methanomicrobia (*14, 23, 53*). Only this last position is significantly supported by all analyses (Fig. 2).

Recent studies with large data sets have solved the issue of the position of the Class Methanopyri. This group was initially found to be an early diverging lineage within Euryarchaeota based on SSU rRNA and indel analyses (*19, 24, 25*). Later studies of the complete genome of its only known representative, *Methanopyrus kandleri* (*26*) and its translation and transcription apparatus (*27*) have shown, instead, that it is more closely related to other methanogens, such as the Methanococci and Methanobacteria. This relation was confirmed by genome-scale analyses with different tree-building methods (supertrees, ML, Bayesian) (*17, 18, 53*).

All of the previous evidence points toward the presence of two main clusters within Euryarchaeota: one formed by the classes Methanomicrobia, Halobacteria, Archaeoglobi, and Thermoplasmata, and another by the classes Methanococci, Methanobacteria, and Methanopyri. These clusters are the same found in a study of sequences from 25 core proteins shared by 218 prokaryote species (including archaebacteria and eubacteria) (*53*). In that analysis, all available species with a complete genome were included and a complete matrix of genes and species was constructed. Single gene trees were manually screened for orthology and vertical inheritance (i.e., genes not showing paraphyly of archaebacteria and eubacteria or significantly supported deep-nesting of one class within another). Site homology of the multiple sequence alignment was established with GBlocks (*28*) and nonconserved sites were deleted. ML (*29*) and Bayesian (*30*) 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 an identical archaebacterial topology. ML and Bayesian phylogenies were found to be identical for archaebacteria with high bootstrap

values (>70%) for the majority of the nodes and are also identical to previous studies, except for the position of Nanoarchaeota (*14*).

As is the case for eubacteria, there have been very few molecular clock studies applied to archaebacteria (*19, 31–33*). For this reason, we estimated divergence times (Table 1) and constructed a molecular timetree (Fig. 2) for 12 families and one phylum with a Bayesian timing method (*34*) and using the only two calibration points available within this superkingdom: (i) a minimum of 3460 million years (Ma) for the origin of methanogenesis based on isotopically light carbon (*23, 35*) and (ii) a maximum of 4200 Ma for the first divergence within archaebacteria based on the midpoint of the range of the last ocean-vaporizing impact (*36*). Besides the initial evidence for biological methane production at 3460 Ma, there is additional evidence for that metabolism later in the Archean, at ~2700 Ma (*37–41*). The topology of the timetree is taken from our latest phylogenetic analysis of sequences from 25 core proteins (*53*), although it is similar to earlier studies of complete genome sequences (e.g., *19*). Divergence times from another study (*32*), using SSU rRNA sequences, a global clock method, a single calibration point, and uncorrected distances, are shown for comparison at relevant nodes in Table 1.

Although we estimate an early divergence between Crenarchaeota and Euryarchaeota (4187, 4199–4163 Ma), all divergences among classes are later in the Archean (3500–2500 Ma), with the divergence of Halobacteria and Methanomicrobia occurring near the Archean–Proterozoic boundary (2430; 2596–2256 Ma). The Phylum Nanoarchaeota is basal to all archaebacteria in our phylogeny and its divergence from other archaebacteria is estimated at 4193 Ma (4200–4176 Ma). However, given its uncertain phylogenetic position this time estimate should be considered with caution. Only three classes (Methanomicrobia, Methanococci, and Thermoplasmata) have representatives of more than one family. These family-level divergences are within the Proterozoic, between 2216 and 992 Ma (CI, 2394–829 Ma).

The deepest branches of both Crenarchaeota and Euryarchaeota are occupied by hyperthermophilic organisms. Although our knowledge of the diversity of archaebacteria is limited and it is possible that mesophilic deep-branching species will be discovered, the current phylogenetic pattern suggests that the ancestor of this superkingdom was adapted to high temperature environments.

The distribution of methanogenesis among families supports multiple losses of this metabolism during evolution (e.g., Thermoplasmata, Halobacteria). The common ancestor of all methanogens (Methanobacteriaceae, Methanocaldococcaceae, Methanococcaceae, Methanosarcinaceae, Methanospirillaceae, and Methanopyraceae) is estimated to have evolved by the mid-Archean regardless of the calibration used for the molecular clock (i.e., eukaryotic or archaebacterial) (*19, 53*). This early evolution of methanogenesis is not only in agreement with the geologic record (*42*) but also lends support to one of the hypotheses addressing the faint young sun paradox (*43*). These suggest that a greenhouse effect was present in the early history of Earth to compensate for the lower luminosity of the Sun. Among the greenhouse gases proposed are carbon dioxide (*44*) and methane (*45–47*) with the latter, according to our timetree, being of biologic origin.

Genes involved in methanogenesis and methylotrophy (i.e., methanopterin and methanofuran-linked C_1 transfer genes) are shared by methanogens and at least two groups of eubacteria (the Phylum Proteobacteria and the Class Planctomycetacia) (*48*). Contrary to a previous hypothesis (*49*), the late divergence of the eubacterial Planctomycetacia (*53*) suggests horizontal gene transfer (HGT) from archaebacteria to eubacteria as a possible cause of their current distribution. An alternative possibility is the presence of this pathway in the ancestor of eubacteria and archaebacteria. This cannot be discarded with the current information, especially in light of the recent discovery of these genes in yet to be classified lineages (*49, 50*).

Another ecological innovation that evolved in archaebacteria is the adaptation to thermoacidophilic environments (i.e., pH < 3; temperature > 50°C). Strict thermoacidophiles are present only in the Class Thermoplasmata (Families Thermoplasmataceae, Ferroplasmaceae, and Picrophilaceae) and some Thermoprotei (e.g., Sulfolobaceae) (*51, 52*) with evidence for extensive HGTs between these two classes. Because the only representative of Thermoprotei is a member of the Family Sulfolobaceae, it is not possible to constrain the time estimate for this metabolism with an upper limit. However, the divergence of Thermoplasmataceae and Picrophilaceae at 992 Ma (1174–829 Ma) sets a minimum time for the origin of this metabolism.

Compared to the Superkingdom Eubacteria, the Superkingdom Archaebacteria is not as well known in terms of its taxonomic and environmental diversity. Nonetheless, the current timetree shows an early origin

of these lineages and evolutionary innovations (e.g., methanogenesis) that are likely to have played a fundamental role in the habitability and colonization of the planet.

Acknowledgment

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. P. Forterre, C. Brochier, H. Philippe, *Theoret. Pop. Biol.* **61**, 409 (2002).
2. K. Kashefi, D. R. Lovley, *Science* **301**, 934 (2003).
3. C. Schleper, G. Puhler, H. P. Klenk, W. Zillig, *Int. J. Syst. Bact.* **46**, 814 (1996).
4. K. J. Edwards, P. L. Bond, T. M. Gihring, J. F. Banfield, *Science* **287**, 1796 (2000).
5. M. T. Madigan, J. M. Martinko, J. Parker, *Brock Biology of Microorganisms* (Prentice-Hall, New Jersey, 10th ed., 2003).
6. D. A. Bryant, Frigaard, N.-U., *Trends Microbiol.* **14**, 488 (2007).
7. G. M. Garrity *et al.* Taxonomic outline of the Bacteria and Archaea Release 7.7 (Michigan State University, East Lansing, Michigan, 2007).
8. T. A. Auchtung, C. D. Takacs-Vesbach, C. M. Cavanaugh, *Appl. Environ. Microb.* **72**, 5077 (2006).
9. C. Brochier, S. Gribaldo, Y. Zivanovic, F. Confalonieri, P. Forterre, *Genome Biol.* **6**, R42 (2005).
10. G. Webster *et al.*, *FEMS Microbiol. Ecol.* **62**, 78 (2007).
11. N. R. Pace, *Science* **276**, 734 (1997).
12. S. M. Barns, C. F. Delwiche, J. D. Palmer, N. R. Pace, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9188 (1996).
13. E. Waters *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12984 (2003).
14. S. Gribaldo, C. Brochier-Armanet, *Philos. Trans. Roy. Soc. Lond. B* **361**, 1007 (2006).
15. B. Gao, R. S. Gupta, *BMC Genomics* **8**, 86 (2007).
16. K. S. Makarova, E. V. Koonin, *Curr. Opin. Microbiol.* **8**, 586 (2005).
17. D. Pisani, J. A. Cotton, J. O. McInerney, *Mol. Biol. Evol.* **24**, 1752 (2007).
18. F. D. Ciccarelli *et al.*, *Science* **311**, 1283 (2006).
19. F. U. Battistuzzi, A. Feijão, S. B. Hedges, *BMC Evol. Biol.* **4**, 44 (2004).
20. B. Runnegar, C. H. House, S. T. Fitz-Gibbon, *Geobiology* **1**, 15 (2003).
21. Y. I. Wolf, I. B. Rogozin, N. V. Grishin, E. V. Koonin, *Trends Genet.* **18**, 472 (2002).
22. G. M. Garrity, Ed., *Bergey's Manual of Systematic Bacteriology* (Springer, New York, 2001).
23. E. Bapteste, C. Brochier, Y. Boucher, *Archaea* **1**, 353 (2005).
24. S. Burggraf, K. O. Stetter, P. Rouviere, C. R. Woese, *Syst. Appl. Microbiol.* **14**, 346 (1991).
25. M. C. Rivera, J. A. Lake, *Int. J. Syst. Bact.* **46**, 348 (1996).
26. A. I. Slesarev *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4644 (2002).
27. C. Brochier, P. Forterre, S. Gribaldo, *Genome Biol.* **5**, R17 (2004).
28. J. Castresana, *Mol. Biol. Evol.* **17**, 540 (2000).
29. A. Stamatakis, *Bioinformatics* **22**, 2688 (2006).
30. F. Ronquist, J. P. Huelsenbeck, *Bioinformatics* **19**, 1572 (2003).
31. D.-F. Feng, G. Cho, R. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13028 (1997).
32. P. P. Sheridan, K. H. Freeman, J. E. Brenchley, *Geomicrobiol. J.* **20**, 1 (2003).
33. S. B. Hedges *et al.*, *BMC Evol. Biol.* **1**, 4 (2001).
34. J. L. Thorne, H. Kishino, *Syst. Biol.* **51**, 689 (2002).
35. Y. Ueno, K. Yamada, N. Yoshida, S. Maruyama, Y. Isozaki, *Nature* **440**, 516 (2006).
36. N. H. Sleep, K. J. Zahnle, J. F. Kasting, H. J. Morowitz, *Nature* **342**, 139 (1989).
37. J. M. Hayes, in *Earth's Earliest Biosphere: Its Origin and Evolution,* J. W. Schopf, Ed. (Princeton University Press, Princeton, New Jersey, 1983) pp. 291–301.
38. J. M. Hayes, in *Early Life on Earth*, S. Bengston, Ed. (Columbia University Press, New York, 1994), pp. 220–236.
39. K. U. Hinrichs, J. M. Hayes, S. P. Sylva, P. G. Brewer, E. F. DeLong, *Nature* **398**, 802 (1999).
40. K.-U. Hinirichs, *11th Annual V.M. Goldschmidt Conference Program and Abstracts* **abs. 3461** (2001).
41. K. U. Hinrichs, *Geochem. Geophys. Geosyst.* **3,** 1042 (2002).
42. Y. Ueno, K. Yamada, N. Yoshida, S. Maruyama, Y. Isozaki, *Nature* **440**, 516 (2006).
43. J. F. Kasting, D. Catling, *Ann. Rev. Astron. Astrophys.* **41**, 429 (2003).
44. H. Ohmoto, Y. Watanabe, K. Kumazawa, *Nature* **429**, 395 (2004).
45. A. A. Pavlov, J. F. Kasting, L. L. Brown, K. A. Rages, R. Freedman, *J. Geophys. Res.* **105**, 11981 (2000).
46. A. A. Pavlov, M. T. Hurtgen, J. F. Kasting, M. A. Arthur, *Geology* **31**, 87 (2003).
47. J. F. Kasting, A. A. Pavlov, J. L. Siefert, *Orig. Life Evol. Biosph.* **31**, 271 (2001).
48. L. Chistoserdova, S. W. Chen, A. Lapidus, M. E. Lidstrom, *J. Bacteriol.* **185**, 2980 (2003).
49. L. Chistoserdova *et al.*, *Mol. Biol. Evol.* **21**, 1234 (2004).
50. M. G. Kalyuzhnaya, L. Chistoserdova, *Methods Enzymol.* **397**, 443 (2005).
51. C. Bertoldo, C. Dock, G. Antranikian, *Eng. Life Sci.* **4**, 521 (2004).
52. A. Angelov, W. Liebl, *J. Biotechnol.* **126**, 3 (2006).
53. F. U. Battistuzzi, S. B. Hedges, *Mol. Biol. Evol.* **26,** 335 (2009).