TUNNELING THROUGH TIME: HORIZONTAL GENE TRANSFER CONSTRAINS THE TIMING OF METHANOGEN EVOLUTION

Joanna M. Wolfe[a,1] and Gregory P. Fournier[a]
[a] Department of Earth, Atmospheric & Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[1] To whom correspondence should be addressed. Email: jowolfe@mit.edu.

Keywords: Archaea, Cyanobacteria, Divergence times, Horizontal gene transfer, Methanogenesis

Abstract ABSTRACT

Archaeal methane production is a major component of the modern carbon cycle. It has been proposed that the metabolism of methanogenic Archaea also contributed to a “methane greenhouse” during the early Archaean Eon, a hypothesis requiring evidence for the evolution of methanogenesis at or before this time. Molecular clock models are frequently used to estimate divergence times of organismal lineages in phylogenies, as well as the order of character acquisition. However, estimating the timing of microbial evolutionary events, especially ones as ancient as the Archaean, is challenged by the lack of diagnostic fossils. Other methods are thus required to calibrate the ages of these clades. Horizontal gene transfers (HGTs) are ubiquitous evolutionary events throughout the Tree of Life, complicating phylogenomic inference by introducing topological conflicts between different gene families across taxa. As HGTs also convey relative timing information between lineages, they can be harnessed to provide geological age constraints for clades lacking a fossil record. We derive a valuable temporal constraint on the timing of the evolution of methanogenesis from a single HGT event from within archaeal methanogen lineages to the ancestor of Cyanobacteria, one of the few microbial clades with recognized crown group fossils. Results of molecular clock analyses using this HGT predict methanogens most likely diverging within Euryarchaeota no later than 3.51 Ga, and methanogenesis itself likely evolving substantially earlier. This timing provides independent support for scenarios wherein microbial methane production has a substantial role in maintaining temperatures on the early Earth.

SIGNIFICANCE

Methanogenic Archaea are the only organisms known to provide biogeochemically relevant sources of methane on Earth today. While this metabolism is undoubtedly ancient, the oldest geochemical evidence is too young to constrain the emergence of microbial methane, and there is a paucity of reliable microbial fossils to suggest the presence of methanogenic lineages. Molecular clock analyses of methanogenic Archaea, with age constraints derived from an HGT from within methanogens to the ancestor of Cyanobacteria, provide independent support for the hypothesis of an Eoarchaean biogenic methane greenhouse. This approach has broad implications for estimating the ages of microbial clades across the entire Tree of Life, a critical yet largely unexplored frontier of natural history.

Introduction

Methane is a greenhouse gas implicated in current and past climate change. Accumulation of atmospheric methane during the Archaean Eon has been proposed as one solution to the “faint young sun paradox”, contributing to increased global temperatures enough to maintain a liquid hydrosphere despite the lower luminosity of the sun at the time (1–6). While microbial methanogenesis is generally assumed to be an extremely ancient pathway due to its phylogenetic distribution across much of Euryarchaeota (7), there is only limited geochemical evidence for microbial methane production in the Archaean, in the form of carbon isotopic composition of kerogens ~2.7 Ga (8) and methane-bearing fluid inclusions ~3.46 Ga (9). Therefore, the time of the onset of microbial methane production and the relative contributions of microbial and abiogenic sources to Archaean atmospheric methane remain uncertain. The case for a microbial methane contribution would be strengthened by molecular clock estimates showing the divergence of methane-producing Archaea predates their proposed geochemical signature. Few such studies have been conducted, resulting in a range of dates for the origin of microbial methanogenesis spanning the majority of the Precambrian, such as 3.05-4.49 Ga (4), 3.46-3.49 Ga (10), ~3.45 Ga (11), 2.97-3.33 Ga (12), and a much younger 1.26-1.31 Ga (13). These studies were based exclusively on information from molecular sequence data, without fossil calibrations, as these are absent for all Archaea. In this work, we employ an HGT event from methanogens to a microbial clade with a fossil record (Cyanobacteria) to enable the use of fossil-calibrated molecular clocks and methods comparable to those validated in metazoans. As calibrations from the rock record are essential for accurate molecular clock inferences, particularly for ancient splits (13), their inclusion permits more accurate and precise dating of the earliest methanogens.

Estimating divergence times requires calibration points from the geological record (14): body or trace fossils attributable to a clade's crown group, preferably by phylogenetic analysis (15, 16), or (more controversially) preserved traces of organic biomarkers that may be diagnostic for certain clades (17, 18). There is no such direct evidence, however, for Archaea in deep time (let alone methanogens nested deeply within Euryarchaeota (19)); thus no dating analysis has previously been conducted using independent calibrations in this clade. Without geological constraints, confidence in divergence estimates rests entirely on the unconstrained rate models used, which are sensitive to lineage-specific rate changes (20), and cannot be internally cross-validated.

HGT events represent temporal intrusions between genomes, where the recipient clade acquires genetic information not inherited from its direct ancestor. These relationships determine the relative age of the donor (older) and recipient (younger) clades (21). Previous work has argued for the relative ages of clades (22–25), or has used HGT events en masse as secondary calibrations for molecular clock studies (26, 27). Caution must be applied when importing secondary divergence estimates from prior molecular clock studies, as they may propagate errors associated with the original estimate, leading to false precision (28, 29). Furthermore, basing a molecular clock solely on donor-recipient logic fails to incorporate the observed reticulating branch length. This is relevant as it is impossible to ascertain whether the HGT occurred near the divergence of the recipient's total group, near the diversification of its crown group, or at any time along its stem lineage. As stem lineages can represent very long time intervals for major microbial clades, their omission may dramatically impact date inferences. Here we extend the observation that HGT events can convey relative temporal evolutionary information, and employ previously identified genes (30, 31) transferred from the clade of interest (in this case, methanogenic Archaea) into clades with a diagnostic fossil record (and thus a direct age estimate; in this case, for Cyanobacteria, the oldest fossils with likely crown-group affinities known in the entire Tree of Life), improving the accuracy and precision of molecular clock models.

Results and Discussion

Horizontal Transfer of SMC Complex. We focused on the HGT of SMC proteins, which, together with ScpA and ScpB proteins, form a complex required for chromosome condensation. The SMC gene was transferred to the root of Cyanobacteria only once; this is demonstrated by a gene tree including all cyanobacterial SMC sequences in GenBank (n=307), excluding sequences horizontally transferred from other bacterial clades, which may be secondary or parallel events (**SI Appendix, Fig. S1**). The most parsimonious explanation, that all three SMC complex genes were transferred to Cyanobacteria from the same donor, was supported by the correspondence of each HGT gene tree topology (**SI Appendix, Figs. S2-S4**).

Previous analyses show the donor lineage as a close relative of Methanosarcinales, Halobacteriales, and Archaeoglobales (31). Our increased taxon sampling within Euryarchaeota weakly supported the same possible donor clades, with the addition of Methanomicrobiales and Methanocellales (which had not been included previously, but are closely related to Methanosarcinales). Halobacteriales have an established strong compositional bias, which may result in long branch attraction (32–34). Omitting Halobacteriales from the concatenated SMC complex alignment (**SI Appendix, Table S1**) significantly improved phylogenetic support that the transfer of SMC proteins to the ancestor of Cyanobacteria was most likely from a sister lineage of Methanomicrobiales (**Fig. 1**).

Validation of Alignment and Dating Approach. To link the HGT event with the species topology of both methanogens and Cyanobacteria, we concatenated 1) aligned SMC complex sequences for Cyanobacteria and Euryarchaeota with 2) ribosomal sequences for Euryarchaeota (expected to reconstruct the Euryarchaeota species tree) and 3) ribosomal sequences for Cyanobacteria as three separate partitions in a 'meta-alignment'. This meta-alignment allows the reticulating branch length for SMC complex evolution to be included in dating analyses, while the topology and branch lengths of the respective euryarchaeal and cyanobacterial clades are inferred from the far more extensive site information within ribosomal datasets. The meta-alignment maximally captures the sequence information required to infer divergences of the donor and recipient lineage, as well as sequence information supporting the length and placement of the reticulating branch.

Previous studies have shown Bayesian dating approaches may be robust to extensive missing sequence data (37). We further explored the suitability of meta-alignments with large blocks of missing data (where entire clades lack all 30 ribosomal sequences) using simulations in evolver (**SI Appendix, Methods** and **Fig. S7**). The mean age estimates for crown Cyanobacteria were slightly yet significantly older when missing data were included, increasing the age of crown Cyanobacteria by 2.6% (two-sample Z-test, $p=0.0003$, $alpha<0.05$); however, the mean age estimates for the donor clade were not affected ($p=0.248$). Therefore, missing data have a small impact on age estimates, but this level of significance does not propagate to deeper nodes.

The accuracy of divergence times estimated individually from the Euryarchaeota species tree (uncalibrated relaxed clock only) differed substantially from the SMC gene tree (incorporating the HGT into the analysis) and the meta-alignment result (**Fig. 2A**). Based on the species tree, methanogens would have diverged within the Paleoarchaean (mean 3.53 Ga ± SD of 163 Myr, minimum 3.24 Ga). Analyses of the SMC complex alone (mean 3.96 Ga ± 236 Myr, minimum 3.46

Ga) and the meta-alignment (mean 3.94 Ga ± 228 Myr, minimum 3.51 Ga) both yield older age estimates for methanogens, in the Eoarchaean. Precision of the latter two analyses is similar for deeper nodes and slightly lower in the meta-alignment for Cyanobacteria. In the Euryarchaeota species tree alone, the donor node (Methanosarcinales + Methanomicrobiales) was not significantly older (2.46 Ga ± 158 Myr) than estimates for crown Cyanobacteria from other analyses (mean 2.32 Ga ± 180 Myr; below), which is unlikely as the donor node *must* be older than the recipient (21). This necessary adjustment makes up for the slight decrease in accuracy when the HGT partition is added. The advantage of adding ribosomal alignment blocks to the HGT is thus the incorporation of taxa and outgroups that lack the SMC complex genes, allowing us to infer the ages of more ancient nodes, including the methanogen ancestor and its deepest splits (e.g. Methanopyrales, Methanobacteriales).

Effect of Fossil Calibrations. Divergence time estimates calibrated by a 2.0 Ga fossil akinete (rod-like resting cell (38) are extremely old (**Fig. 2B**), with the age of Cyanobacteria (mean 2.93 Ga ± 161 Myr, minimum 2.62 Ga) substantially predating the Great Oxygenation Event (GOE; 2.33 Ga (39)), and with the age of microbial methanogenesis tipping into the Hadean or Eoarchaean (mean 4.33 Ga ± 240 Myr, minimum 3.88 Ga). In this analysis, the age of Euryarchaeota (mean 4.53 Ga ± 252 Myr, minimum 4.09 Ga) violates the maximum prior applied to the root, estimating a most likely ancestor age older than the oldest zircons (4.38 Ga (40)), and possibly older than the Earth itself. The maximum plausible fossil age for total-group Nostocales (before resulting mean estimates violate the root prior; **Fig. 2B**) corresponds to ~1.7 Ga, a similar age to proposed akinete material from the McArthur Group of Northern Australia (41). As the validity of the 2.0 Ga microfossil has been questioned (42) we also calibrated the same node on our tree with a younger 1.2 Ga fossil akinete (43), which has greater morphological evidence (42), and is the most conservative estimate discussed more extensively below (**Fig. 3**). This calibration results in age estimates for Cyanobacteria (mean 2.32 Ga ± 180 Myr, minimum 1.97 Ga) very close to and potentially younger than GOE, microbial methanogenesis in the Eoarchaean (mean 3.94 Ga ± 228 Myr, minimum 3.51 Ga), and a correspondingly early age for Euryarchaeota (mean 4.17 Ga ± 228 Myr, minimum 3.67 Ga).

Although only a single fossil calibration was used for this analysis, it may still improve accuracy and precision where among-lineage rate variation is accounted for jointly with the root prior (44). In simulations, age estimates for Cyanobacteria are substantially more accurate when a fossil calibration is added (**Fig. 2B**), while deeper nodes in Euryarchaeota are less influenced. The 95% confidence intervals calculated from empirical fossils (1.2 and 2.0 Ga) overlap for the ages of Euryarchaeota and microbial methanogenesis, but not for Cyanobacteria, illustrating the importance of sensitivity analysis for clades such as Cyanobacteria with ghost ranges dependent upon phylogenetic interpretation of fossil discoveries (45)) and the use of (relatively) “safe but late” constraints (28)).

Age of Microbial Methanogenesis. Within Euryarchaeota, hydrogenotrophic methanogenesis appears to be the ancestral pathway, with Methanopyrales, Methanococcales, and Methanobacteriales diverging earlier than Methanomicrobiales, Methanosarcinales, and their relatives (**Fig. 3**). We conservatively estimate the emergence of methanogens as 3.94 Ga ± 228 Myr at the youngest, and the split between Methanosarcinales and Methanomicrobiales (the closest split to the HGT) at 3.10 Ga ± 195 Myr. Therefore, any proposed scenario of late microbial methanogenesis in the Mesoarchaean through Proterozoic (12, 13, 46)) violates the youngest possible calibrated molecular clock estimate (95% CI younger bound of 3.51 Ga), in addition to geochemical evidence (9). Recently, archaeal clades outside of Euryarchaeota, the uncultured Bathyarchaeota and Verstraetearchaeota, were found to possess genes involved in methane metabolism (47, 48)), thus the absolute origin of microbial methanogenesis could

be substantially older than Euryarchaeota. Although our analysis is agnostic regarding the ancestral metabolism of Archaea, an older evolutionary history for microbial methanogenesis does not refute the hypothesis of an Archaean microbial methane greenhouse.

A substantial microbial methane greenhouse likely only contributes to Archaean warming of the Earth if 1) methanogenesis evolved early enough, which is consistent with our age estimates, and 2) the divergence time of methanogens predates that of the diversification of poorly characterized microbial taxa involved in anaerobic oxidation of methane (AOM; usually comprising communities of Bacteria and Archaea living together). AOM taxa are of interest because their metabolism can also alter the carbon isotope signature of methane produced by microbes (9, 49). Furthermore, AOM removes a substantial fraction of methane from sediments. Since AOM affects the carbon isotope signature of methane in the opposite direction from microbial methanogenesis, this could effectively erase any geochemical signature of Eorchaean microbial methanogenesis (50, 51). Divergence time estimates for AOM taxa alone could not directly support the existence of microbial methanogenesis, as they can also metabolize abiotic sources of methane (52). Thus comprehensive estimates of divergence times for both methanogenic Euryarchaeota and AOM taxa could together constrain the Archaean "methane greenhouse window", by permitting a narrower, independent interpretation of isotopic data.

Conclusion

As in biostratigraphy, in which index fossils are used to correlate and calibrate rock formations worldwide, a well-supported HGT event from a clade of interest into a fossil-bearing clade permits a direct link between divergence estimates and geological history. Combining data from genes with both reticulate and vertical histories into a single alignment complements other recent developments in calibrating microbial evolution (25, 27, 53–55). Our results strongly support the appearance of major methanogen lineages predating the emergence of crown group Cyanobacteria. Our divergence estimates for Euryarchaeota are consistent with previous hypotheses proposing a role for microbial methane in warming the Archaean Earth. With the growing importance of time-calibrated phylogenies in evolutionary inference (56–62), these methodological developments help to overcome the limitations of the sparse microbial geologic record, and indicate their potential utility in resolving the comparative natural history of microbial clades across the entire Tree of Life.

Methods

Data Matrix Construction. The SMC, ScpA and ScpB proteins form a complex required for chromosome condensation in many microbial groups. Genes encoding these proteins within Cyanobacteria have previously been identified as having been transferred from within Archaea (30, 31). We queried NCBI's nr database using BLASTp for homologs of SMC, ScpA, and ScpB proteins in each member of Euryarchaeota with a sequenced genome (except the species-rich Halobacteriales, for which we selected eight representatives), and representative Cyanobacteria from all orders. Previously reported SMC homologs within Aquificales likely representing an additional HGT from Thermococcales (31) were also included. No ScpB sequences were found in Aquificales or Halobacteriales. Protein sequences for each homolog were individually aligned in MUSCLE v3.7 (63). The SMC protein contained two large poorly aligned regions, representing coiled-coil domains (30, 64). These regions were removed via alignment masking using GUIDANCE (65), leaving 729 aligned sites. For the two Scp proteins (which are much shorter, with limited phylogenetic informativeness; **SI Appendix, Table S2**), we elected not to mask poorly aligned regions in light of recent work indicating

trees resulting from this process may be of decreased quality (66).

Phylogenetic Analysis. Individual gene trees were constructed with RaxML v1.8.9 (67) using the LG4M + G substitution model (68). All three genes were concatenated with FASconCAT v1.0 (69), analyzed in RaxML with 100 bootstrap replicates, and in PhyloBayes v3.3f using two chains and the CAT20 site-dependent model (70, 71). The CAT20 model was used because preliminary analyses using the full CAT model did not reach convergence. An automatic stopping rule was implemented, with tests of convergence every 100 cycles, until the default criteria of effective sizes and parameter discrepancies between chains were met (50 and 0.3, respectively). Trees and posterior probability support values were then generated from completed chains after the initial 20% of sampled generations were discarded as burn-in.

Meta-Alignment. A meta-alignment was constructed to date the origin of methanogens, by concatenating 1) aligned SMC complex sequences for Cyanobacteria and Euryarchaeota with 2) ribosomal sequences for Euryarchaeota (adding representatives of clades without identified SMC homologs, i.e. Methanobacteriales, Methanocellales, Methanopyrales, and Thermoplasmatales) and 3) ribosomal sequences for Cyanobacteria, as three separate partitions. Specifically, 30 ribosomal proteins (**SI Appendix, Table S3**) were identified by BLASTp, aligned separately in MUSCLE, then concatenated. Thus only SMC complex sequences determine cyanobacterial placement 'within' Euryarchaeota along the reticulating HGT branch. Note that SMC sequences from Aquificales were omitted from these analyses, as this additional putative HGT event is uninformative in this investigation. The concatenated topology was estimated with RaxML using the LG4M + G model and PhyloBayes using CAT20.

Fossil Calibration. To produce a divergence estimate, we applied a time constraint within Cyanobacteria, derived from fossil resting cells (akinetes; genus *Archaeoellipsoides*) similar to the cyanobacterial clades Nostocales (morphological subsection IV) and Stigonematales (subsection V), from the 2.0 Ga Franceville Group of Gabon (38, 41, 72, 73). There are too few morphological characters to determine a crown-group position of this fossil (42), so we assigned the fossil minimum age to total-group Nostocales (i.e. the clade in our tree including Nostocales and Stigonematales as shown in (42), and their sister group Chroococcidiopsidales). As the affinities of Paleoproterozoic *Archaeoellipsoides* have been questioned (42, 74), we also tested a less controversial younger fossil of the same genus with a more similar size to members of total-group Nostocales, from the 1.2 Ga Dismal Lakes Group of Northwest Canada (43, 74). To measure the effect of using different *Archaeoellipsoides* fossil ages on divergence time estimates, we simulated calibrations for total-group Nostocales at 100 Myr intervals between 1.3 and 2.3 Ga (in addition to the empirical fossil dates at 1.2 and 2.0 Ga).

Divergence Time Estimation. Divergence times were estimated in PhyloBayes using a fixed topology and branch lengths from the RaxML meta-alignment result and the uncorrelated gamma multipliers (UGM) relaxed clock model (20). The UGM model allows substitution rates to vary across the tree, and makes no assumptions about autocorrelation of evolutionary rates across deep branches (20), also accounting for genes evolving at a fixed ratio among lineages (35). Therefore, this model is suited to modeling rate changes associated with HGT events along a reticulating branch.

The root was calibrated with a gamma distributed prior with a mean of 3.9 Ga and SD 230 Myr; this constraint was calculated as the mean of the maximum root age of 4.38 Ga (oldest zircons,

approximating the age of habitable Earth: (40)) and minimum of 3.46 Ga (oldest traces of microbial methane: (9)). Each fossil age (above) was used as a hard-bound minimum constraint on a uniform age prior, which is appropriate due to the extreme antiquity and limited character information from these calibrations. Other validatory analyses, including varying the molecular clock model and prior distributions (including comparisons of estimated CIs to the joint prior (75) by removing sequence data using the -prior flag in PhyloBayes) resulted in minimal changes to divergence time estimates.

Data Deposition. The results from this paper are available at Dryad (provisional doi:10.5061/dryad.m371v).

Acknowledgments. We thank D. Gruen, C. Magnabosco, D. Rothman, and B. Schirrmeister for discussions, and G. Shomo for assistance with the Engaging Cluster at MGHPCC. We acknowledge support from Simons Foundation Collaboration on the Origin of Life #339603 and NSF EAR-1615426.

REFERENCES

1. Gough DO (1981) Solar interior structure and luminosity variations. *Physics of Solar Variations* (Springer), pp 21–34.
2. Catling DC, Zahnle KJ, McKay CP (2001) Biogenic Methane, Hydrogen Escape, and the Irreversible Oxidation of Early Earth. *Science* 293:839–843.
3. Pavlov AA, Hurtgen MT, Kasting JF, Arthur MA (2003) Methane-rich Proterozoic atmosphere? *Geology* 31(1):87–90.
4. Battistuzzi FU, Feijao 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(1):44.
5. Haqq-Misra JD, Domagal-Goldman SD, Kasting PJ, Kasting JF (2008) A Revised, Hazy Methane Greenhouse for the Archean Earth. *Astrobiology* 8(6):1127–1137.
6. Wolf ET, Toon OB (2014) Controls on the Archean Climate System Investigated with a Global Climate Model. *Astrobiology* 14(3):241–253.
7. Gao B, Gupta RS (2007) Phylogenomic analysis of proteins that are distinctive of Archaea and its main subgroups and the origin of methanogenesis. *BMC Genomics* 8(1):86.
8. Hinrichs K-U (2002) Microbial fixation of methane carbon at 2.7 Ga: Was an anaerobic mechanism possible? *Geochem Geophys Geosystems* 3(7):1–10.
9. 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(7083):516–519.
10. Marin J, Battistuzzi FU, Brown AC, Hedges SB (2017) The Timetree of Prokaryotes: New Insights into Their Evolution and Speciation. *Mol Biol Evol*:msw245.
11. Battistuzzi FU, Hedges SB (2009) A major clade of prokaryotes with ancient adaptations to life on land. *Mol Biol Evol* 26(2):335–343.
12. Sheridan PP, Freeman KH, Brenchley JE (2003) Estimated Minimal Divergence Times of the Major Bacterial and Archaeal Phyla. *Geomicrobiol J* 20(1):1–14.
13. Blank CE (2009) Not so old Archaea - the antiquity of biogeochemical processes in the archaeal domain of life. *Geobiology* 7(5):495–514.
14. Pisani D, Liu AG (2015) Animal Evolution: Only Rocks Can Set the Clock. *Curr Biol* 25(22):R1079–R1081.
15. Parham JF, et al. (2012) Best Practices for Justifying Fossil Calibrations. *Syst Biol* 61(2):346–359.
16. Wolfe JM, Daley AC, Legg DA, Edgecombe GD (2016) Fossil calibrations for the arthropod Tree of Life. *Earth-Sci Rev* 160:43–110.
17. Brocks JJ, Pearson A (2005) Building the biomarker tree of life. *Rev Mineral Geochem* 59(1):233–258.
18. Rasmussen B, Fletcher IR, Brocks JJ, Kilburn MR (2008) Reassessing the first appearance of eukaryotes and cyanobacteria. *Nature* 455(7216):1101–1104.
19. Borrel G, et al. (2013) Phylogenomic Data Support a Seventh Order of Methylotrophic Methanogens and Provide Insights into the Evolution of Methanogenesis. *Genome Biol Evol* 5(10):1769–1780.
20. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed Phylogenetics and Dating with Confidence. *PLoS Biol* 4(5):e88.
21. Gogarten JP, Murphey RD, Olendzenski L (1999) Horizontal gene transfer: pitfalls and promises. *Biol Bull* 196(3):359–362.
22. Huang J, Xu Y, Gogarten JP (2005) The Presence of a Haloarchaeal Type Tyrosyl-tRNA Synthetase Marks the Opisthokonts as Monophyletic. *Mol Biol Evol* 22(11):2142–2146.

23. Huang J, Gogarten P (2009) Ancient Gene Transfer as a Tool in Phylogenetic Reconstruction. *Horizontal Gene Transfer*, Methods in Molecular Biology., eds Gogarten MB, Gogarten JP, Olendzenski LC (Humana Press, Totowa, NJ), pp 127–139.
24. Petitjean C, Moreira D, López-García P, Brochier-Armanet C (2012) Horizontal gene transfer of a chloroplast DnaJ-Fer protein to Thaumarchaeota and the evolutionary history of the DnaK chaperone system in Archaea. *BMC Evol Biol* 12(1):226.
25. Szöllősi GJ, Boussau B, Abby SS, Tannier E, Daubin V (2012) Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations. *Proc Natl Acad Sci* 109(43):17513–17518.
26. Szöllősi GJ, Tannier E, Lartillot N, Daubin V (2013) Lateral Gene Transfer from the Dead. *Syst Biol* 62(3):386–397.
27. Rothman DH, et al. (2014) Methanogenic burst in the end-Permian carbon cycle. *Proc Natl Acad Sci* 111(15):5462–5467.
28. Sauquet H, et al. (2012) Testing the Impact of Calibration on Molecular Divergence Times Using a Fossil-Rich Group: The Case of *Nothofagus* (Fagales). *Syst Biol* 61(2):289–313.
29. Schenk JJ (2016) Consequences of Secondary Calibrations on Divergence Time Estimates. *PLoS ONE* 11(1):e0148228.
30. Soppa J (2001) Prokaryotic structural maintenance of chromosomes (SMC) proteins: distribution, phylogeny, and comparison with MukBs and additional prokaryotic and eukaryotic coiled-coil proteins. *Gene* 278(1):253–264.
31. Cobbe N, Heck MMS (2003) The Evolution of SMC Proteins: Phylogenetic Analysis and Structural Implications. *Mol Biol Evol* 21(2):332–347.
32. Fukuchi S, Yoshimune K, Wakayama M, Moriguchi M, Nishikawa K (2003) Unique Amino Acid Composition of Proteins in Halophilic Bacteria. *J Mol Biol* 327(2):347–357.
33. Paul S, Bag SK, Das S, Harvill ET, Dutta C (2008) Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. *Genome Biol* 9(4):R70.
34. Lasek-Nesselquist E, Gogarten JP (2013) The effects of model choice and mitigating bias on the ribosomal tree of life. *Mol Phylogenet Evol* 69(1):17–38.
35. Ho SYW, Duchêne S (2014) Molecular-clock methods for estimating evolutionary rates and timescales. *Mol Ecol* 23(24):5947–5965.
36. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591.
37. Zheng Y, Wiens JJ (2015) Do missing data influence the accuracy of divergence-time estimation with BEAST? *Mol Phylogenet Evol* 85:41–49.
38. Amard B, Bertrand-Sarfati J (1997) Microfossils in 2000 Ma old cherty stromatolites of the Franceville Group, Gabon. *Precambrian Res* 81:197–221.
39. Luo G, et al. (2016) Rapid oxygenation of Earth's atmosphere 2.33 billion years ago. *Sci Adv* 2(5):e1600134–e1600134.
40. Valley JW, et al. (2014) Hadean age for a post-magma-ocean zircon confirmed by atom-probe tomography. *Nat Geosci* 7(3):219–223.
41. Tomitani A, Knoll AH, Cavanaugh CM, Ohno T (2006) The evolutionary diversification of cyanobacteria: molecular–phylogenetic and paleontological perspectives. *Proc Natl Acad Sci* 103(14):5442–5447.
42. Butterfield NJ (2015) Proterozoic photosynthesis - a critical review. *Palaeontology* 58(6):953–972.
43. Horodyski RJ, Donaldson JA (1980) Microfossils from the Middle Proterozoic Dismal Lakes Group, Arctic Canada. *Precambrian Res* 11:125–159.

44. Duchêne S, Lanfear R, Ho SYW (2014) The impact of calibration and clock-model choice on molecular estimates of divergence times. *Mol Phylogenet Evol* 78:277–289.
45. Toussaint EF, Condamine FL (2016) To what extent do new fossil discoveries change our understanding of clade evolution? A cautionary tale from burying beetles (Coleoptera: *Nicrophorus*). *Biol J Linn Soc* 117(4):686–704.
46. 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(1):15–26.
47. Evans PN, et al. (2015) Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* 350:434–438.
48. Vanwonterghem I, et al. (2016) Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat Microbiol* 1:16170.
49. Barker JF, Fritz P (1981) Carbon isotope fractionation during microbial methane oxidation. *Nature* 293:289–291.
50. Reeburgh WS (2007) Oceanic Methane Biogeochemistry. *Chem Rev* 107(2):486–513.
51. Holler T, et al. (2009) Substantial 13C/12C and D/H fractionation during anaerobic oxidation of methane by marine consortia enriched in vitro. *Environ Microbiol Rep* 1(5):370–376.
52. Suda K, et al. (2014) Origin of methane in serpentinite-hosted hydrothermal systems: The CH4–H2–H2O hydrogen isotope systematics of the Hakuba Happo hot spring. *Earth Planet Sci Lett* 386:112–125.
53. Blank CE (2009) Phylogenomic Dating—The Relative Antiquity of Archaeal Metabolic and Physiological Traits. *Astrobiology* 9(2):193–219.
54. David LA, Alm EJ (2011) Rapid evolutionary innovation during an Archaean genetic expansion. *Nature* 469(7328):93–96.
55. Shih PM, Matzke NJ (2013) Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins. *Proc Natl Acad Sci* 110(30):12355–12360.
56. Harmon LJ, et al. (2013) Arbor: comparative analysis workflows for the Tree of Life. *PLoS Curr* 5. doi:10.1371/currents.tol.099161de5eabdee073fd3d21a44518dc.
57. Pyron RA, Burbrink FT (2013) Phylogenetic estimates of speciation and extinction rates for testing ecological and evolutionary hypotheses. *Trends Ecol Evol* 28(12):729–736.
58. Ng J, Smith SD (2014) How traits shape trees: new approaches for detecting character state-dependent lineage diversification. *J Evol Biol* 27(10):2035–2045.
59. Rabosky DL (2014) Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE* 9(2):e89543.
60. Revell LJ (2014) Ancestral character estimation under the threshold model from quantitative genetics. *Evolution* 68(3):743–759.
61. Stadler T, Rabosky DL, Ricklefs RE, Bokma F (2014) On age and species richness of higher taxa. *Am Nat* 184(4):447–455.
62. Uyeda JC, Harmon LJ, Blank CE (2016) A comprehensive study of cyanobacterial morphological and ecological evolutionary dynamics through deep geologic time. *PLOS ONE* 11(9):e0162539.
63. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.
64. Melby TE, Ciampaglio CN, Briscoe G, Erickson HP (1998) The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: long, antiparallel coiled coils, folded at a flexible hinge. *J Cell Biol* 142(6):1595–1604.
65. Penn O, et al. (2010) GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic Acids Res* 38(S2):W23–W28.

66. Tan G, et al. (2015) Current Methods for Automated Filtering of Multiple Sequence Alignments Frequently Worsen Single-Gene Phylogenetic Inference. *Syst Biol* 64(5):778–791.
67. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
68. Le SQ, Dang CC, Gascuel O (2012) Modeling Protein Evolution with Several Amino Acid Replacement Matrices Depending on Site Rates. *Mol Biol Evol* 29(10):2921–2936.
69. Kück P, Meusemann K (2010) FASconCAT: Convenient handling of data matrices. *Mol Phylogenet Evol* 56(3):1115–1118.
70. Lartillot N, Lepage T, Blanquart S (2009) PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25(17):2286–2288.
71. Quang LS, Gascuel O, Lartillot N (2008) Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* 24(20):2317–2323.
72. Schirrmeister BE, de Vos JM, Antonelli A, Bagheri HC (2013) Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event. *Proc Natl Acad Sci* 110(5):1791–1796.
73. Schirrmeister BE, Gugger M, Donoghue PCJ (2015) Cyanobacteria and the Great Oxidation Event: evidence from genes and fossils. *Palaeontology* 58(5):769–785.
74. Schirrmeister BE, Sanchez-Baracaldo P, Wacey D (2016) Cyanobacterial evolution during the Precambrian. *Int J Astrobiol*:1–18.
75. Warnock RCM, Yang Z, Donoghue PCJ (2012) Exploring uncertainty in the calibration of the molecular clock. *Biol Lett* 8(1):156–159.

FIGURES

Fig. 1. Concatenated PhyloBayes gene tree of SMC, ScpA, and ScpB for Euryarchaeota (methanogenic lineages in red), with HGT to Aquificales (blue), and Cyanobacteria (green). Numbers at nodes represent bootstrap percentages (unlabeled nodes have 100% support) / posterior probabilities (unlabeled nodes have pp = 1.00). Monophyly of the methanogen node was recovered by PhyloBayes (pp = 0.95), and was not recovered by RaxML (bootstrap = 52% for alternative topology), hence the Bayesian topology is shown. Nodes not supported in the Bayesian topology are indicated by a dash (-).

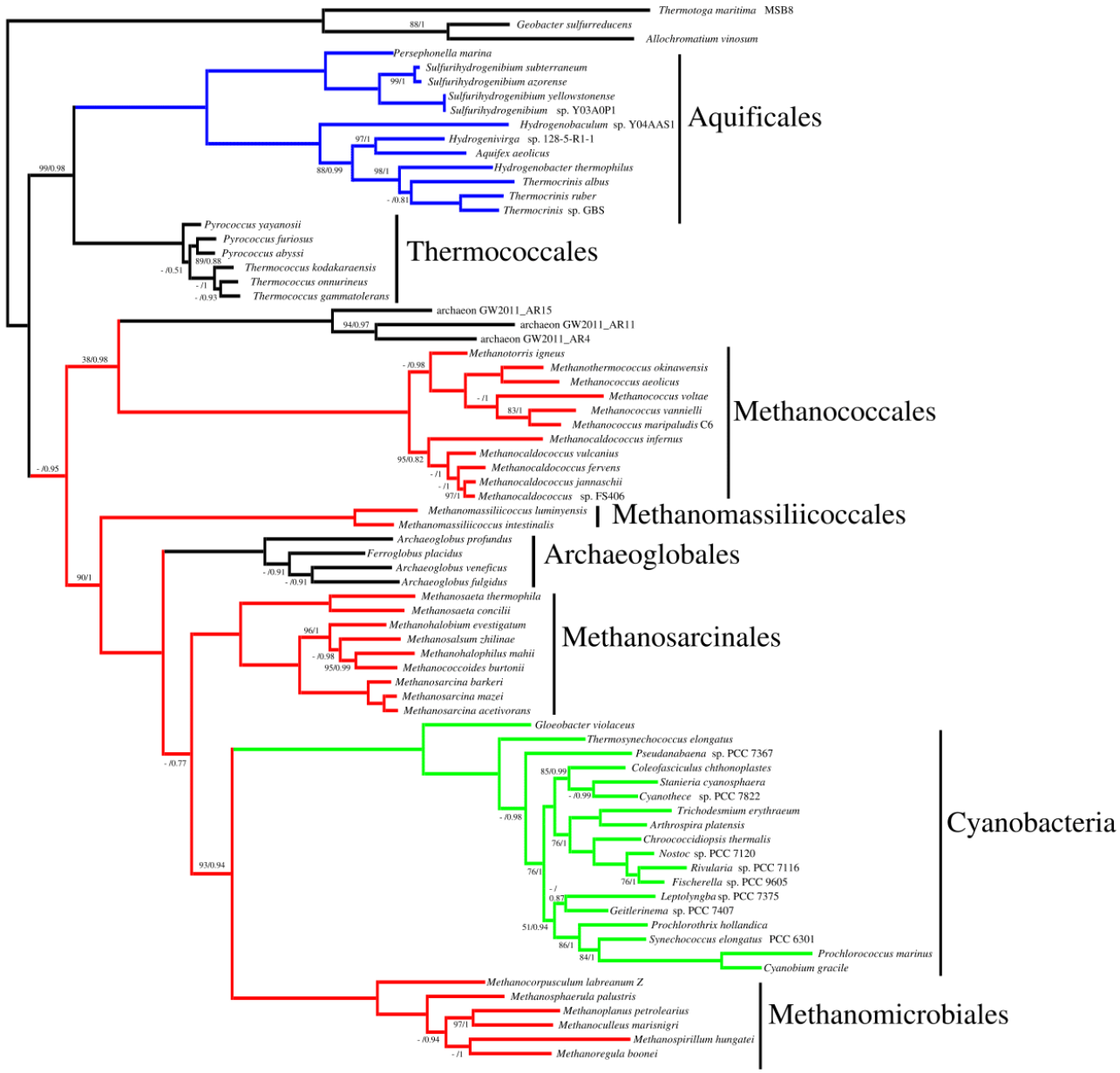

Fig. 2. Comparisons of 95% CI date estimates for Cyanobacteria (green), methanogenic Euryarchaeota (red), and crown Euryarchaeota (grey). Lower dotted horizontal line represents the GOE (39); upper dotted line represents the oldest zircons (40). **(A)** Separate effects of the Euryarchaeota species tree (does not include Cyanobacteria), HGT gene tree (does not include a root estimate, because the SMC complex is not found in all methanogens), and meta-alignment. The HGT gene tree and meta-alignment are calibrated with 1.2 Ga akinete fossils (43). **(B)** Effect of simulated fossil constraints. Filled circles indicate empirical fossil ages (38, 43).

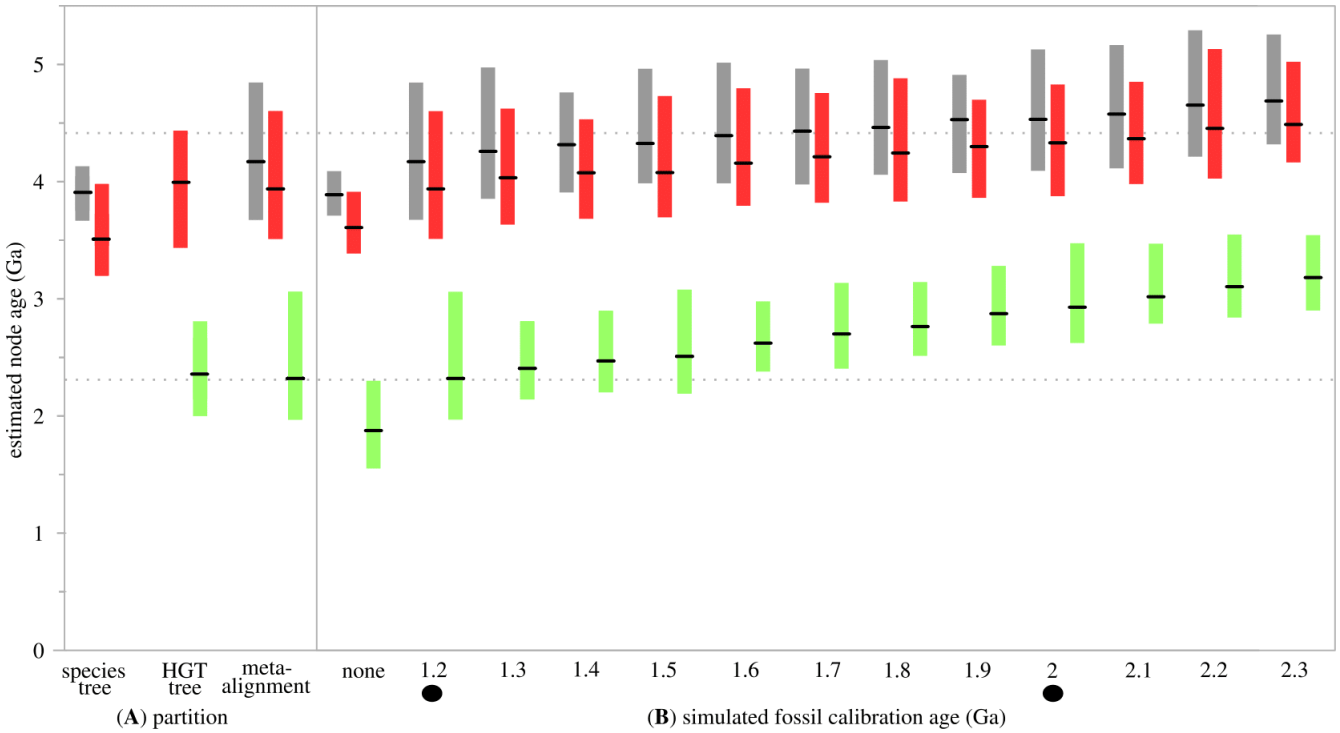

Fig. 3. Most conservative divergence time estimates of Euryarchaeota + Cyanobacteria from meta-alignment. Branch lengths for Euryarchaeota and Cyanobacteria are determined by ribosomal alignment partitions; the branch length of the cyanobacterial stem is determined by the SMC (HGT) partition. The fossil calibration from 1.2 Ga akinetes (43) is indicated by a filled circle. Bars on nodes indicate 95% confidence intervals. Note the GOE age is based on new sulfur isotope measurements from (39), and is thus younger than the base of the Proterozoic.

SI APPENDIX

Supplemental Methods

Pairwise Distances. Pairwise distances between all shared taxa were extracted from the phylogenies generated from the SMC complex (HGT) and meta-alignments (**Fig. 1** and topology of **Fig. 3**, respectively) using T-REX (1). A plot of these distances (**Fig. S5**) shows a generally linear relationship between alignments for methanogen and cyanobacterial groups, with consistently slightly longer branches within the SMC complex gene tree.

HGT Branch Length Simulations. We tested whether the branch length of the HGT itself had a significant effect on the assessed divergence times. Ten simulations were generated for each of two trees, with the same topology as the meta-alignment (i.e. that depicted in **Fig. 3**), but in which the reticulating branch length was altered, by either doubling or halving its length. In this way, the effect of rate changes along a reticulating branch induced via HGT on the molecular clock model could be observed. For each branch length simulation, a divergence time analysis was performed on the simulated sequences in PhyloBayes (2), using the same parameters as the empirical data and the fossil akinete calibration from 1.2 Ga (3). All runs on simulated data converged.

Missing Data Simulations. To test the role of missing data in the meta-alignment, a control simulation was created. Ten simulated alignments were generated using the PAML4 module evolver (4), using the LG model with eight discrete gamma-distributed categories and a shape parameter of alpha=0.69, as best fit to the concatenated ribosomal-SMC protein meta-alignment by ProtTest (5). Amino acid frequencies were matched to those observed within the meta-alignment. Sequences were simulated along the inferred maximum likelihood tree recovered from the concatenated dataset (i.e. the topology in **Fig. 3**). Sequences were initially simulated for 11223 sites, to match the full number of sites within the meta-alignment that have less than 50% gaps within sequence blocks.

For each simulated alignment, blocks of of equal length and taxonomic distribution to the regions of missing data within the observed alignment (e.g. the blocks between euryarchaeal and cyanobacterial ribosomal proteins) within meta-alignments were replaced with gaps. 6060 sites in the Cyanobacteria ribosomal partition were replaced with gaps, corresponding to euryarchaeal ribosomal sites absent in Cyanobacteria; 575 sites in the Euryarchaeota ribosomal partition were replaced with gaps corresponding to euryarchaeal taxa that do not contain an included SMC homolog; and 4588 sites in the Euryarchaeota ribosomal partition were replaced with gaps corresponding to cyanobacterial ribosomal sites absent in Euryarchaeota. For each control and missing simulation, a divergence time analysis was performed on the simulated sequences in PhyloBayes, using the same parameters as the empirical data and the fossil akinete calibration from 1.2 Ga (3).

SI APPENDIX REFERENCES

1. Boc A, Diallo AB, Makarenkov V (2012) T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res* 40(W1):W573–W579.
2. Lartillot N, Lepage T, Blanquart S (2009) PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25(17):2286–2288.
3. Horodyski RJ, Donaldson JA (1980) Microfossils from the Middle Proterozoic Dismal Lakes Group, Arctic Canada. *Precambrian Res* 11:125–159.
4. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591.
5. Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21(9):2104–2105.
6. Ho SYW, Duchêne S (2014) Molecular-clock methods for estimating evolutionary rates and timescales. *Mol Ecol* 23(24):5947–5965.

SI FIGURES AND TABLES

Fig. S1. RaxML gene tree of SMC for all 307 Cyanobacteria taxa available in GenBank, excluding sequences transferred from other bacterial clades.

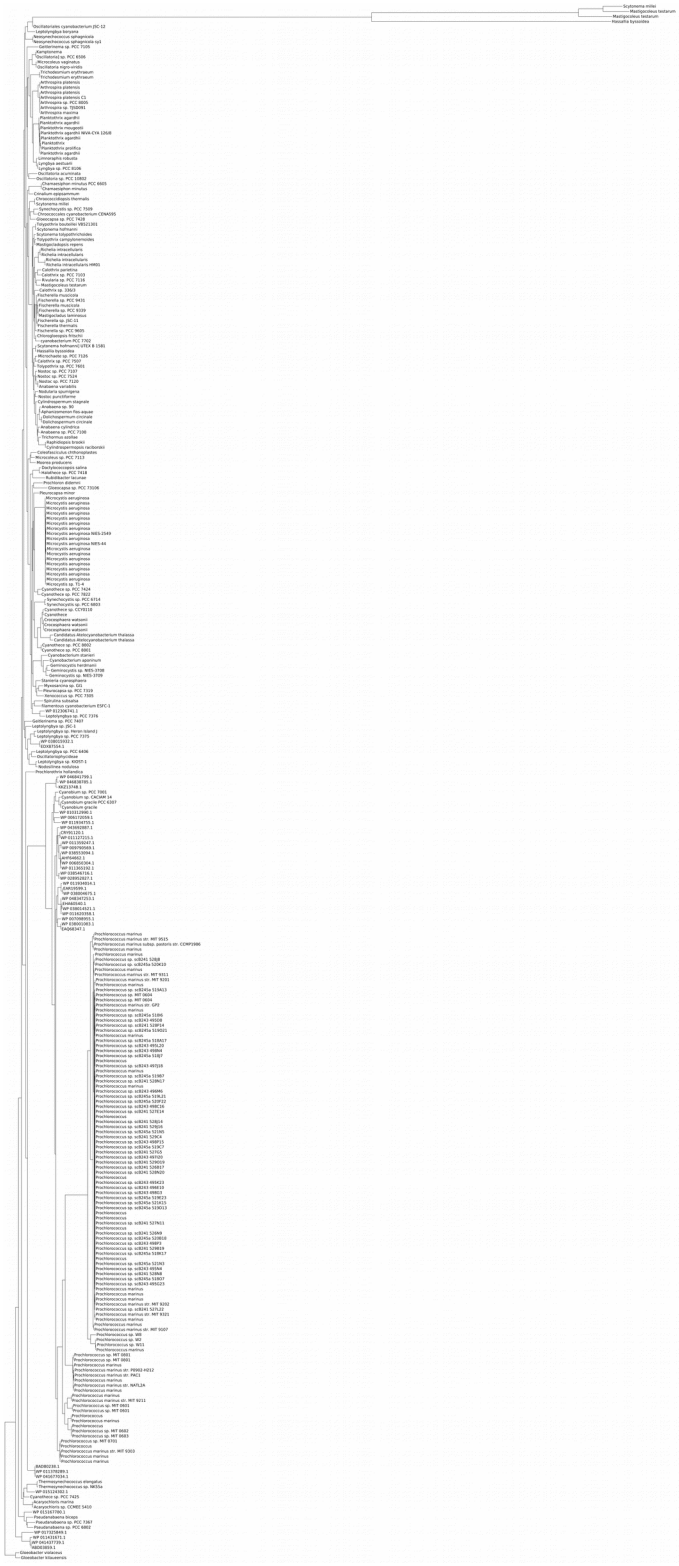

Fig. S2. RaxML gene tree of SMC for Euryarchaeota (methanogenic lineages in red), with HGT to Aquificales (blue), and Cyanobacteria (green).

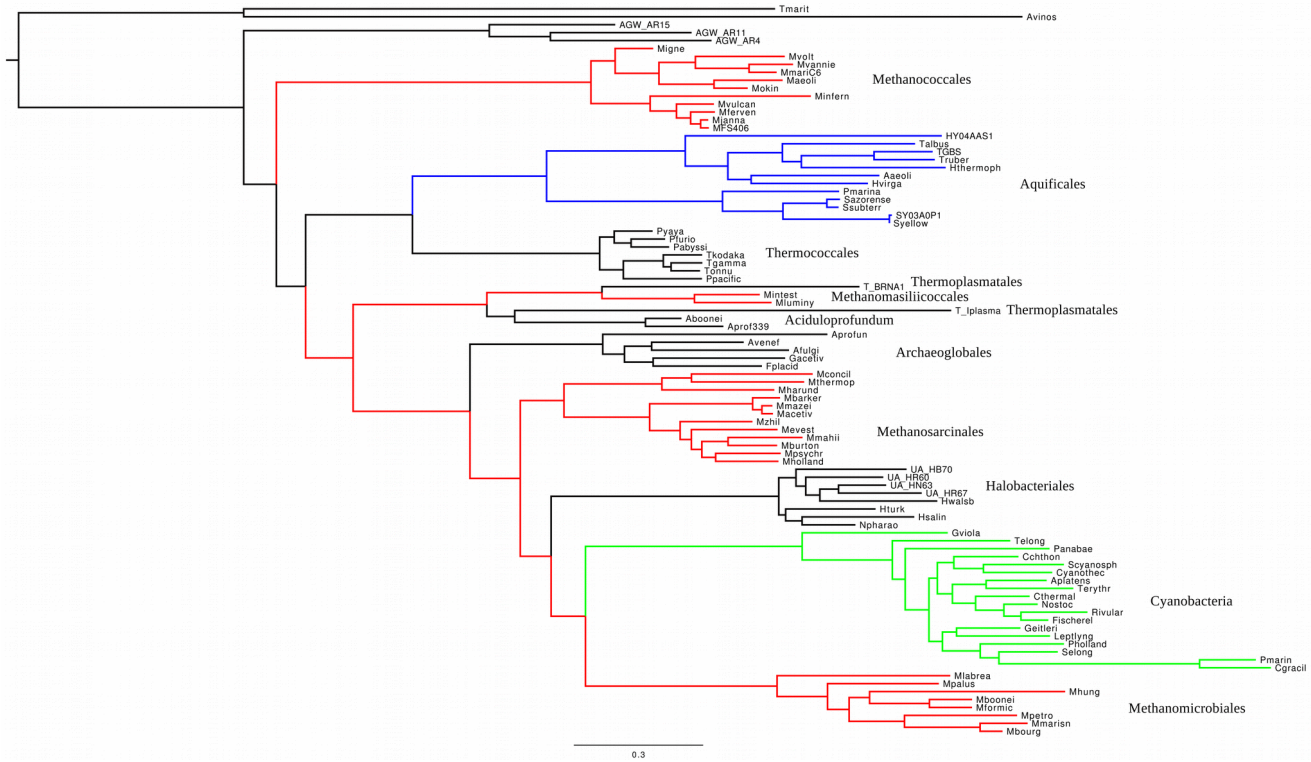

Fig. S3. RaxML gene tree of ScpA for Euryarchaeota (methanogenic lineages in red), with HGT to Aquificales (blue), and Cyanobacteria (green).

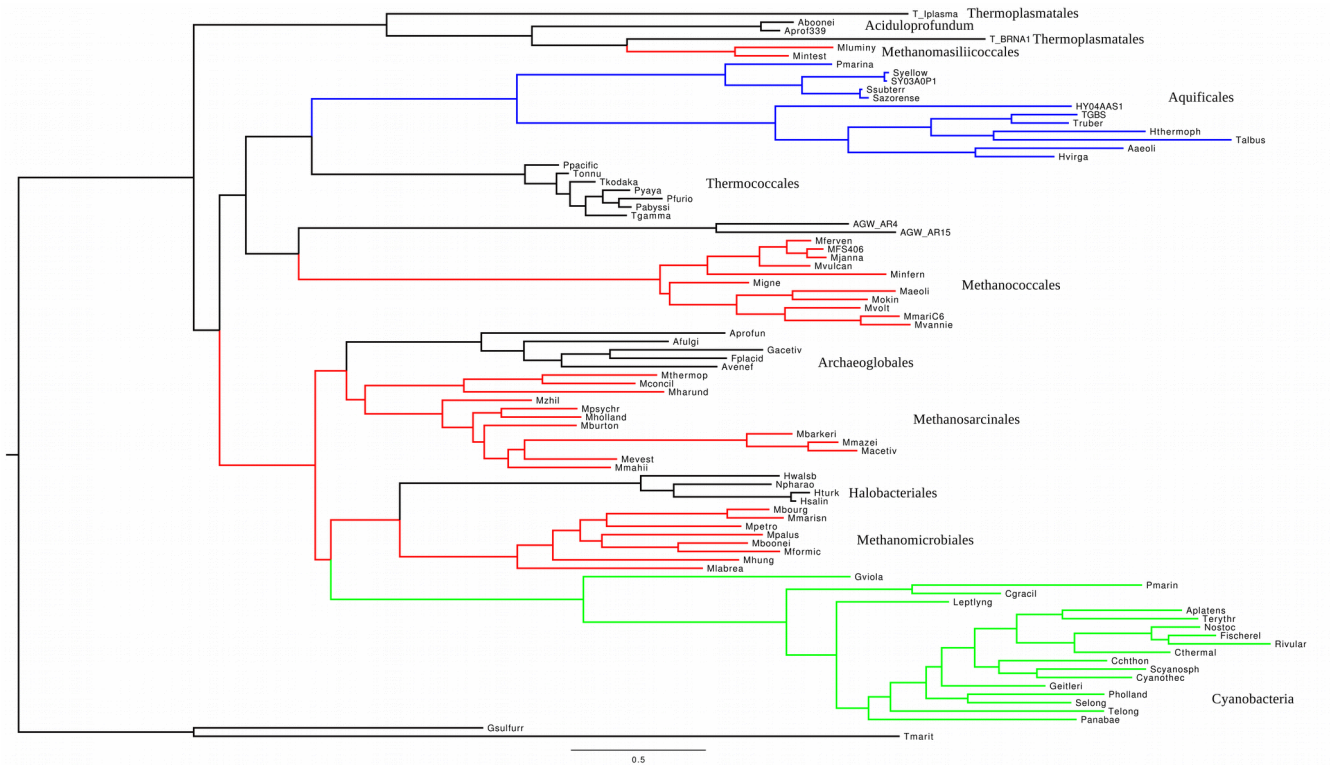

Fig. S4. RaxML gene tree of ScpB for Euryarchaeota (methanogenic lineages in red), with HGT to Cyanobacteria (green).

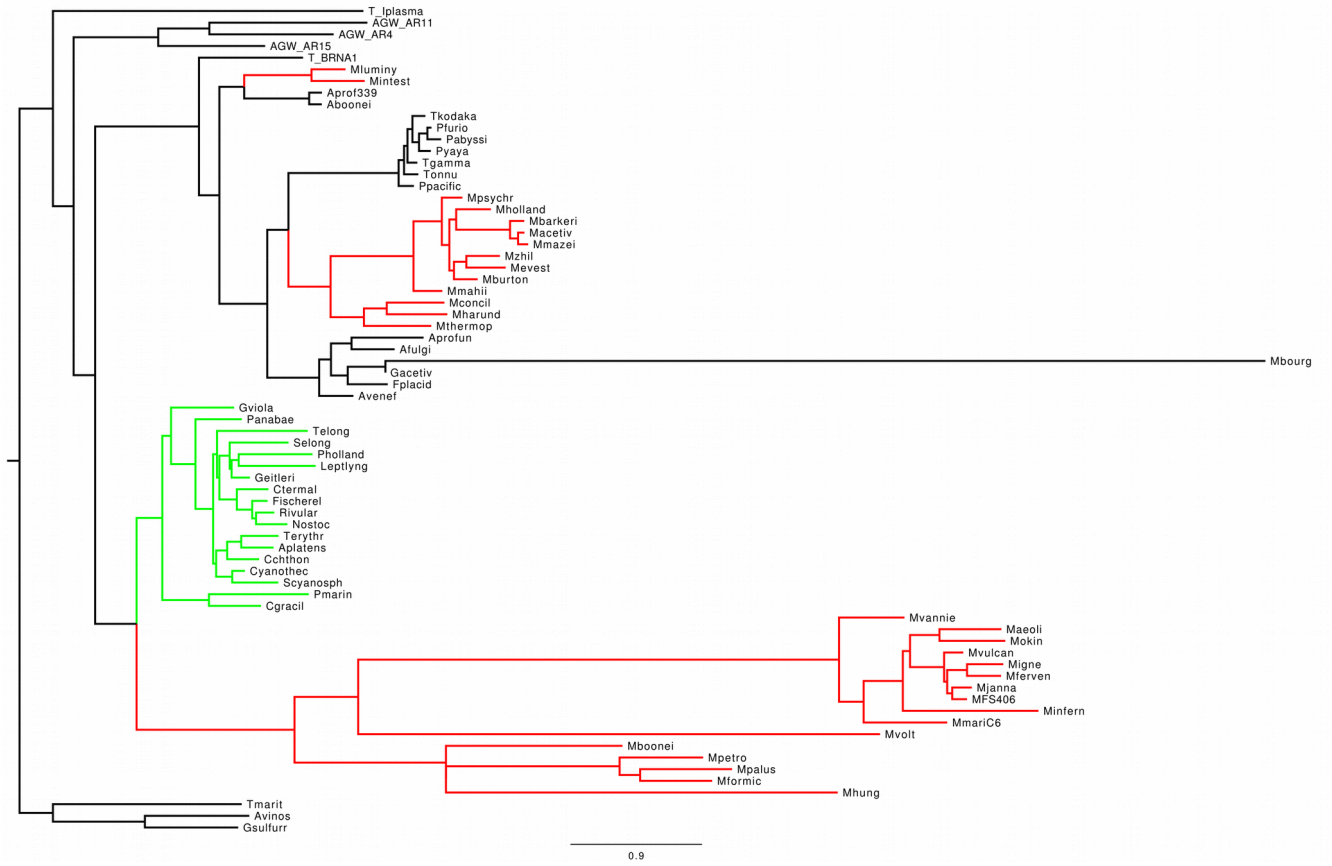

Table S1. Bootstrap bipartitions for the concatenated alignment of SMC + ScpA + ScpB, calculated in RaxML.

Cyanobacteria sister group	**Bootstrap %**	**Bootstrap % with Halobacteriales removed**
Methanomicrobiales	33	93
Methanomicrobiales + Halobacteriales	20	0
Halobacteriales	17	0
Methanosarcinales + Halobacteriales	15	0
Methanosarcinales + Halobacteriales + Archaeoglobales	7	0
Methanomicrobiales + Methanosarcinales + Halobacteriales	4	0
Methanomicrobiales + Methanosarcinales + Halobacteriales + Archaeoglobales	2	0
Archaeoglobales	2	0
Methanosarcinales	0	6
Methanomicrobiales + Methanosarcinales	0	1

Fig. S5. Pairwise distances between branches of the SMC complex gene tree and the meta-alignment concatenated tree. Both the root of methanogens and cyanobacterial recipient (green) clades show the same trend, excepting Methanomicrobiales (orange), which are on a disproportionately long branch within the SMC tree. The stem lineage branch lengths are also plotted for Cyanobacteria, Methanomicrobiales, and Methanosarcinales, showing that the long cyanobacterial stem and short Methanosarcinales (pink) stem are proportional between trees and fall on this diagonal (thus no lineage effects are observed among these clades (6)), while the Methanomicrobiales (donor lineage) stem does not (thus this clade may have a lineage specific rate).

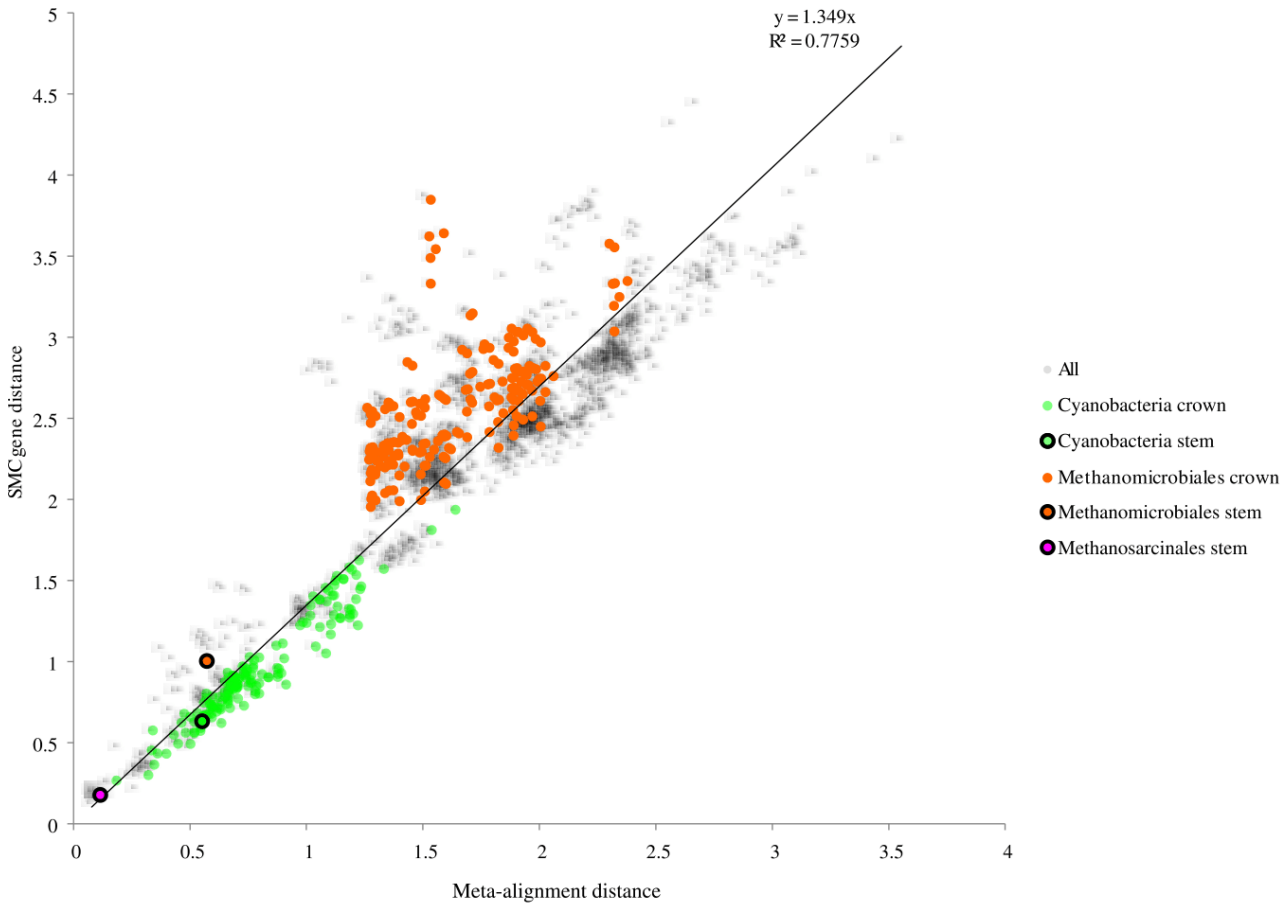

Fig. S6. Comparisons of simulated 95% CI date estimates for Cyanobacteria (green), the reticulating node (yellow), and the methanogen donor lineage (orange). Simulations are depicted with the full observed reticulating branch length, with the reticulating branch length half of that observed empirically, and with the reticulating branch length double that observed. Only the mean of 10 simulations is depicted for each mean age and 95% CI.

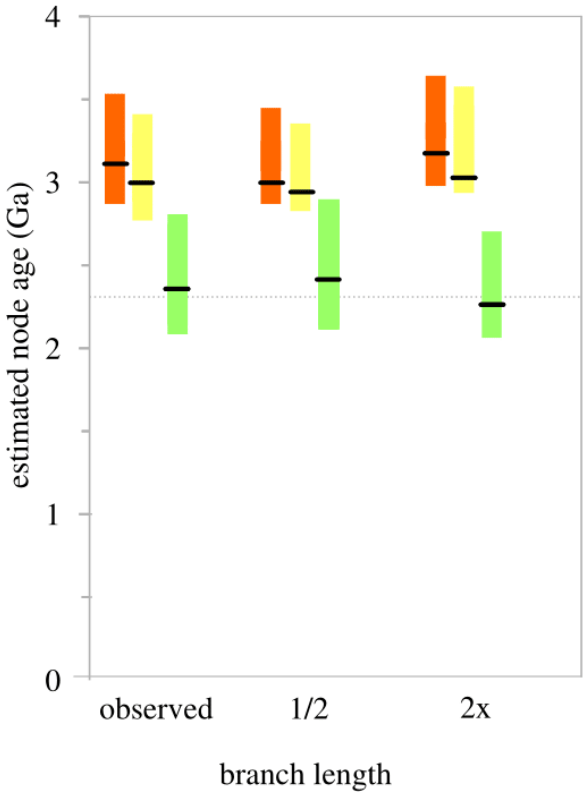

Fig. S7. Comparisons of 95% CI date estimates for simulated full length alignments and with blocks of missing data to represent that observed in the meta-alignment. All 10 simulations of each dataset depicted have empirical branch lengths from the meta-alignment. **A)** Cyanobacteria (green) simulated age. **B)** The reticulating node (yellow) simulated age. **C)** Methanomicrobiales (orange) simulated age.

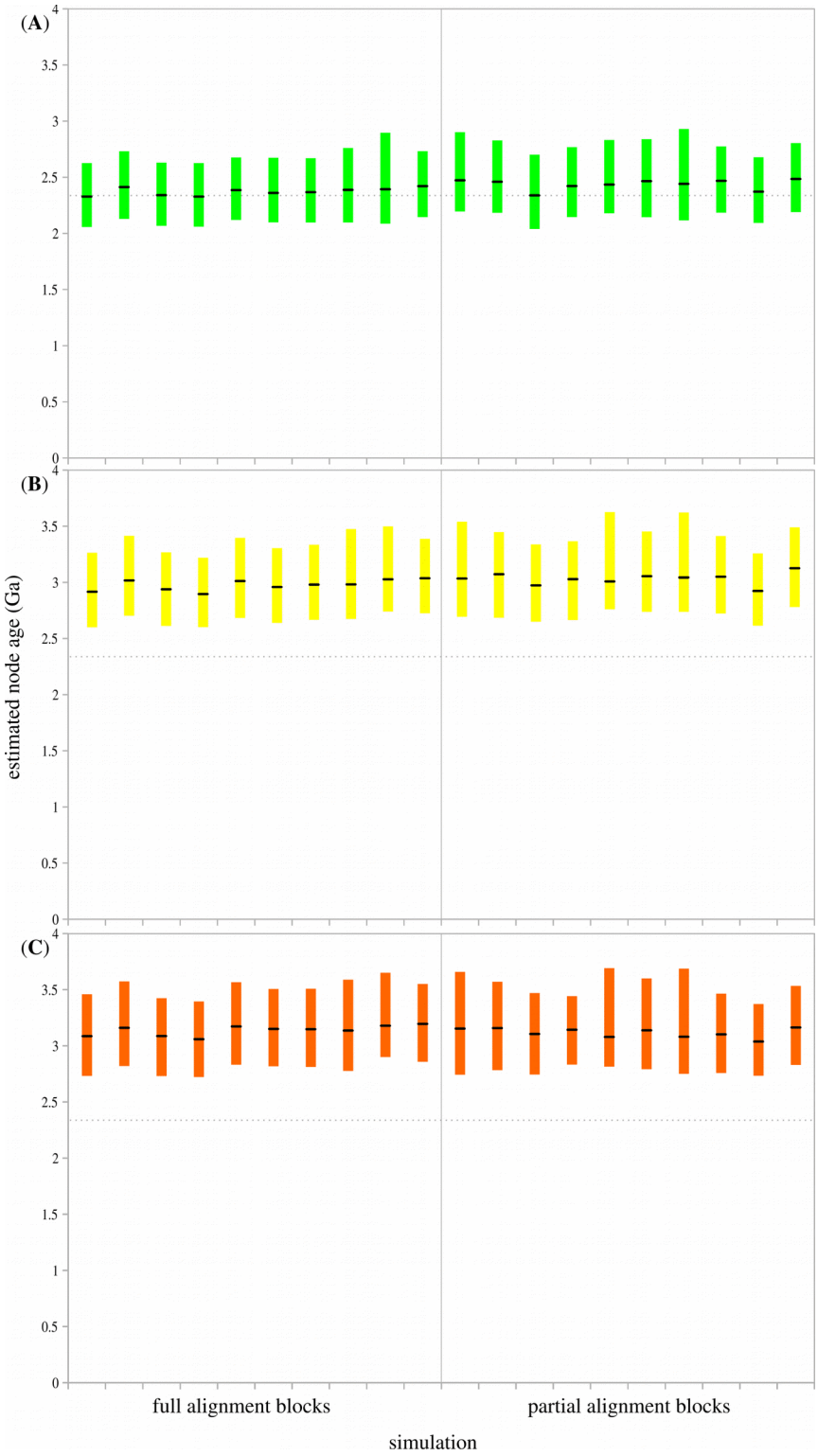

Table S2. Number of amino acid sites in each single-gene alignment before and after masking with GUIDANCE. Number of sites found in >4 taxa in parentheses.

Gene	Unmasked	Masked
SMC	1576 (1323)	729 (588)
ScpA	560 (423)	136 (96)
ScpB	489 (408)	100 (0)

Table S3. List of ribosomal proteins. These were aligned and concatenated for Euryarchaeota, and separately for Cyanobacteria, to build the meta-alignment.

Large subunit (50S)	Small subunit (30S)
L1	S2
L2	S3
L3	S4
L4	S5
L5	S7
L6	S8
L10	S9
L13	S10
L14	S11
L15	S12
L18	S13
L22	S14
L23	S15
L24	S17
L29	S19