LETTERS

Palaeotemperature trend for Precambrian life inferred from resurrected proteins

Eric A. Gaucher[1], Sridhar Govindarajan[2] & Omjoy K. Ganesh[3]

Biosignatures and structures in the geological record indicate that microbial life has inhabited Earth for the past 3.5 billion years or so[1,2]. Research in the physical sciences has been able to generate statements about the ancient environment that hosted this life[3–6]. These include the chemical compositions and temperatures of the early ocean and atmosphere. Only recently have the natural sciences been able to provide experimental results describing the environments of ancient life. Our previous work with resurrected proteins indicated that ancient life lived in a hot environment[7,8]. Here we expand the timescale of resurrected proteins to provide a palaeotemperature trend of the environments that hosted life from 3.5 to 0.5 billion years ago. The thermostability of more than 25 phylogenetically dispersed ancestral elongation factors suggest that the environment supporting ancient life cooled progressively by 30 °C during that period. Here we show that our results are robust to potential statistical bias associated with the posterior distribution of inferred character states, phylogenetic ambiguity, and uncertainties in the amino-acid equilibrium frequencies used by evolutionary models. Our results are further supported by a nearly identical cooling trend for the ancient ocean as inferred from the deposition of oxygen isotopes. The convergence of results from natural and physical sciences suggest that ancient life has continually adapted to changes in environmental temperatures throughout its evolutionary history.

Computational reconstruction and laboratory resurrection of ancestral sequences provide an opportunity to rewind the 'tape of life'. This form of time travel has been exploited to improve our understanding of the molecular adaptation of substrate specificity, the response of organisms to external stimuli, and the environmental temperatures of ancient organisms, among others (reviewed in ref. 9).

Ancestral sequence reconstruction uses standard statistical theory to generate posterior probabilities of different reconstructions given the data at a site from aligned sequences. For each site of the inferred sequence at a phylogenetic node, posterior values for all 20 amino acids are calculated and represent the probability that a particular amino acid occupied a specific site in the protein during its evolutionary history. This posterior probability distribution is calculated from patterns of amino acids in modern sequences as described by a phylogeny, a matrix of amino-acid replacement probabilities, amino-acid equilibrium (stationary) frequencies, phylogenetic branch lengths and site-specific replacement rates. The most probabilistic ancestral sequence (M-PAS) uses the amino acid with the highest posterior probability at each site within the distribution.

Despite insightful studies, the field of ancestral sequence reconstruction is encumbered by its inability to know whether inferred sequences truly recapitulate ancestral forms[10]. Practitioners in the field acknowledge a certain degree of inaccuracy associated with reconstructing ancestral sequences. The concern is not necessarily whether the resurrected form has the exact composition (genotype) of the true ancestral form, but rather that the resurrected form displays the exact behaviour (phenotype). A reconstructed sequence can be considered a consensus of a gene distributed throughout a population before species divergence or before gene duplication. Inaccuracies in a reconstructed sequence can result from sequence variation in the gene itself within an ancient population. If one assumes that the variants of a homologous gene within a population had the same phenotype at a specific geological time, it does not necessarily matter which individual genotype is reconstructed.

This assumption is invalid if recombination of individual genotypes generates new phenotypes and if the reconstructed ancestral gene itself is a consensus of those genotypes. Additional concerns arise if the reconstruction process generates inaccurate sequences, either because of bias in the evolutionary models used to infer ancestral states or because of phylogenetic conditions such as long branches and incorrect branching patterns[10–12].

Bias in the reconstruction process can, for example, lead to a preponderance of hydrophobic amino acids in an ancestral sequence. This bias results from long branches in a phylogeny combined with the fact that hydrophobic residues have high equilibrium frequencies in the amino-acid replacement matrices used to infer ancestral states. An increased proportion of hydrophobic residues in an ancestral protein has the potential effect of spuriously increasing thermostability. A similar form of bias is produced when the equilibrium frequencies themselves are incorrect. Equilibrium frequencies are derived from amino-acid occurrence in modern proteins. It may be incorrect to use these frequencies when reconstructing ancestral sequences if amino-acid occurrence has evolved over time, resulting in an inhomogeneous process[13].

We address these multiple concerns. First, we generate weighted random sequences sampled from the posterior distribution of ancestral character states to address whether bias in the amino-acid equilibrium frequencies affects the phenotypes of the inferred proteins. Second, we calculate ancestral amino-acid equilibrium frequencies and use these as an alternative to modern equilibrium frequencies to determine potential effects on ancestral phenotypes. Third, we reconstruct ancestral proteins across two competing bacterial phylogenies to determine the effects of topology on the ancestral phenotypes. For both phylogenies we assume the root of the tree to lie between bacteria and archaeans/eukaryotes, despite the suggestion that bacteria may be paraphyletic[14].

Elongation factor (EF) Tu (Bacteria)/1A (Archaea and Eukarya) is a suitable protein family with which to address the above concerns. The thermal stabilities of EFs are correlated with the growth temperature of their host organisms. Thus, EFs are optimally stable at temperatures of 20–45 °C, 45–80 °C and more than 80 °C when isolated from mesophiles, thermophiles and hyperthermophiles, respectively.

[1]Foundation for Applied Molecular Evolution, Gainesville, Florida 32601, USA. [2]DNA2.0, Inc., Menlo Park, California 94025, USA. [3]Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, Florida 32610, USA.

This relationship is consistent with a correlation coefficient of 0.91 between the melting temperatures (T_m values) of proteins and environmental temperatures of their host organisms[15].

Reconstructions of ancestral EF sequences were computed across two bacterial phylogenies selected from the literature[16,17]. Both phylogenies were constructed from the concatenation of numerous gene families and are therefore less susceptible to systematic error than phylogenies based on single genes. The two phylogenies capture the main competing views for bacterial relationships. One model posits that hyperthermophilic lineages occupy basal branches of the bacterial tree, whereas the other places these lineages in a more derived portion of the tree. To accommodate the latter model, a phylogeny was selected in which the Firmicute lineage (void of hyperthermophiles) was located at the base of the bacterial tree, although other topologies have been suggested[18].

The thermostability of modern and ancestral EF proteins was monitored by means of circular dichroism spectroscopy. The T_m values of two modern EFs, from *Escherichia coli* and *Thermus thermophilus* (HB8), were determined as 42.8 °C and 76.7 °C, respectively. These values highlight the relationship between EF stability and the optimal growth temperature of their respective hosts, about 40 °C and about 74 °C (see Supplementary Information)[19].

T_m values for ancestral EF proteins were determined across the two phylogenies (Fig. 1). The thermostability profiles of the ancestral proteins display the same general trend even though the two phylogenies represent competing hypotheses. Ancestral EF proteins resurrected at basal nodes are compatible with thermophilic environments, whereas ancestral proteins from more derived nodes are compatible with cooler environments. Consistent with this temperature trend is the observation that the node representing the presumed last common ancestor of bacteria (and thus the oldest) had the most thermostable protein within each phylogeny (64.8 °C and 73.3 °C). The similarity in thermostability (less than 9 °C difference) between these two ancestral proteins is significant because the sequences were identical across only 78% of the amino-acid sites (see Supplementary Information).

Systematic bias in the reconstruction of ancestral character states has the potential to generate incorrect inferences of ancient biomolecules[10]. As discussed above, we considered two potential sources of bias: first, incorrect equilibrium frequencies in the amino-acid replacement matrix, and second, selecting the M-PAS versus weighted random samplings from the posterior distribution of ancestral states. Ancestral amino-acid frequencies from a set of 31 protein families present in the last common ancestor of bacteria were

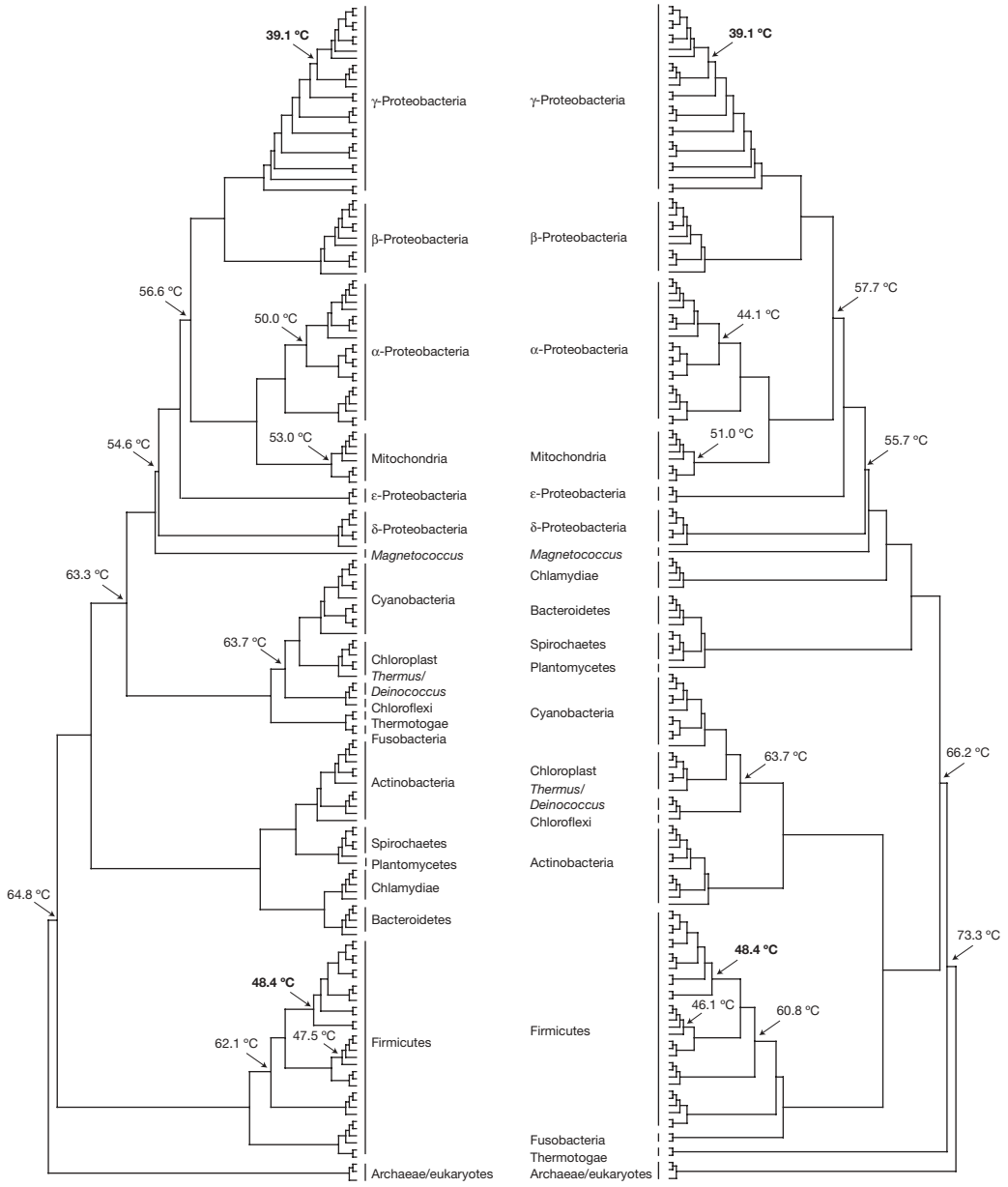

Figure 1 | Cladograms of the trees used to resurrect EF proteins. The phylogeny on the left follows ref. 17, whereas that on the right follows ref. 16. Melting temperatures for ancestral EF proteins are shown at their corresponding nodes. Temperatures in bold represent identical sequences at analogous nodes between the two trees. Errors associated with measurements of circular dichroism are negligible (see Supplementary Information).

estimated as described previously[13]. These 31 families were extracted from 16 species (8 mesophiles and 8 thermophiles) whose optimal growth temperatures ranged from 25 °C to 95 °C. The inferred ancestral frequencies then served as equilibrium frequencies in an amino-acid replacement matrix, and an ancient EF was reconstructed at the node representing the last common ancestor of bacteria from the phylogeny from ref. 17 by using this matrix. The ancient protein had a T_m of 61.4 °C, in contrast with T_m = 64.8 °C for the analogous ancient protein when standard frequencies were applied (Fig. 2). These two proteins differed at 14 amino-acid sites.

A random distribution of 10,000 sequences weighted in accordance with the posterior distribution for the node presenting the last common ancestor of bacteria from the phylogeny from ref. 17 was generated computationally to determine whether the EF data set and/or phylogenetic parameters led to bias when selecting the most probable ancestral character states. Five of the 10,000 sequences were then randomly selected and synthesized in the laboratory. The melting temperatures for these five sequences ranged from 60.0 °C to 66.3 °C, in contrast with T_m = 64.8 °C for the M-PAS at this node (Fig. 2). The number of amino-acid differences between any of the five random sequences compared with the M-PAS ranged from 7 to 18 sites.

The environmental temperature of ancient bacteria inferred from resurrected EF proteins can be connected to divergence times of major bacterial lineages to gain a more detailed understanding of temperature trends for Precambrian life[16]. Divergence estimates from ref. 16 were applied to nodes in the current study. Figure 3 highlights the progressive cooling trend of ancient EF proteins from about 3.5 Gyr ago to 500 Myr ago. This temperature trend is strikingly similar to the temperature trend of the ancient ocean inferred from the deposition of oxygen and silicon isotopes[3–5].

Reconstruction of ancestral EF proteins throughout the bacterial domain of life suggests that the organisms that hosted these extinct biomolecules lived in environments that have cooled progressively for about 3 Gyr. This evidence is predicated on multiple assumptions. For instance, it assumes that the reconstruction of ancestral sequences recapitulates ancient phenotypes and that phylogenies and divergence dates capture the evolutionary relationships and timing of bacterial divergences.

The observation that five samples from the posterior distribution had equivalent thermostability profiles to that of the M-PAS (Fig. 2) suggests that ancestral resurrections are robust for phenotype even when uncertainties exist in the ancestral sequences themselves. Further, the observation that ancestral amino-acid equilibrium frequencies produce an ancient protein with a phenotype equivalent to that of an ancient protein derived from modern amino-acid frequencies (Fig. 2) demonstrates that ancestral phenotypes can be robust to violations of a priori parameters contained within evolutionary models.

The inability (other than by time travel) to know the true relationships of bacterial lineages and their divergence times should not preclude attempts to understand Precambrian life. Rather, a coherent description of ancient life can be generated when empirical evidence from diverse studies converge on analogous conclusions. For instance, the same palaeotemperature trend was observed for ancestral EF proteins regardless of the phylogeny. For the phylogeny with divergence dates, this trend was substantiated when aligned with the inferred palaeotemperature curve of the ancient ocean.

These descriptions are particularly useful when they have predictive value. For instance, the last common ancestor of the mitochondrial bacterium is estimated to have lived 1.66–1.88 Gyr ago, on the basis of the T_m values for ancestral EF proteins from the node representing the origins of mitochondria (51.0–53.0 °C) (see Supplementary Information). This is consistent with the origins of mitochondria estimated at 1.8 Gyr ago on the basis of a molecular clock[20], despite the controversial nature of the clock[21] and assuming that the last common mitochondrial bacterium lived at a time close to the endosymbiotic event between α-Proteobacteria and eukaryotic cells.

Our results suggest that early life lived at an environmental temperature similar to those of today's hot springs. Particular geological

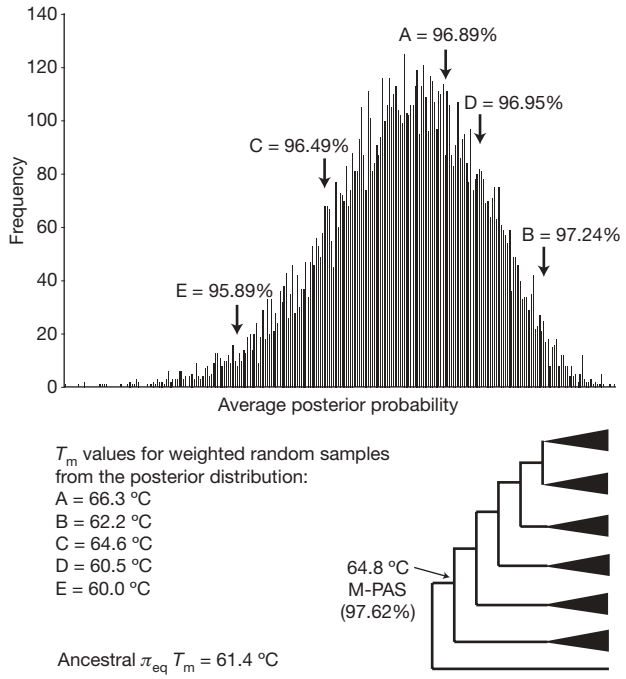

Figure 2 | The EF phenotypes associated with samples drawn from the posterior distribution and ancestral equilibrium frequencies. Both analyses focus on the node corresponding to the last common ancestor (LCA) of bacteria on the phylogeny from ref. 17. The figure shows the distribution of 10,000 weighted random sequences sampled from the posterior distribution associated with the LCA node. For each of the 10,000 sequences, the average posterior probability across all 394 sites is presented on the *x*-axis and the frequency of each average is presented on the *y*-axis. The distribution mean is 96.64%. Five sequences (A–E) were randomly selected from the distribution and synthesized. The corresponding T_m values of the five encoded proteins are presented below the graph, and the average posterior probability across all sites for each sequence is indicated above an arrow. The T_m for the LCA node inferred from ancestral equilibrium frequencies (π_{eq}) is also shown.

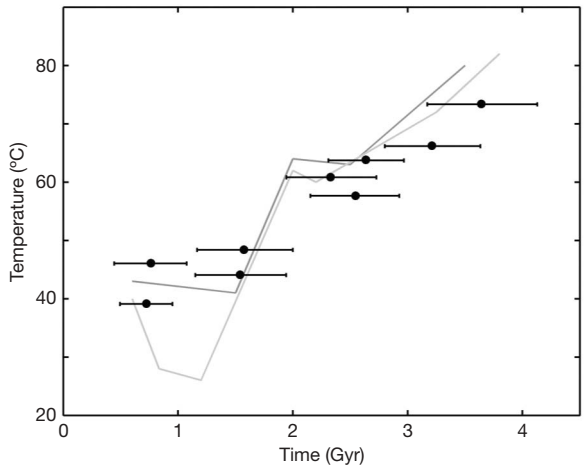

Figure 3 | Plot of ancestral EF melting temperatures against geological time. Molecular clock estimates are shown with their confidence intervals (horizontal bars) from ref. 16, using a 2.3-Ga minimum constraint for the Great Oxidation Event. Solid lines are temperature curves of the ancient ocean inferred from maximum $\delta^{18}O$ (light grey[3,4], dark grey[5]). Although not shown, an analogous trend is seen with $\delta^{30}Si$ isotopes[5].

theory and evidence suggest that the ancient ocean also had temperatures similar to those of hot springs[3,4,22]. As the ocean cooled from 3.5 to 0.5 Gyr ago, life may have responded by adapting its range of growth temperatures to correspond to its environment. This connection assumes that early life lived in the ancient ocean, which seems practical on the basis of geological and biological constraints such as ocean depth and circulation, land mass exposed to the atmosphere, and susceptibility to desiccation and ultraviolet radiation. Alternatively, it is possible that the inferred trend in palaeotemperature reflects an ecological trajectory as ancient bacteria made the transition from hot springs and thermal vents to the open ocean.

We note that correlating isotope ratios ($\delta^{18}O$ and $\delta^{30}Si$) with ancient ocean temperatures is contentious[23,24]. In particular, the correlation could be invalid if isotope ratios were caused by variation in seawater composition alone. This would translate into a more temperate ancient ocean and would be consistent with ancient glaciation events. However, the similarity in palaeotemperature trends inferred from $\delta^{18}O$, $\delta^{30}Si$ and ancient EF proteins is striking. Further, the overall trend is compatible with biological evolution. For instance, the thermostability of ancient EFs suggests that the origins of cyanobacteria occurred at an environmental temperature close to 63.7 °C (Fig. 1). This is consistent with an upper temperature limit of typical cyanobacterial mats in hot springs (about 65 °C)[25].

Overall, the results demonstrate that ancient EF thermostability profiles (phenotypes) are robust to uncertainties and potential biases associated with inferring ancestral character states (genotypes). The results also show how ancestral sequence reconstruction can connect the physical and natural sciences. As an extension, we have determined that certain ancestral EFs are indeed able to participate in peptide elongation when substituted for *E. coli* EFTu in a reconstituted *in vitro* translation system composed of *E. coli* components (data not shown). We expect that this type of assay will allow us to determine some of the underlying molecular mechanisms governing EF proteins as they evolved adaptively along particular branches during their evolutionary history.

METHODS SUMMARY

Bacterial EF homologues were retrieved from GenBank, phylogenetic analysis was performed with MrBayes when necessary[26], and ancestral sequence reconstruction was calculated with PAML[27]. Ancestral amino-acid equilibrium frequencies were calculated as described previously[13].

Codon optimization and gene synthesis were conducted as described previously[28,29]. Ancestral and modern EF proteins were expressed by autoinduction and were purified by affinity chromatography[30]. Thermostability curves of EF proteins were determined by circular dichroism monitored at a wavelength of 222 nm. Data were analysed with MATLAB version 7.4.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 October; accepted 3 December 2007.

1. Buick, R., Dunlop, J. S. R. & Groves, D. I. Stromatolite recognition in ancient rocks—an appraisal of irregularly laminated structures in an early Archean chert–barite unit from North Pole, Western Australia. *Alcheringa* **5,** 161–181 (1981).
2. Hofmann, H. J., Grey, K., Hickman, A. H. & Thorpe, R. I. Origin of 3.45 Ga coniform stromatolites in Warrawoona Group, Western Australia. *Geol. Soc. Am. Bull.* **111,** 1256–1262 (1999).
3. Knauth, L. P. & Lowe, D. R. Oxygen ISOTOPE GEOCHEMISTRY OF CHERTS from Onverwacht Group (3.4 billion years), Transvaal, South Africa, with implications for secular variations in isotopic composition of cherts. *Earth Planet. Sci. Lett.* **41,** 209–222 (1978).
4. Knauth, L. P. & Lowe, D. R. High Archean climatic temperature inferred from oxygen isotope geochemistry of cherts in the 3.5 Ga Swaziland Supergroup, South Africa. *Geol. Soc. Am. Bull.* **115,** 566–580 (2003).
5. Robert, F. & Chaussidon, M. A palaeotemperature curve for the Precambrian oceans based on silicon isotopes in cherts. *Nature* **443,** 969–972 (2006).
6. Shen, Y., Buick, R. & Canfield, D. E. Isotopic evidence for microbial sulphate reduction in the early Archaean era. *Nature* **410,** 77–81 (2001).
7. Gaucher, E. A. in *Ancestral Sequence Reconstruction* (ed. Liberles, D.A.) 20–33 (Oxford Univ. Press, Oxford, 2007).
8. Gaucher, E. A., Thomson, J. M., Burgan, M. F. & Benner, S. A. Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature* **425,** 285–288 (2003).
9. Liberles, D. A. *Ancestral Sequence Reconstruction* (Oxford Univ. Press, Oxford, 2007).
10. Williams, P. D., Pollock, D. D., Blackburne, B. P. & Goldstein, R. A. Assessing the accuracy of ancestral protein reconstruction methods. *PLoS Comput. Biol.* **2,** e69 (2006).
11. Felsenstein, J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27,** 401–410 (1978).
12. Kelchner, S. A. & Thomas, M. A. Model use in phylogenetics: nine key questions. *Trends Ecol. Evol.* **22,** 87–94 (2007).
13. Brooks, D. J. & Gaucher, E. A. in *Ancestral Sequence Reconstruction* (ed. Liberles, D.A.) 200–207 (Oxford Univ. Press, Oxford, 2007).
14. Cavalier-Smith, T. The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int. J. Syst. Evol. Microbiol.* **52,** 7–76 (2002).
15. Gromiha, M. M., Oobatake, M. & Sarai, A. Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins. *Biophys. Chem.* **82,** 51–67 (1999).
16. Battistuzzi, F. U., Feijao, A. & Hedges, S. B. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol. Biol.* **4,** 44 (2004).
17. Ciccarelli, F. D. *et al.* Toward automatic reconstruction of a highly resolved tree of life. *Science* **311,** 1283–1287 (2006).
18. Brochier, C. & Philippe, H. Phylogeny: a non-hyperthermophilic ancestor for Bacteria. *Nature* **417,** 244 (2002).
19. Williams, R. A. D. & da Costa, M. S. in *The Prokaryotes* (eds Balows, A., Truper, H.G., Dworkin, M., Harder, W. & Schleifer, K.-H.) 3745–3753 (Springer, New York, 1992).
20. Hedges, S. B. *et al.* A genomic timescale for the origin of eukaryotes. *BMC Evol. Biol.* **1,** 4 (2001).
21. Graur, D. & Martin, W. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends Genet.* **20,** 80–86 (2004).
22. Hoyle, F. History of Earth. *Q. J. R. Astron. Soc.* **13,** 328–345 (1972).
23. Jaffres, J. B. D., Shields, G. A. & Wallmann, K. The oxygen isotope evolution of seawater: a critical review of a long-standing controversy and an improved geological water cycle model for the past 3.4 billion years. *Earth Sci. Rev.* **83,** 83–122 (2007).
24. Kasting, J. F. *et al.* Paleoclimates, ocean depth, and the oxygen isotopic composition of seawater. *Earth Planet. Sci. Lett.* **252,** 82–93 (2006).
25. Ward, D. M., Ferris, M. J., Nold, S. C. & Bateson, M. M. A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. *Microbiol. Mol. Biol. Rev.* **62,** 1353–1370 (1998).
26. Altekar, G., Dwarkadas, S., Huelsenbeck, J. P. & Ronquist, F. Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* **20,** 407–415 (2004).
27. Yang, Z. H. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13,** 555–556 (1997).
28. Dillon, P. J. & Rosen, C. A. A rapid method for the construction of synthetic genes using the polymerase chain reaction. *Biotechniques* **9,** 298–300 (1990).
29. Villalobos, A., Ness, J. E., Gustafsson, C., Minshull, J. & Govindarajan, S. Gene Designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics* **7,** 285 (2006).
30. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Knoll for his comments on this research, and D. Brooks, F. Battistuzzi, R. Davis, K. Josephson, S. Sassi, R. Shaw, J. Szostak and S. Benner for their assistance. E.A.G. acknowledges support from the NASA Exobiology program. O.G. was supported by a NIH/NCRR grant to A. S. Edison and a National Science Foundation grant to S. J. Hagan. S.G. was supported financially by DNA2.0.

Author Contributions E.A.G. designed the study, performed the evolutionary analyses and circular dichroism experiments, analysed the results and wrote the manuscript. S.G. performed gene synthesis. O.G. performed circular dichroism experiments, fitted the data and analysed the results. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.A.G. (egaucher@ffame.org).

METHODS

Phylogenetic analysis. The NCBI genome BLAST server was used to identify 160 phylogenetically diverse bacterial EF homologues that consisted of 394 unambiguously aligned amino-acid positions. The phylogenetic relationships between major bacterial lineages (phyla; and classes for proteobacteria) followed those of previous multigene studies[16,17]. All other relationships, branching patterns and phylogenetic parameters were determined with the parallel MPI version of MrBayes version 3.1.2 (ref. 26). This analysis used the following: Jones–Taylor–Thornton amino-acid replacement matrix, invariant-plus-gamma rates with eight categories, 4,000,000 generations with sampling at 100 intervals, and two runs with four chains per run. The first 1,000 trees were discarded as burn-in. Optimal chain-swapping occurred at a temperature of 0.02.

Ancestral reconstructions. Ancestral amino-acid character states were reconstructed computationally with PAML version 3.14 (ref. 27). Reconstructions of the most probabilistic ancestral sequences were inferred as described previously[8], with some exceptions. Notably, the present study also estimated the ancestral amino-acid frequencies derived from a set of 31 inferred proteins (from 8 mesophiles and 8 thermophiles) at a node representing the last common ancestor of bacteria as calculated from a multigene data set by using the EMAPI software[13]. These ancestral frequencies in turn served as equilibrium state frequencies in the amino-acid replacement matrix used to infer ancestral EF sequences. The present study also generated weighted random samplings derived from the posterior probability distribution for an ancestral EF node from the phylogeny from ref. 17. The distribution of posterior probabilities across all sites for all random sequences was equal to the posterior distribution calculated by PAML for the particular node of interest.

Gene synthesis, protein expression and purification. Genes were codon-optimized for expression in *E. coli* by using Gene Designer[29]. Gene synthesis was conducted as described previously[28], and genes were cloned into pET-21a vector (Novagen). *E. coli* Tuner (DE3) (Novagen) cells harbouring ancestral genes were grown for 16 h in 50 ml of autoinduction medium (1 mM of each soytone, yeast extract and 16 × P-salts, 20 mM succinate, 2 mM $MgSO_4$, 10 μM $FeCl_3$, 0.5% glycerol, 0.05% glucose, 0.02% lactose) with 100 μg ml^{-1} carbenicillin. Cells were harvested and disrupted by using BugBuster with Benzonase (Novagen) in accordance with the manufacturer's protocol, and were supplemented with 1 mM phenylmethylsulphonyl fluoride (PMSF) and 5 mM imidazole. Ancestral proteins were isolated with Ni^{2+} columns (Qiagen) in accordance with the manufacturer's protocol with the following exception: wash and elution buffers contained 80 and 500 mM imidazole, respectively. Purified proteins were dialysed twice against a 1,000× buffer volume (20 mM Tris, 100 mM KCl, 10 mM $MgCl_2$ and 5 μM guanosine diphosphate pH 7.5) and stored at 4 °C. Removal of imidazole was crucial to prevent interference during spectrophotometer analysis.

Circular dichroism. Thermostability of modern and ancestral EF proteins were characterized by using far-ultraviolet circular dichroism (CD) on an Aviv model 400 spectrometer. Protein samples were loaded into 1-mm pathlength cells for all experiments. Temperature scans at 222 nm were typically taken from 25 °C to 95 °C in increments of 2 °C. Temperature was increased 0.5 °C min^{-1} and samples were allowed to equilibrate for 1 min at each temperature step. Wavelength scans were performed before and after all temperature scans to assess changes in secondary structure. Wavelength scans were taken at 25 °C, from 250 nm to 190 nm in increments of 1 nm. All data points were collected and averaged over 5 s, at a rate of 10,000 data points per second. CD analyses were performed in buffer void of stabilizers to ensure accurate melting profiles. Stabilizers, such as glycerol, can increase thermostability beyond physiologically permissible temperatures. For instance, CD analysis of *E. coli* EFTu in 33% glycerol increases protein thermostability (T_m = 52.2 °C with glycerol compared with T_m = 42.8 °C without glycerol; data not shown).

Data analysis. A single-exponential function was used to fit the data, assuming a structured/unstructured transition as the EF proteins were heated. The relationship between temperature and denaturation at 222 nm seems sigmoidal, allowing a sigmoid curve to be fitted to the CD data (see Supplementary Information):

$$a + b(1 + e^{-(x-c)/d})^{-1}$$

The four parameters *a*, *b*, *c* and *d* define the curve, with *c* being the midpoint of the transition. CD data obtained for each protein were fitted by the curve to obtain the melting profiles. A derivative of the sigmoid fit was used to represent a resolution function, which was scaled and applied to each curve. This permitted the calculation of full-width at half maximum, standard deviation and confidence intervals. T_m values of ancestral and modern EFs are defined here as the lower temperature bound of the 95% confidence interval derived from the sigmoid curve and were unaffected by errors in CD measurement. These temperatures correspond to the initiation of structural perturbations of EF proteins and are correlated with the optimal growth temperatures of their host organisms[15].