Title: Phenotypic reconstruction of the last universal common ancestor reveals a complex cell.

Fouad El Baidouri[1,2,*], Chris Venditti[3], Sei Suzuki[1,4], Andrew Meade[3] & Stuart Humphries[1,*]

Author Affiliation:

[1] School of Life Sciences University of Lincoln, Joseph Banks Laboratories, Green Lane, Lincoln LN6 7DL, UK.

[2] Current address: Department of Botany, University of Hawaii at Manoa, Honolulu, HI 96822, USA.

[3] School of Biological Sciences, University of Reading, Reading RG6 6BX, UK.

[4] Current address: Centre for Ocean Life, Technical University of Denmark, Kemitorvet, 2800 Kgs. Lyngby, Denmark.

Author ORCID ID:

Fouad El Baidouri: 0000-0001-5204-6244

Chris Venditti: 0000-0002-6776-2355

Andrew Meade: 0000-0001-7095-7711

Stuart Humphries: 0000-0001-9766-6404

[*] Corresponding Authors:

Fouad El Baidouri: Department of Botany, University of Hawaii at Manoa, Honolulu, HI 96822, USA. Email: baidouri@hawaii.edu.

Stuart Humphries: School of Life Sciences University of Lincoln, Joseph Banks Laboratories, Green Lane, Lincoln LN6 7DL, UK. Email: shumphries@lincoln.ac.uk. Phone: +44 (0)1522 886018.

Abstract

A fundamental concept in evolutionary theory is the last universal common ancestor (LUCA) from which all living organisms. While some authors have suggested a relatively complex LUCA [1] it is still widely assumed that LUCA must have been a very simple cell and that life has subsequently increased in complexity through time [2,3]. However, while current thought does tend towards a general increase in complexity through time in Eukaryotes [4,5], there is increasing evidence that bacteria and archaea have undergone considerable genome reduction during their evolution [6,7]. This raises the surprising possibility that LUCA, as the ancestor of bacteria and archaea may have been a considerably complex cell. While hypotheses regarding the phenotype of LUCA do exist, all are founded on gene presence/absence [1–3]. Yet, despite recent attempts to link genes and phenotypic traits in prokaryotes [8,9], it is still inherently difficult to predict phenotype based on the presence or absence of genes alone [8]. In response to this, we used Bayesian phylogenetic comparative methods [10,11] to predict ancestral traits. Testing for robustness to horizontal gene transfer (HGT) we inferred the phenotypic traits of LUCA using two robust published phylogenetic trees [12,13] and a dataset of 3,128 bacterial and archaeal species (Supplementary Information). Our results depict LUCA as a far more complex cell than has previously been proposed, challenging the evolutionary model of increased complexity through time in prokaryotes. Given current estimates for the emergence of LUCA we suggest that early life very rapidly evolved considerable cellular complexity.

Results and Discussion

Using phylogenetic comparative methods [16,17] and the largest compilation to-date of bacterial and archaeal phenotypic data (Fig. 1), along with two robust prokaryotic phylogenetic trees [12,13] we show that the last universal common ancestor of all living organisms was a complex cell with at least 22 reconstructed phenotypic traits just as intricate as those of many modern bacteria and archaea (Fig. 2 and Table 1, and see SI for estimates of uncertainty). It is now largely accepted that LUCA is the common ancestor of bacteria and archaea[14,15] with eukaryotes originating from within the archaea in the two domains tree of life[18]. Given current estimates for the time of Earth's formation and the presence of liquid water offering habitable conditions [19] our results suggest that early life must have very quickly (<500 Myr) evolved considerable cellular complexity.

Our probabilistic Bayesian phylogenetic models show clear phylogenetic signal in all 22 phenotypic traits we investigate (Table 1, Table S3 and Table S6). Where this is easily quantified (five continuous traits), there is a minimum value of λ of 0.95 across the two trees (Tables 1 and Table S3). This signal indicates vertical inheritance of complex microbial traits and support previous findings that trait (as opposed to gene) conservatism is widespread in bacteria and archaea [20,21]. Genetically complex traits, such as motility, shape, and the cell envelope are controlled by multiple genes [22–25] and while unlikely to be HGT free, the extent of this mechanism in the traits we study is not strong enough to blur their phylogenetic signal (see below and Methods).

We reveal LUCA as actively motile with a cell wall and a single cell membrane (monodermic cell plan, Figs. 2 & 3). Our reconstruction of LUCA with a cell wall is supported by the recent suggestion that it had genes associated with cell wall synthesis [2]. While the presence and nature of a cell membrane in LUCA has been debated [26–28], recent evidence points towards a mixed membrane of archaeal- and bacterial-like lipids [29,30]. We are confident that LUCA was halotolerant with a preferred salinity range below that of modern seawater, hyperthermophilic (existing at temperatures above 70 °C) and living freely in an aquatic environment of neutral pH (Fig. 2, Tables 1, S1-S3). However, while we provide support for hyperthermophilia there is still debate over the presence or absence of reverse gyrase (a hyperthermophile-specific enzyme) in LUCA [2,31]. A chemolithotroph, using inorganic molecules to drive its metabolism, LUCA was adapted to anaerobic conditions.

Many of these reconstructed traits are indicative of the environment around hydrothermal vents, a habitat suggested for both the origins of life and the emergence of LUCA [2,32]. While modern vent systems are not generally of neutral pH [33], recent estimates for the early ocean suggest slightly acidic to neutral pH conditions [34] at the time of LUCA [35,36] (~ 4.0 Ga) with values similar to our estimates (Table 1). Although an anaerobe, our reconstructions show that LUCA possessed both catalase and oxidase enzymes, suggesting an ability to resist oxidative stress. The presence of these enzymes might seem surprising given current evidence for free oxygen on early earth at concentrations less than 0.001% of those today [37], but is supported by genomic studies of Cytochrome oxidase genes[38], but see Dibrova at al. [39]. It has been hypothesized that LUCA might have possessed pathways to remove reactive oxygen species [40] (ROS) which is supported by our results, in particular, the presence of catalase [41]. While genes involved in resistance to oxidative stress might have arisen and been horizontally transferred between many ancestors after LUCA, recently Weiss et al. [2] identified eight of their 355 inferred LUCA proteins as enzymes involved in oxygen-dependent reactions.

Although bacteria and archaea exhibit a variety of morphologies, shape is usually well conserved within species [42]. Based on categorical data ('rod-like', 'ovoid' or 'coccoid') we provide evidence that LUCA was an ovoid cell (Fig. 2, Tables 1 & S1), suggesting that it had a complex cell cycle including an elongation phase (*sensu* 22). Our estimates of cell dimensions also support an elongated cell with a length of 1.72 μm and width of 0.51 μm (Table 1). We were unable to draw a clear picture of pleomorphism in LUCA, but aggregation of individuals into clusters was not frequent (Fig. 2, Tables S1 & S2). The ability to form spores, while hypothesized as ancestral [44], was unclear, likely owing to the uncertainty in the deep phylogenetic branching order of the two trees [12,13] (Tables S1 & S2, Fig. S3-S4). While some evidence exists for HGT of sporulation-related genes within the firmicutes [45], these transfers occurred locally and did not remove the global phylogenetic signal we recover from our data (Table S6).

The complex phenotypic picture we draw of LUCA implies that it must also have possessed a complex genome. We estimated LUCA's ancestral genome size at 2.49 Mbp (Table 1), larger than the phylogenetically corrected average genome size in modern prokaryotes from our dataset (~1.94 Mbp). Models allowing both gene losses (genome reduction) and gains (HGT) suggest that high levels of HGT lead to ancestral genome size estimates far smaller than modern genomes, while with limited HGT ancestral genome sizes become unrealistically large [46]. Unlike eukaryotes, prokaryotes show a consistent relationship between genome size and gene number [47]. Our estimate of the number of genes (2629) in LUCA (Table 1), assuming an average gene length for prokaryotes of 1024 bp and 13% of non-coding DNA in prokaryotic genomes [48], implies a genome size of around 2.34 Mbp, compared to our independent reconstruction of 2.49 Mbp (Methods).

A point we wish to highlight is that our approach differs fundamentally from gene-based studies in that the inherent complexity of the traits we examine acts to isolate them from the effects of HGT. The involvement of multiple genes and gene-clusters in determining the phenotypic values of many traits reduces the likelihood that they have been subject to HGT [49–51] and the close correspondence between the data and the trees (indicated by high λ, Table 1) gives us confidence that comparative phylogenetic methods can be used to understand the evolution of prokaryotic phenotypes and accurately estimate ancestral states (as we have previously shown [17]).

While HGT could have disrupted any phylogenetic signal in the trees we use, the majority of taxonomic groups are seen to be supported by both trees. Deep branching in the two trees does differ, but our approach of accounting for phylogenetic uncertainty (Supplementary Information) gives us confidence that these differences are unlikely to be due to HGT.

Irrespective of any network structure due to HGT, we recover a strong phylogenetic signal in all traits from both trees, implying that HGT does not rapidly act to change whole phenotypes or that it occurs mostly between very closely related lineages. While a positive correlation between the frequency of HGTs in distant lineages and genome size has been suggested [52], several studies indicate that HGT in prokaryotes also depends on both physical and genetic proximity [53–55]. Moreover, there is accumulating evidence to suggest that some of the traits that we present such as motility, cell envelope and sporulation, were subject to little or rare HGT events during their evolutionary history in a number of prokaryotic lineages [22,23,25,45,55].

As a further check we also performed simulations of HGT for each trait (See Methods, Fig. S2, SI) showing that our method is robust to HGT biases when estimating ancestral states and we were able to reconstruct ancestral nodes even in the presence of unrealistically high [56] rates of HGT. Thus, the strong phylogenetic signal that we and others [20,21,51,57] find for prokaryote phenotypic traits demonstrates that the underlying phylogenetic relationships, even into deep time, are often conserved and that HGT does not always act to erase the phylogenetic signal for many complex traits.
Based on all of the above, we conclude that LUCA, the cenancestor, was far more than a "half-alive" progenote [58,59], as suggested by Weiss et al. [2], and show that it was a complex "prokaryotic" cell resembling modern archaea and bacteria. The complex phenotypic picture we depict of LUCA implies a complex genome, which is supported by our estimates of genome size and gene numbers. These results challenge the common assumption of increasing complexity through time, suggesting instead that cellular complexity arose near the very beginning of life and was retained or even lost through the evolution of the prokaryote lineage.

LUCA itself gave rise to two diverse and important domains of life, bacteria and archaea, whose ancestral phenotypic characteristics we also reconstructed. Our results suggest that both the last bacterial common ancestor (LBCA) and the last common ancestor of archaea (LACA) were similar to LUCA, in being free-living, motile organisms with a cell wall (Fig. 2). We were unable to draw a clear picture of pleomorphism or determine the pH conditions under which LBCA and LACA lived (Tables 1 & S1) but like LUCA neither were able to form aggregates (Fig. 2). While LACA was non-spore forming it is unclear whether LBCA was able to do so, likely owing to the uncertainty in deep phylogenetic branching of bacterial phyla between the two trees, (Tables S1-S2). As for LUCA, while there is some evidence for HGT of sporulation-related genes within the firmicutes, these transfers occurred locally and did not remove the global phylogenetic signal we recover from our data (Table S6). Like LUCA, these progenitors of the bacterial and archaeal domains were chemolithotrophic, halotolerant anaerobes with an ability to resist oxidative stress as suggested by the presence of catalase and oxidase (Fig. 2).

In agreement with previous work [60–62] our results suggest that LBCA was a rod-like cell, implying that it had a complex life cycle with an elongation phase (Figs. 2b & 3, Table 1). Our results support the suggestion that LBCA was gram positive [61], with a monodermic cell plan (Fig. 2, Table S1). In contrast to LBCA, LACA was a slightly elongated ovoid cell (Figs. 2c & 3, Tables 1, S1), living in extremely hot aquatic environments with optimal temperatures between 73 and 74 °C (Fig. 2, Table 1). Very similar temperatures (~73-76 °C) have been recently estimated for LACA based on a different approach [63]. As with LUCA, the complex phenotypic characters of both LBCA and LACA suggest that they too had large genomes, supported (as before) by our estimates of genome size and gene numbers (Tables 1 & S3). However, Williams et al. [63] estimated a genome half the size (~1090 to 1328 genes) of our estimates for LACA, with a dataset of 62 archaeal genomes compared to 205 archaeal species in our study.

Conclusion

While LACA, a descendant of LUCA, has been suggested to be a relatively complex organism based on gene number [63,64], our results push cellular complexity back to the very beginnings of life. Barring panspermia [65], these results imply that complex phenotypic traits arose far earlier in the history of life than previously thought. Accumulating lines of evidence suggest that LUCA appeared very quickly during Earth's history, perhaps as early as four billion years ago [35,36] during the Hadean or early Archaean eons. Given that Earth formed around 4.5 Billion years ago [66], life probably arose in less than 500 million years (perhaps within 200 million years [67]). This indicates that early life must have very quickly evolved considerable cellular complexity. We thus reveal LUCA as a complex cell possessing a genetic code more intricate than many modern bacteria and archaea.

Materials and Methods

Experimental Design

1. Data collection

1.1. Phenotypic data

To reconstruct the ancestral states for LUCA, we collected all the relevant phenotypic data in the species description section from the Bergey's Manual of Systematic Bacteriology that were consistently reported for many phyla and genera. The phenotypic traits (Data S1) that were consistently reported for all the species included shape, cell size, pleomorphism, cell aggregation, motility, habitat, cell wall and spore formation, as well as catalase and oxidase activity, oxygen, temperature, pH and NaCl requirements and nutritional mode for 3,128 prokaryotic species from the phylogenies of Chai et al. [12] (large tree) and Segata et al. [13] (small tree). These two trees were chosen to account for phylogenetic uncertainties as they were reconstructed from two different data sets and have different deep branching orders and different positions of the root. The large tree and the phylogenetic distribution of phenotypic data on it are shown in Fig. 1. Data for most species was collated from the five volumes of Bergey's Manual of Systematic Bacteriology [68–72], but for species described after their publication data it was collected from the primary literature based on the List of Prokaryotic Names with Standing in Nomenclature (LPSN, Supplementary Text) [73]. Data for outlier strains (potentially misclassified species), were not included in the analysis (for example, *Clostridium difficile* strain P28 did not cluster with members of the same species). Bacterial and archaeal that have been delineated based on genomic data but have not been described phenotypically (not cultured, e.g. bacterial candidate phyla radiation (CPR) or Asgard archaea) were not included in the present study as no phenotypic data was available.

1.2. Genomic data

Genome size and gene number data were collected for completely sequenced bacterial and archaeal genomes from the GOLD database [74]. Average genome size and average gene number per species were calculated from all the available strains of a given species. To independently estimate LUCA's genome size we multiplied our estimated number of genes for LUCA (2629) by the estimated average gene length for prokaryotes from our data (1024 bp). Average gene length was estimated by regressing the genome size (Mbp) on the number of genes using Phylogenetic generalized least-squares (PGLS) implemented in the CAPER package to account for species non-independence due to shared ancestry [75]. Estimated genome size (gene number × gene length = 2.69 Mbp) was multiplied by the average ratio of coding versus non-coding sequences in prokaryotic genomes (0.87) based on Xu [48] to correct for non-coding sequences in the genome. We note that our estimates do not include CPR, known for their small genome size [76], as no phenotypically described species existed for this group at the time of data collection. However, the position of the CPR in the bacteria tree of life) [77] suggests that they mostly symbiotic and are derived (in particular in reference to their small genome) and therefore unlikely to influence the results of our reconstructions.

2. Phenotypic characterization

2.1. Cell morphology

Shape characterization in the species description sections from Bergey's Manual of Systematic Bacteriology and the primary literature is qualitative, often subjective, and is not geometrically precise. A more reliable description would ideally be based on size measurements from individual cells. However, generally only cell length and width (diameter for coccoid cells) are available in the literature. These dimensions do however allow calculation of cell aspect ratio (AR) as length divided by width (Supplementary Text). While AR can be used to distinguish between coccoid and non-coccoid species, it cannot be used to distinguish categories within elongated, cylindrical cells (e.g. rod versus ovoid). Based on qualitative descriptions of cell morphology, we defined three broad shape categories for individual cells – rod, ovoid and coccoid (see Supplementary Text for a detailed description). For all the species with one shape category for which AR was available (2005 species) the delineation between coccoid and non-coccoid species based on qualitative descriptions was largely congruent with the AR data (Fig. S1, Supplementary Information).

2.2. Pleomorphism

Pleomorphism describes the ability of some prokaryotes to alter their shape or size in response to environmental conditions [78]. To investigate the extent and evolution of pleomorphism we classified bacteria and archaea as either monomorphic or non-monomorphic.

2.3. Motility

Motility was defined for the purpose of this study as the ability to move. Species were classified as motile or non-motile regardless of motility type (i.e. swarming, swimming, gliding or twitching motility).

2.4. Cell aggregation

We classified species as either aggregating or non-aggregating, with cell aggregation defined as the ability to form an association between two or more cells. Species with cells occurring in pairs, tetrads, chains, sarcina, clusters or V-form, Y-form, T-form and Palisade groups or species able to form multicellular structures (e.g. fruiting bodies) were considered as aggregating. Species occurring only as single cells were coded as non-aggregating, while those occurring as single cells but with the ability to associate were coded as aggregating.

2.5. Habitat

Due to limited availability of information on microenvironments we used three broad categorizations [17] of habitat types based on macro-environmental descriptions (i.e. the different locations where the organism naturally lives and grows and from which it could be recovered and isolated). Where habitat was not known, the first isolation site was used (e.g. human tissue, soil, shallow water, etc.). The first categorization set (habitat a) was a simple division into free-living (i.e. living independently in the environment) and non-free-living species (i.e. those associated with a host, manufactured products or isolated from industrial environments). Both free and non-free living species were further classified as Terrestrial or Aquatic (habitat b). For instance, species living in soil, fields or isolated from terrestrial sediments, were coded as Terrestrial, while an aquatic habitat was defined by living in marine or freshwater environments. Finally, we used a third habitat classification (habitat c) that included three categories: Terrestrial, Freshwater or Marine (see Supplementary Text for further details).

2.6. Cell envelope

To test whether ancestral forms had a cell wall we classified species based on cell envelope type. Species with a description of a cell wall ultrastructure and/or described as Gram positive or Gram negative were all coded as having a cell wall. However, wall-less species (e.g. *Mycoplasma bovis*, see Supplementary Text) were considered as lacking a cell wall regardless of Gram stain. To investigate the ancestral state of cell plan organization, we further classified bacteria and archaea based on two cell types: monoderm (single membrane) and diderm (inner and outer membranes) based on information on cell envelope type and Gram reaction. Under this classification scheme species described as Gram positive or Gram negative were coded as having monodermic and didermic cell plans respectively, while those described as having a variable Gram reaction or staining negative but having a positive cell wall structure were coded as monoderms. Wall-less species were coded as monoderm as they possess a cell membrane. Species displaying a variable Gram reaction and without information on cell wall structure were not included in the analysis (see Supplementary Text for further details).

2.7. Spore formation

We classified bacterial species as spore-forming or non-spore-forming based on clear statements in the species description and/or with a description of the spore type (e.g. "coccoid spores are formed centrally"). Archaea were classified as non-spore forming, and those species without information on spore-formation were not included in the analysis (see Supplementary Text for further details).

2.8. Catalase, oxidase and oxygen requirements

We classified bacteria and archaea based on their ability to produce catalase and oxidase enzymes. Species described as having variable catalase or oxidase activities (i.e. positive or negative catalase or oxidase tests, 14 species in total) and those for which catalase activity was described as 'weakly positive' were excluded from the analysis. Species described as not usually producing an enzyme were recorded as not producing it rather than excluded from the analysis.

Species were recorded according to their ability to use oxygen for metabolic processes as aerobic, anaerobic, or both. Facultative aerobes or anaerobes were coded as both aerobic and anaerobic. Microaerophylic species (requiring oxygen to metabolize energy sources but not able to tolerate high concentrations) and microaerotolerant (tolerant of oxygen but unable to use it for growth) were not considered as a distinct category and were excluded from the analyses (42 species).

2.9. Metabolism

The source of energy and electron donor that bacterial and archaeal species exploit was used to classify their metabolism. Data were extracted from species descriptions in the form of, for example, 'chemoorganotroph', 'chemolithotroph' or 'phototroph'. Use of the photo- or chemo- prefix classified species phototrophs or chemotrophs respectively. For analysis of ancestral electron donor sources species described with words such as litho- or organo- were classified as lithotrophs or organotrophs respectively.

3. Physicochemical parameters

For all species data on optimal temperature, pH and NaCl ranges were collected (lower and upper values) and used to estimate ancestral values based on a variable rate model (see Continuous traits). To assess whether these estimates were accurate we used a different approach based on categorical data (see Categorical traits) employing four key categories for temperature, pH and NaCl (See Dataset S1) defined using optimal ranges. Species growing optimally in a temperature range of 5-20 ºC were classified as psychrotolerant, and those at temperatures between 20 to 45 ºC as mesophilic. Thermophilic species were defined as growing optimally at temperatures between 45 to 70 ºC, while hyperthermophiles were those with optimal growth temperatures above 70 ºC (see Supplementary Text). Optimal pH ranges were used to define neutrophiles (pH 5.5-7.5), acidophiles (pH 3-5.5), hyperacidophiles (pH 0-3) and alkaliphiles (pH 7.5-13). Optimal NaCl range was used to define non-halophiles (0-1% w/v), halotolerant (1-5% w/v), halophiles (5-20% w/v) and extreme-halophiles (20-32% w/v). Results from both approaches were compared to assess whether optimal lower and upper estimated ancestral values for temperature, pH and NaCl fell within the corresponding ancestrally reconstructed category (e.g. if the estimated lower and upper optimal temperature values were 72 ºC and 73 ºC, the ancestral reconstructed temperature category was hyperthermophile).

Statistical Analysis

1. Ancestral state reconstruction

1.2. Categorical traits

To allow us to account for model and parameter uncertainty a Bayesian rjMCMC [10] approach was used to reconstruct LUCAs ancestral states for all categorical traits (Table S1.) using the Multistate method in the multicore version of BayesTraits V3.0.1 [11].[79] For the categorical data, species exhibiting more than one state of the same character were coded as having two states or more in BayesTraits (e.g. if a species exhibited both rod 'R' and coccoid 'C' forms it was coded as RC, see Cell morphology and Dataset S1). For all the categorical data species with missing information (NAs, Dataset S1) were excluded from the analyses.

The 'AddMRCA' command was used to estimate the ancestral states of the LBCA and LACA. All rjMCMC analyses were run using both the large [12] and small [13] Maximum Likelihood trees, reconstructed using 400 conserved proteins among 14,727 and 3,737 prokaryotic genomes respectively. All rjMCMC analyses were run in BayesTraits using the same parameters with all priors set to an exponential with a mean of 10. rjMCMC chains were run for 5,000,000 iterations with the first 10% as burn-in and sampling every 1,000 iterations after convergence. The exception was the analysis for pH, which was run for 20 million iterations with the first two million iterations as burn-in and a sampling interval of 2,000 to ensure chain convergence. MCMC chains were run three times per tree and chain mixing and convergence were assessed using Tracer v1.7 [80]. For all the analyses we report the mean and median posterior probability for the root (LUCA), LBCA and LACA, as well as the lower and upper HPD and 95% PI for the probability of the character state.

1.2. *Continuous traits*

Ancestral values for cell size, genome size, and gene number, as well as optimal pH, NaCl and temperature were estimated for LUCA, LBCA and LACA using a variable rates model [81], implemented in BayesTraits, that detects heterogeneous rates of evolutionary changes along the tree within a Bayesian framework. This method can be used to detect evolutionary trends and has been shown to estimate ancestral states more accurately than other methods [82].

Following the methodology in Baker et al. [82] we ran the variable-rates model for each continuous trait (the MCMC chain was run for 1 billion iterations, sampling every 500,000 iterations after a burn-in period of 500 million iterations). This approach detects branches or clades of the tree that have undergone especially fast or slow rates of change, it stretches (increased rate) or compresses (decreased rate) branch lengths by an amount reflecting the inferred rate of evolution in that branch.

We summed all the rate-scaled branches along the phylogenetic path of each species from the root of the tree to the terminal branch (root-to-tip rate) – we use the median rate from the posterior distribution to rate-scale the branches. To test for a trend, we regressed the value of the continuous trait under investigation with root-to-tip rate using a Bayesian MCMC phylogenetic generalized least squares (GLS) regression [75,83] (the MCMC chain was run for 1 billion iterations with a burn-in of 500 million iterations and a sampling interval of 500,000 iterations). We used the proportion of time (Px) the regression coefficient determined the crossed zero to determine significance where Px < 0.05 we declared a significant trend.

We estimate the ancestral values for LUCA, LACA and LBCA for all continuous traits using the phylogenetic prediction method originally described in Organ et al. [84], where we find a significant trend we account for that as in Baker et al. [82]. Each of our MCMC analyses was repeated three times to ensure convergence was achieved – we report the results from one randomly chosen chain for each analysis.

As recommended in Venditti et al. [81], we randomly removed all except one species in a clade where the continuous data were identical. We do this, as from a statistical modelling point of view, the best fit to such a group of species with exactly the same size would be to infer that no evolutionary change occurred in that group. This is an unrealistic scenario. Using this reduced set of species, we also repeated our Categorical analyses for pH, temperature and NaCl, finding concordant results (Tables S4-S5). Continuous data were log10 transformed prior to analysis, except NaCl which were cube-root transformed owing to the presence of zeros, and pH where the original values were used as the scale is logarithmic.

In contrast to rod or ovoid shape species, coccoids are evolutionarily constrained and thus can only divide and not elongate [43]. We therefore excluded data from coccoid species with an Aspect Ratio (AR) = 1 (see Cell morphology and Supplementary Text) prior to estimation of ancestral cell dimensions for LUCA, LACA and LBCA.

Phylogenetic signal

To evaluate the extent of the phylogenetic signal in our data we coded categorical traits into binary data and reconstructed the ancestral states for each trait using Maximum likelihood (Table S6). We then compared the posterior probabilities with a randomized data set from each of the two trees. We concluded that there is phylogenetic signal if we were able to discriminate between the states of a given trait (e.g. P(0)=0.99 and P(1)=0.01) in our data while simultaneously obtaining equal posterior probabilities for these states (i.e. 50/50) of the same trait from the randomized data set (Table S6).
For continuous traits the phylogenetic signal was estimated using Maximum likelihood and the variable rates model and λ values were selected from the best of two competing models - one with λ set to zero and one for which λ was estimated.

2. Horizontal trait transfer simulations

While HGT plays a role in prokaryotic evolution, often seen as obscuring the assumption of vertical descent with modification, the traits under investigation in our study are multi-gene entities and as such our ancestral reconstructions are less likely to be influenced by HGT [49–51]. Furthermore, a number of studies show that HGT occurs predominantly among near phylogenetic relatives and closely interacting species [85–87] and is far less common between distantly related organisms. To test the accuracy of our reconstruction methods to cases of horizontal trait transfer (HTT) we conduct two sets of simulations. We call our simulation process HTT rather than HGT as we simulate the extreme case where the trait itself is transferred between phylogenetic branches – this represents a suite of functional genes associated with a particular trait transferred and instantly fixed in the population. HTT is thus a much more severe test of HGT than a gradual replacement model as it implies the transfer of an entire operon or the entire set of genes coding for the trait.

Using the Chai et al. [12] phylogenetic tree we randomly identified between 0 and 1000 (in 200 increments, see Fig. S2) points along the branches. We simulated a trait via Brownian motion (variance = 1) along the branches. When the simulation reached one of the randomly chosen points, an HTT event occurred from the ‘donor’ (red points in Fig. S2 a and b) to one of the potential ‘recipients’ (yellow squares in Fig. S2 a and b). We ran simulations for two types of HTT events. Firstly, a *Local* version, where HTT occurred between near phylogenetic relatives (Fig. S2a) – here, at the HTT point (and example is provided as a red point in Fig. S2) we selected the recipient branch from the clade defined by the most recent common ancestor (i.e. one phylogenetic node back) of the HTT event (in our example the potential recipient branches are shown with a yellow squares). Where there is more than one recipient branch (as in Fig. S2a) we randomly select one to receive the horizontally transferred trait. At the point of transfer the recipient branch and the donor have the same trait value – following this the Brownian motion process continues. We also simulate a very severe *Global* process which is identical to the *Local* simulation apart from the fact that at the HTT point the potential recipient points can be chosen from all branches at that time point (i.e. from the root of the tree rather than the most recent common ancestor), this is illustrated in Fig. S2b. Through the simulation process we record the trait value at each node of the tree – these are the ‘known’ ancestral states. Following the simulation process, we used the data at the terminal branches (the end point of the simulation at tips of the tree) to infer the ancestral states at each node in the phylogeny. We compare these with the known ancestral states for each simulation condition. We repeat the process 1000 times for each condition for both the global and the local set – a total of 12000 simulations.

We use linear regression to compare the known to inferred ancestral state values. We expect, where the reconstructions are perfectly accurate, to see a slope (regression coefficient) equal to one, an intercept equal to zero, and a coefficient of determination near one. We also recorded the estimated Brownian motion variance from the inference stage – the simulated variance was one. The results show that *Local* HTT does not significantly influence our ability to accurately infer ancestral states even up to a point where there are 1000 such events in the tree (Fig. S2c-f). As expected, the accuracy of the unrealistically severe *Global* simulations reduces with the number of HTT events – specifically the variance in the intercept increases and the variance and magnitude of the Brownian motion variance increases.

References.

1. Ouzounis, C. A., Kunin, V., Darzentas, N. & Goldovsky, L. A minimal estimate for the gene content of the last universal common ancestor - Exobiology from a terrestrial perspective. *Res. Microbiol.* **157**, 57–68 (2006).

2. Weiss, M. C. *et al.* The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **1**, 16116 (2016).

3. Koonin, E. V. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* **1**, 127–136 (2003).

4. Hedges, S. B., Blair, J. E., Venturi, M. L. & Shoe, J. L. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol. Biol.* **4**, 2 (2004).

5. Carroll, S. B. Chance and necessity: The evolution of morphological complexity and diversity. *Nature* **409**, 1102–1109 (2001).

6. Wolf, Y. I. & Koonin, E. V. Genome reduction as the dominant mode of evolution. *BioEssays* **35**, 829–837 (2013).

7. Giovannoni, S. J., Cameron Thrash, J. & Temperton, B. Implications of streamlining theory for microbial ecology. *ISME Journal* **8**, 1553–1565 (2014).

8. Brbić, M. *et al.* The landscape of microbial phenotypic traits and associated genes. *Nucleic Acids Res.* **44**, 10074–10090 (2016).

9. Edgar, R. C. SINAPS: Prediction of microbial traits from marker gene sequences. *bioRxiv* (2017). doi:doi.org/10.1101/124156

10. Pagel, M., Meade, A. & Barker, D. Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* **53**, 673–684 (2004).

11. Pagel, M. & Meade, A. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am. Nat.* **167**, 808–825 (2006).

12. Chai, J., Kora, G., Ahn, T. H., Hyatt, D. & Pan, C. Functional phylogenomics analysis of bacteria and archaea using consistent genome annotation with UniFam. *BMC Evol. Biol.* **14**, 1–13 (2014).

13. Segata, N., Börnigen, D., Morgan, X. C. & Huttenhower, C. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* **4**, 2304 (2013).

14. Williams, T. A., Foster, P. G., Cox, C. J. & Embley, T. M. An archaeal origin of

eukaryotes supports only two primary domains of life. *Nature* **504**, 231–236 (2013).

15. Spang, A. *et al.* Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179 (2015).

16. Cornwell, W. & Nakagawa, S. Phylogenetic comparative methods. *Current Biology* **27**, R333–R336 (2017).

17. El Baidouri, F., Venditti, C. & Humphries, S. Independent evolution of shape and motility allows evolutionary flexibility in Firmicutes bacteria. *Nat. Ecol. Evol.* **1**, 0009 (2016).

18. Williams, T. A., Cox, C. J., Foster, P. G., Szöllősi, G. J. & Embley, T. M. Phylogenomics provides robust support for a two-domains tree of life. *Nat. Ecol. Evol.* **4**, 138–147 (2020).

19. Camprubí, E. *et al.* The Emergence of Life. *Space Sci. Rev.* **215**, 56 (2019).

20. Martiny, A. C., Treseder, K. & Pusch, G. Phylogenetic conservatism of functional traits in microorganisms. *ISME J.* **7**, 830–838 (2013).

21. Goberna, M. & Verdú, M. Predicting microbial traits with phylogenies. *ISME J.* **10**, 959–967 (2016).

22. Liu, R. & Ochman, H. Stepwise formation of the bacterial flagellar system. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 7116–7121 (2007).

23. Desmond, E., Brochier-Armanet, C. & Gribaldo, S. Phylogenomics of the archaeal flagellum: rare horizontal gene transfer in a unique motility structure. *BMC Evol. Biol.* **7**, 106 (2007).

24. Jiang, C., Caccamo, P. D. & Brun, Y. V. Mechanisms of bacterial morphogenesis: Evolutionary cell biology approaches provide new insights. *BioEssays* **37**, 413–425 (2015).

25. Antunes, L. C. S. *et al.* Phylogenomic analysis supports the ancestral presence of LPS-outer membranes in the Firmicutes. *Elife* **5**, (2016).

26. Koga, Y., Kyuragi, T., Nishihara, M. & Sone, N. Did archaeal and bacterial cells arise independently from noncellular precursors? A hypothesis stating that the advent of membrane phospholipid with enantiomeric glycerophosphate backbones caused the separation of the two lines of descent. *J. Mol. Evol.* **46**, 54–63 (1998).

27. Koonin, E. V. & Martin, W. On the origin of genomes and cells within inorganic compartments. *Trends Genet.* **21**, 647–654 (2005).

28. Peretó, J., López-García, P. & Moreira, D. Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem. Sci.* **29**, 469–477 (2004).

29. Villanueva, L. *et al.* Bridging the divide: bacteria synthesizing archaeal membrane lipids. *bioRxiv* 448035 (2018). doi:https://doi.org/10.1101/448035

30. Caforio, A. *et al.* Converting Escherichia coli into an archaebacterium with a hybrid heterochiral membrane. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 3704–3709 (2018).

31. Catchpole, R. J. & Forterre, P. The Evolution of Reverse Gyrase Suggests a Nonhyperthermophilic Last Universal Common Ancestor. *Mol. Biol. Evol.* **36**, 2737–2747 (2019).

32. Baross, J. A. & Hoffman, S. E. Submarine hydrothermal vents and associated gradient environments as sites for the origin and evolution of life. *Orig. Life Evol. Biosph.* **15**, 327–345 (1985).

33. Ding, K. *et al.* The in situ pH of hydrothermal fluids at mid-ocean ridges. *Earth Planet. Sci. Lett.* **237**, 167–174 (2005).

34. Krissansen-Totton, J., Arney, G. N. & Catling, D. C. Constraining the climate and ocean pH of the early Earth with a geological carbon cycle model. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 4105–4110 (2018).

35. Tashiro, T. *et al.* Early trace of life from 3.95 Ga sedimentary rocks in Labrador, Canada. *Nature* **549**, 516–518 (2017).

36. Betts, H. C. *et al.* Integrated genomic and fossil evidence illuminates life’s early evolution and eukaryote origin. *Nat. Ecol. Evol.* **2**, 1556–1562 (2018).

37. Lyons, T. W., Reinhard, C. T. & Planavsky, N. J. The rise of oxygen in Earth’s early ocean and atmosphere. *Nature* **506**, 307–15 (2014).

38. Castresana, J., Lübben, M., Saraste, M. & Higgins, D. G. Evolution of cytochrome oxidase, an enzyme older than atmospheric oxygen. *EMBO J.* **13**, 2516–2525 (1994).

39. Dibrova, D. V., Shalaeva, D. N., Galperin, M. Y. & Mulkidjanian, A. Y. Emergence of cytochrome bc complexes in the context of photosynthesis. *Physiol. Plant.* **161**, 150–170 (2017).

40. Slesak, I., Slesak, H. & Kruk, J. Oxygen and hydrogen peroxide in the early evolution of life on earth: in silico comparative analysis of biochemical pathways. *Astrobiology* **12**, 775–784 (2012).

41. Heck, D. E., Shakarjian, M., Kim, H. D., Laskin, J. D. & Vetrano, A. M. Mechanisms of oxidant generation by catalase. *Ann. N. Y. Acad. Sci.* **1203**, 120–125 (2010).

42. Young, K. D. The Selective Value of Bacterial Shape. *Microbiol. Mol. Biol. Rev.* **70**, 660–703 (2006).

43. Philippe, J., Vernet, T. & Zapun, A. The elongation of ovococci. in *Microbial Drug Resistance* **20**, 215–221 (2014).

44. Tocheva, E. I., Ortega, D. R. & Jensen, G. J. Sporulation, bacterial cell envelopes and the origin of life. *Nature Reviews Microbiology* **14**, 535–542 (2016).

45. Ramos-Silva, P., Serrano, M. & Henriques, A. O. From Root to Tips: Sporulation Evolution and Specialization in Bacillus subtilis and the Intestinal Pathogen Clostridioides difficile. *Mol. Biol. Evol.* **36**, 2714–2736 (2019).

46. Dagan, T. & Martin, W. Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proc. Natl. Acad. Sci.* **104**, 870–875 (2007).

47. Konstantinidis, K. T. & Tiedje, J. M. Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci.* **101**, 3160–3165 (2004).

48. Xu, L. *et al.* Average gene length is highly conserved in prokaryotes and eukaryotes and diverges only between the two kingdoms. *Mol. Biol. Evol.* **23**, 1107–1108 (2006).

49. Jain, R., Rivera, M. C. & Lake, J. A. Horizontal gene transfer among genomes: The complexity hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* (1999). doi:10.1073/pnas.96.7.3801

50. Wellner, A., Lurie, M. N. & Gophna, U. Complexity, connectivity, and duplicability as barriers to lateral gene transfer. *Genome Biol.* **8**, R156 (2007).

51. Martiny, J. B. H., Jones, S. E., Lennon, J. T. & Martiny, A. C. Microbiomes in light of traits: A phylogenetic perspective. *Science (80-.).* **350**, aac9323–aac9323 (2015).

52. Cordero, O. X. & Hogeweg, P. The impact of long-distance horizontal gene transfer on prokaryotic genome size. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 21748–21753 (2009).

53. Kloesges, T., Popa, O., Martin, W. & Dagan, T. Networks of gene sharing among 329 proteobacterial genomes reveal differences in lateral gene transfer frequency

at different phylogenetic depths. *Mol. Biol. Evol.* **28**, 1057–1074 (2011).

54. Williams, D., Gogarten, J. P. & Papke, R. T. Quantifying homologous replacement of loci between haloarchaeal species. *Genome Biol. Evol.* **4**, 1223–1244 (2012).

55. Jeong, H., Arif, B., Caetano-Anollés, G., Kim, K. M. & Nasir, A. Horizontal gene transfer in human-associated microorganisms inferred by phylogenetic reconstruction and reconciliation. *Sci. Rep.* **9**, 5953 (2019).

56. Cohen, O., Gophna, U. & Pupko, T. The Complexity Hypothesis Revisited: Connectivity Rather Than Function Constitutes a Barrier to Horizontal Gene Transfer. *Mol. Biol. Evol.* **28**, 1481–1489 (2011).

57. Barberán, A., Caceres Velazquez, H., Jones, S. & Fierer, N. Hiding in Plain Sight: Mining Bacterial Species Records for Phenotypic Trait Information. *mSphere* **2**, e00237-17 (2017).

58. Woese, C. R. & Fox, G. E. The concept of cellular evolution. *J. Mol. Evol.* (1977). doi:10.1007/BF01796132

59. Gogarten, J. P. & Deamer, D. Is LUCA a thermophilic progenote ? *Nat. Publ. Gr.* **1**, 1–2 (2016).

60. Siefert, J. L. & Fox, G. E. Phylogenetic mapping of bacterial morphology. *Microbiology* **144**, 2803–2808 (1998).

61. Koch, A. L. Were Gram-positive rods the first bacteria? *Trends Microbiol.* **11**, 166–170 (2003).

62. Yulo, P. R. J. & Hendrickson, H. L. The evolution of spherical cell shape; progress and perspective. *Biochemical Society Transactions* **47**, 1621–1634 (2019).

63. Williams, T. A. *et al.* Integrative modeling of gene and genome evolution roots the archaeal tree of life. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E4602–E4611 (2017).

64. Wolf, Y. I., Makarova, K. S., Yutin, N. & Koonin, E. V. Updated clusters of orthologous genes for Archaea: a complex ancestor of the Archaea and the byways of horizontal gene transfer. *Biol. Direct* **7**, 46 (2012).

65. Wesson, P. S. Panspermia, past and present: Astrophysical and biophysical conditions for the dissemination of life in space. *Space Sci. Rev.* **156**, 239–252 (2010).

66. Jackson, M. G. *et al.* Evidence for the survival of the oldest terrestrial mantle reservoir. *Nature* **466**, 853–856 (2010).

67. Weiss, M. C., Preiner, M., Xavier, J. C., Zimorski, V. & Martin, W. F. The last

universal common ancestor between ancient Earth chemistry and the onset of genetics. *PLOS Genet.* **14**, e1007518 (2018).

68. Kuever, J., Rainey, F. A. & Widdel, F. *Bergey's Manual® of Systematic Bacteriology*. *Bergey's Manual® of Systematic Bacteriology* (2001). doi:10.1007/978-0-387-21609-6

69. Brenner, D. J., Krieg, N. R. & Staley, J. T. *Bergey's Manual® of Systematic Bacteriology*. *Bergey's Manual® of Systematic Bacteriology* (2005). doi:10.1007/0-387-29298-5

70. De Vos, P. *et al. Bergey's manual of systematic bacteriology Volume Three The Firmicutes*. *Bergey's Manual of Systematic Bacteriology* **3**, (Springer Science & Business Media, 2009).

71. Krieg, N. R. *et al. Bergey's Manual® of Systematic Bacteriology*. *Bergey's Manual® of Systematic Bacteriology* (2010). doi:10.1007/978-0-387-68572-4

72. Bornstein, B. T. & Barker, H. A. The Nutrition of Clostridium kluyveri. *J. Bacteriol.* **55**, 223–230 (1948).

73. Parte, A. C. LPSN - List of prokaryotic names with standing in nomenclature (Bacterio.net), 20 years on. *International Journal of Systematic and Evolutionary Microbiology* **68**, 1825–1829 (2018).

74. Mukherjee, S. *et al.* Genomes OnLine Database (GOLD) v.6: Data updates and feature enhancements. *Nucleic Acids Res.* **45**, D446–D456 (2017).

75. Pagel, M. Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884 (1999).

76. Brown, C. T. *et al.* Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208–211 (2015).

77. Castelle, C. J. & Banfield, J. F. Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. *Cell* **172**, 1181–1197 (2018).

78. Justice, S. S., Hunstad, D. A., Cegelski, L. & Hultgren, S. J. Morphological plasticity as a bacterial survival strategy. *Nat. Rev. Microbiol.* **6**, 162–168 (2008).

79. Stone, E. A. Why the phylogenetic regression appears robust to tree misspecification. *Syst. Biol.* **60**, 245–260 (2011).

80. Rambaut, A. & Drummond, A. J. Tracer v1.6. (2013).

81. Venditti, C., Meade, A. & Pagel, M. Multiple routes to mammalian diversity. *Nature* **479**, 393–396 (2011).

82. Baker, J., Meade, A., Pagel, M. & Venditti, C. Adaptive evolution toward larger size in mammals. *Proc. Natl. Acad. Sci.* **112**, 5093–5098 (2015).

83. Freckleton, Harvey & Pagel. Phylogenetic Analysis and Comparative Data: A Test and Review of Evidence. *Am. Nat.* **160**, 712 (2017).

84. Organ, C. L., Shedlock, A. M., Meade, A., Pagel, M. & Edwards, S. V. Origin of avian genome size and structure in non-avian dinosaurs. *Nature* **446**, 180–184 (2007).

85. Soucy, S. M., Huang, J. & Gogarten, J. P. Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.* **16**, 472 (2015).

86. Wiedenbeck, J. & Cohan, F. M. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol. Rev.* **35**, 957–976 (2011).

87. Caro-Quintero, A. & Konstantinidis, K. T. Inter-phylum HGT has shaped the metabolism of many mesophilic and anaerobic bacteria. *ISME J.* **9**, 958–967 (2015).

ACKNOWLEDGMENTS: We thank, O. Guadayol Roig, M. El Baidouri and R. Schuech, Ciara O'Donovan, Manabu Sakamoto and Joanna Baker for helpful comments and advice on the draft MS; Funding: This work is financially supported by a Leverhulme Trust Research Leadership Award (RL-2012-022), granted to S.H. and a Leverhulme Trust Research Project Grant (RPG-2013-185) to C.V.

Author contributions: F.E.B., C.V. and S.H designed the study; F.E.B., S.S, and S.H. developed the protocol for the data collection; F.E.B. and S.S. collected the data and F.E.B., C.V., S.S., A.M. and S.H. analysed the data. F.E.B., C.V., and S.H. wrote the first draft of the manuscript, and all authors contributed substantially to revisions.

Competing interests statement: The authors declare no conflict of interest.

Figures and Tables.

Fig. 1. Distribution of phenotypic traits on the larger of the two phylogenetic trees used in the study. To aid clarity only species with complete data for all traits are included (n = 674). Tips on the tree are colour coded according to cell shape, while the **inner** to **outer** rings are coded according to motility status (**inner**), habitat (**second**), oxygen requirement (**third**), spore formation (**fourth**) and cell wall type (**outer**). The outer bars are proportional to genome size in Mbp.

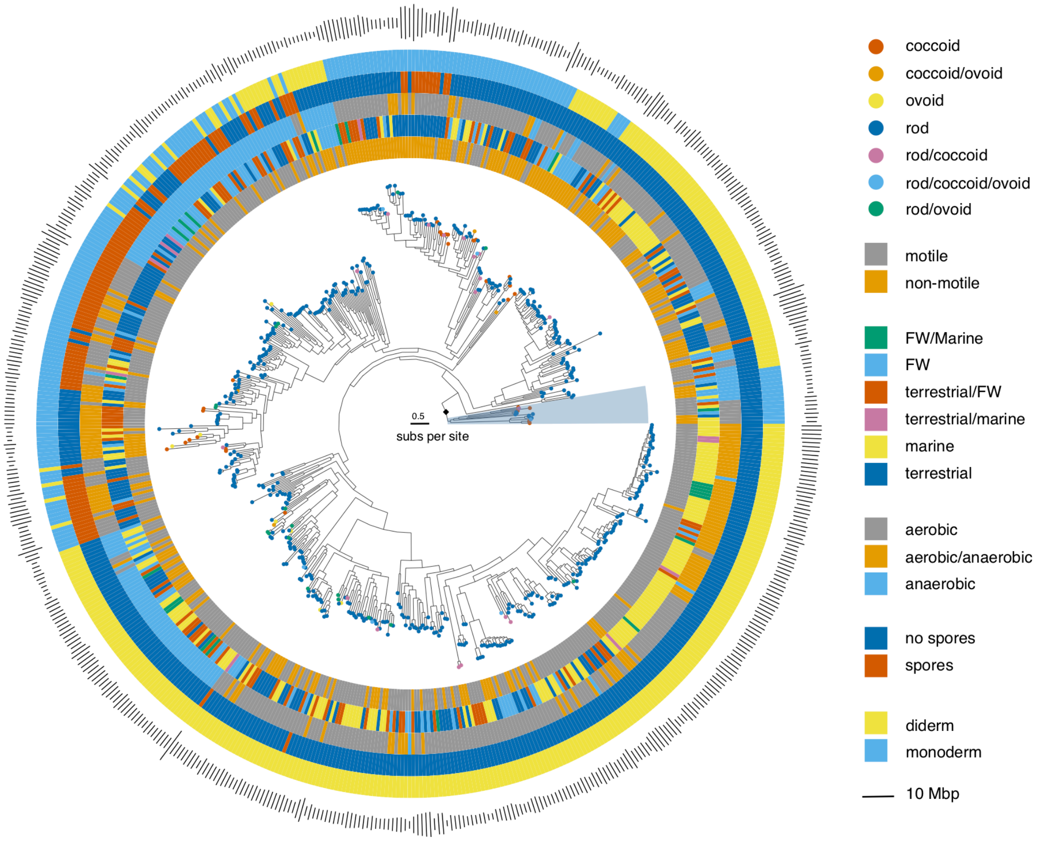

Fig. 2. Bayesian estimates of the ancestral state for categorical traits. Mean and median posterior probabilities (PP) for each character state for LUCA, LBCA and LACA with colour intensity proportional to the value of the PP. On the right side the character state with the highest PP for each character is shown (e.g. Motility: Motile). When the ancestral character state differs between LUCA, LBCA or LACA, alternative character states are shown within thick boxes (i.e. shape, temperature and pH). Values are from the large tree, and only consistent results for ancestral character states between the large and the small trees are reported. Supplementary Table 1 contains the full results for all characters. Abbreviations: LUCA - last universal common ancestor; LBCA - last bacterial common ancestor; LACA - last archaeal common ancestor; PP – posterior probability.

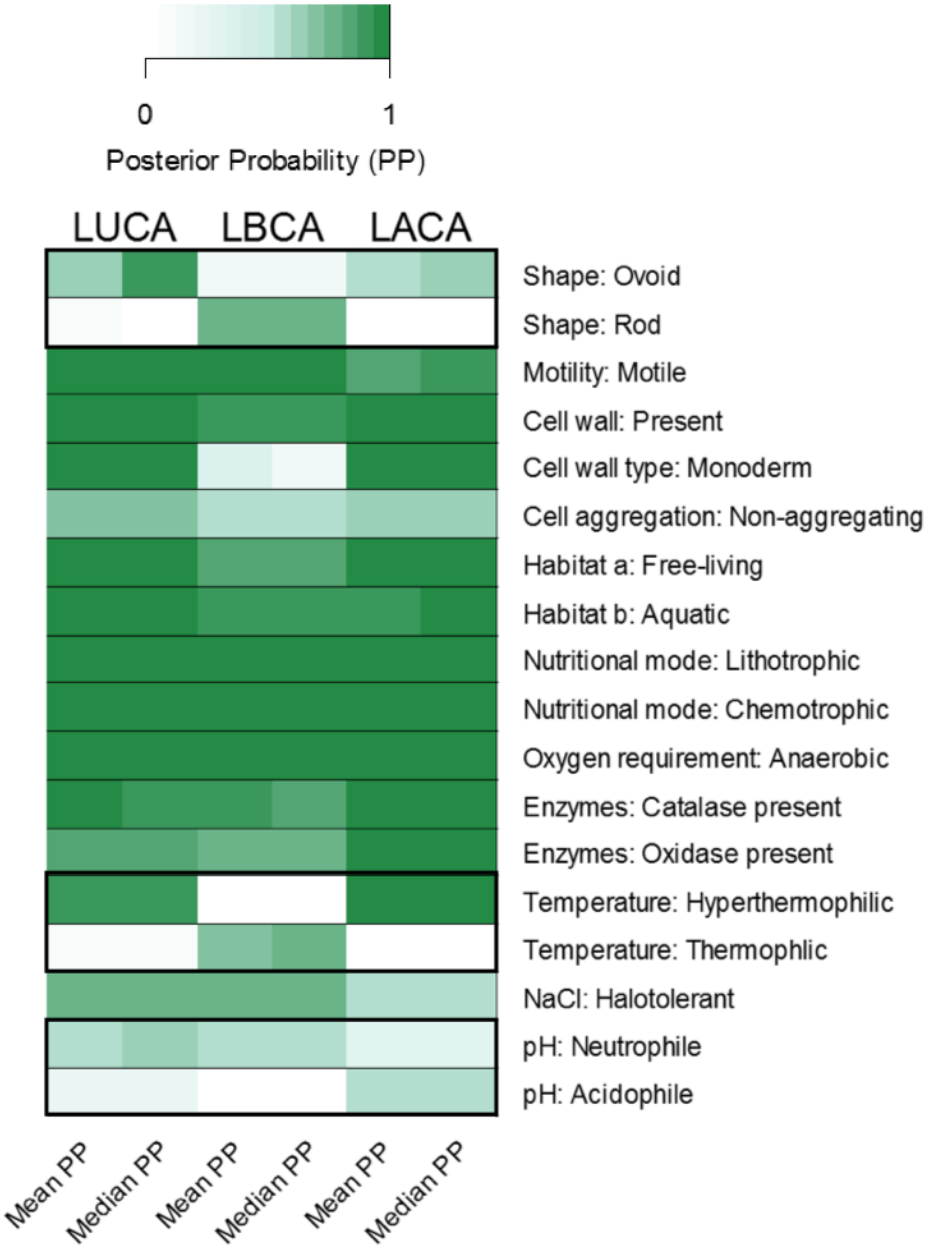

Fig. 3. Phenotypic reconstructions of ancestral cells. (A) LUCA, (B) LBCA, (C) LACA. Flagella-like appendages indicate motility, green outer layer a cell wall and red inner layer a cell membrane. Where available, grey shading on habitat categories represents continuous data estimation while black rectangles bracket the categorical estimates. Histograms give posterior probability distributions for the continuous trait values.

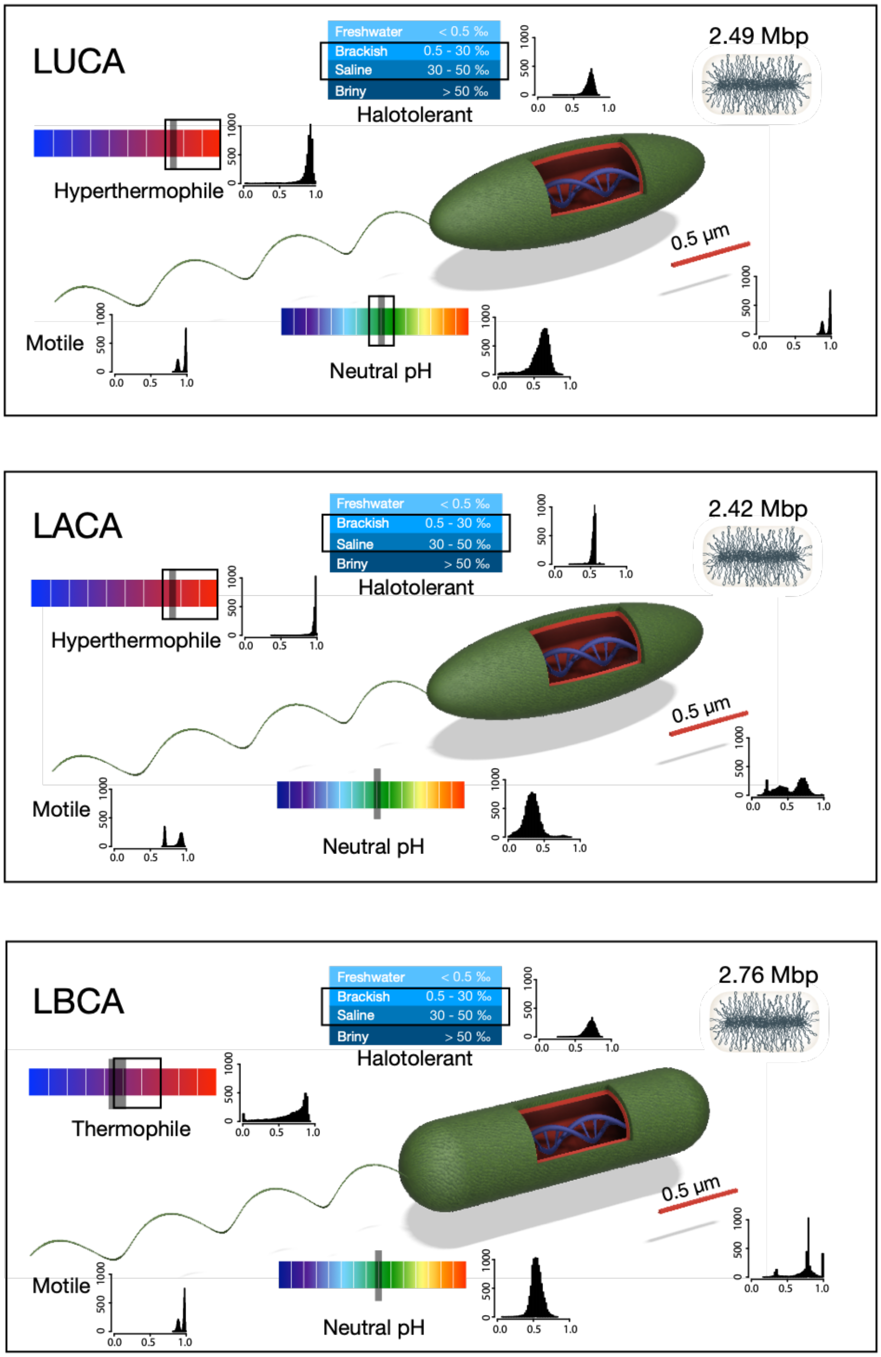

Table 1. Bayesian values estimates for continuous traits for LUCA, LBCA and LACA from the large tree. Ancestral state estimates for cell and genome sizes, optimum temperature, NaCl and pH from the best of two competing models - the random walk and the directional models. Phylogenetic signal (λ) is estimated for each trait. Abbreviations: LUCA: last universal common ancestor, LBCA: last bacterial common ancestor, LACA: last archaeal common ancestor. HPD - Highest Posterior Density; 95% PI - 95% Probability Interval.

	Number of species	**λ Lower HPD**	**λ Upper HPD**	**λ Lower 95% PI**	**λ Upper 95% PI**	**λ**	**LUCA**	**LBCA**	**LACA**
Cell size (µm)									
Average width	1705	0.99	1.00	0.99	1.00	1.00	0.55	0.48	0.52
Average length	1656	0.98	0.99	0.98	0.99	0.99	1.72	1.87	1.73
Genome									
Average genome size (Mb)	2978	1.00	1.00	1.00	1.00	1.00	2.49	2.76	2.42
Average gene number	2965	1.00	1.00	1.00	1.00	1.00	2629	2704	2588
Temperature (°C)									
Optimum lower	1448	1.00	1.00	1.00	1.00	1.00	72	51	73
Optimum upper	1432	1.00	1.00	1.00	1.00	1.00	73	43	74
pH									
Optimum lower	896	0.97	0.97	0.97	0.97	0.97	6.93	7.03	6.91
Optimum upper	879	0.96	0.96	0.96	0.96	0.96	7.05	7.16	6.03
NaCl (% w/v)									
Optimum lower	470	0.99	0.99	0.99	0.99	0.99	2.31	2.33	2.21
Optimum upper	479	0.98	0.98	0.98	0.98	0.98	2.76	2.83	2.70