# Extensive transfer of membrane lipid biosynthetic genes between Archaea and Bacteria

**Gareth A. Coleman[1], Richard D. Pancost[2], Tom A. Williams[1]***

**1. School of Biological Sciences, University of Bristol, BS8 1TH, Bristol, UK**
**2. School of Earth Sciences, University of Bristol, BS8 1TH, Bristol, UK**

***corresponding author: tom.a.williams@bristol.ac.uk**

## Abstract

The divergence between Bacteria and Archaea may represent the deepest split in the tree of life. One of the key differences between the two domains are their membrane lipids, which are synthesised by distinct biosynthetic pathways with non-homologous enzymes. This 'lipid divide' has important implications for the early evolution of cells, and motivates the hypothesis that the last universal common ancestor (LUCA) may have lacked a modern cell membrane. However, we still know surprisingly little about the natural diversity of prokaryotic lipids in modern environments, or the evolutionary origins of the genes that produce them. In particular, the discovery of environmental lipids, such as glycerol dialkyl glycerol tetraethers with a mixture of classically archaeal and bacterial features, suggest that the 'lipid divide' may be less clear cut than previously assumed. Here, we investigated the distribution and evolutionary history of membrane lipid biosynthesis genes across the two domains. Our analyses reveal extensive inter-domain horizontal transfer of core lipid biosynthetic genes, and suggest that many modern Bacteria and Archaea have the capability to biosynthesize membrane lipids of the opposite "type". Gene tree rooting further suggests that the canonical archaeal pathway could be older than the bacterial pathway, and could have been present in LUCA.

## Introduction

Archaea and Bacteria form the two primary domains of life (see review Williams et al. 2013). While overwhelming similarities in their fundamental genetics and biochemistry and evidence of homology in a near-universally conserved core of genes (Weiss et al. 2016) strongly suggest that Archaea and Bacteria descend from a universal common ancestor (LUCA), they also differ in ways that have important implications for the early evolution of cellular life. These differences include DNA replication (Kelman and Kelman 2014), transcription (Bell and Jackson 1998), DNA packaging (Reeve et al. 1997), and cell wall compositions (Kandler 1995). One striking difference is in the lipid composition of the cell membranes (Fig. 1), which is particularly important for understanding the origin of cellular life. Canonically, Archaea have isoprenoid chains attached to a glycerol-1-phosphate (G1P) backbone via ether bonds, and can have either membrane spanning or bilayer-forming lipids (Lombard et al. 2012). Most bacteria, as well as eukaryotes, classically have acyl (fatty acid) chains attached to a glycerol-3-phosphate (G3P) backbone via ester bonds and form bilayers (Lombard et al. 2012), although a number of exceptions have been documented (Sinninghe Damsté et al. 2002a; Weijers et al. 2006; Sinninghe Damsté et al. 2007; Goldfine 2010). These lipids are synthesised by non-homologous enzymes via different

biosynthetic pathways (Fig. 1). This so-called 'lipid divide' (Koga 2011) raises some important questions regarding the early evolution of cellular life, including the number of times cell membranes evolved, whether LUCA had a cell membrane, and if so, what the nature of that membrane was.

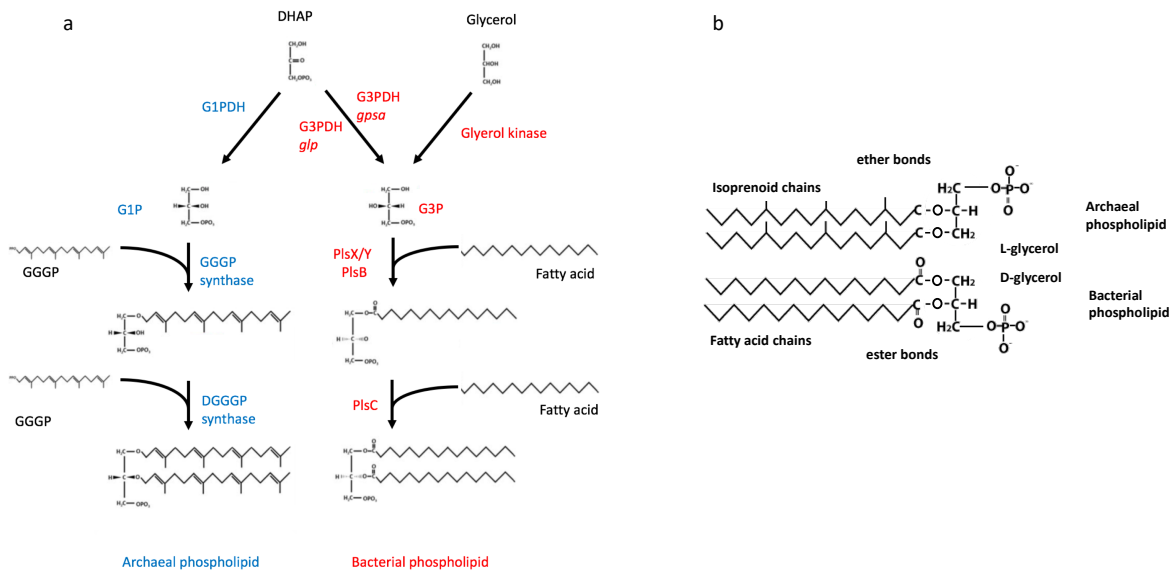


**Figure 1. a)** The canonical phospholipid biosynthetic pathways in Archaea and Bacteria. Archaeal enzymes in blue, bacterial enzymes in red. **b)** Composition of bacterial and archaeal phospholipids. In Archaea, G1P is synthesised from dihydroxyacetone phosphate (DHAP) using the enzymes glycerol-1-phosphate dehydrogenase (G1PDH). The first and second isoprenoid chains (GGGPs) are added by garenygarenylglcerol synthase (GGGPS) and digarenylgarenylgecerol synthase (DGGGPS) respectively. In Bacteria, G3P is synthesised by glycerol-3-phosphate dehydrogenase (G3PDH) from DHAP. There are two forms of this enzyme, encoded by the *gpsA* and *glp* genes respectively. G3P may also be produced form glycerol by glycerol kinase (glpK). In certain Bacteria, such as Gammaproteoteobacteria, the first fatty-acid chain is added by a version of glycerol-3-phosphate acyltransferase encoded by the *PlsB* gene. Other Bacteria, including most gram positive bacteria, use a system another glycerol-3-phosphate acyltransferase encoded by *PlsY*, in conjunction with an enzymes encoded by *PlsX* (Yoa and Rock 2013; Parsons and Rock 2013). The second fatty-acid chain is attached by 1-acylglycerol-3-phosphate O-acyltransferase, encode by *PlsC*.

The observation that lipid biosynthesis in Bacteria and Archaea is non-homologous has motivated the hypotheses that LUCA had no, or only a rudimentary, membrane, with modern ion-tight membranes evolving in parallel along the stem lineage of each prokaryotic domain (Koga et al. 1998; Martin and Russell 2003; Sousa et al. 2013; Sojo et al. 2014; Sousa and Martin 2014). In these scenarios, the acellular LUCA is envisaged as a non-free living entity that lived on a pyrite substrate (Wächtershäuser 1988; Wächtershäuser 1992) or inhabited mineral compartments within alkaline hydrothermal vents (Martin and Russell 2003; Sousa et al. 2013); the latter hypothesis gains additional explanatory strength from similarities between the geochemistry at modern vents and the biochemistry of one of the most ancient pathways for biological carbon fixation, the Wood Ljungdahl pathway (Sousa et al. 2013; Weiss et al. 2016; Williams et al. 2017; Adam et al. 2018). An alternative view is that LUCA may have had a fully modern, ion-tight membrane, which was heterochiral with respect to

membrane stereochemistry (Wächtershäuser 2003), with later independent transitions to homochirality in Bacteria and Archaea driven by increased membrane stability. However, the available experimental evidence - including the recent engineering of an *Escherichia coli* cell with a heterochiral membrane (Caforio et al. 2018) - suggests that homochiral membranes are not necessarily more stable than heterochiral ones (Fan et al. 1995; Shimada and Yamagishi 2011; Caforio et al. 2018), requiring some other explanation for the loss of ancestral heterochirality.

The lipid divide has also played an important role in debates about the origin of eukaryotic cells, whose membrane lipids are predominantly (though not exclusively (Goldfine 2010; Tan et al. 2012)) of the bacterial type. Thus, scenarios for eukaryote origins that invoke a symbiosis between an archaeal host cell and a bacterial endosymbiont – the class of scenarios that has received the greatest phylogenetic and comparative genomic support (Martin and Muller 1998; Embley and Martin 2006; Cox et al. 2008; Guy and Ettema 2011; Williams et al. 2013; Martin et al. 2015; Williams and Embley 2015; Spang et al. 2015; Eme et al. 2017; Zaremba-Niedzwiedzka et al. 2017) – require a corollary to explain the change in host membrane composition. Several possibilities have been proposed. Gould et al. (2016) posit that vesicle secretions from the mitochondria were incorporated into the cell membrane, altering its composition. Baum and Baum (2014), as part of their 'inside-out' model, suggest that eukaryotes could have acquired bacterial lipids from mitochondria via traffic across ER-mitochondrial contact sites. Bacteria may have adopted 'bacterial' lipids due to living in mesophilic environments and no longer needing membranes adapted for thermophilic conditions. Other endosymbiotic models, such as the syntrophy model (López-García and Moreira 2006), or non-endosymbiotic models have been devised to explain bacterial lipids in eukaryotes (Woese et al. 1990; Kandler 1995).

Despite the importance of the lipid divide for our understanding of early cellular evolution, membrane lipid stereochemistry has been experimentally determined for a surprisingly limited range of Bacteria and Archaea, perhaps owing to the expense and difficulty of characterisation (Sinninghe Damsté, et al. 2002a; Weijers et al. 2006; Sinninghe Damsté et al. 2014). Interestingly, the limited data that are available suggest that the distinction between bacterial and archaeal membranes may be less clear-cut than was previously thought. For example, the model bacterium *Bacillus subtilis* has been shown to synthesise both bacterial- and archaeal-type lipids (Guldan et al. 2008; Guldan et al. 2011). Further, some characterised lipids possess a mixture of archaeal and bacterial features, including the plasmalogens of animals and anaerobic bacteria, which include an ether bond (Goldfine 2010), and the branched glycerol dialkyl glyercol tetra-ether (brGDGT) lipids of Bacteria (Sinninghe Damsté et al. 2002b). brGDGTs have bacterial stereochemistry and branched rather than isoprenoidal alkyl chains, but they also contain ether bonds and span the membrane, as observed for canonical archaeal lipids. These brGDGTs are particularly abundant in peat bogs and were thought to be produced by Bacteria as adaptations to low pH environments (Sinninghe Damsté et al. 2002b; Weijers et al. 2006; Sinninghe Damsté et al. 2007), but are now known to occur in a wide range of soils and aquatic settings (Schouten et al. 2013a). The enzymes responsible for their synthesis are currently unknown. On the other side of the "lipid divide", some Archaea have been shown to produce membrane lipids with fatty acid chains and ester bonds (Gattinger et al. 2002). The biosynthetic pathways for all of these mixed-type membrane lipids remain unclear. However, given the frequency with which prokaryotes undergo horizontal gene transfer (Garcia-Vallvé

et al. 2000), one possibility is that these mixed biochemical properties reflect biosynthetic pathways of mixed bacterial and archaeal origin. This prompted us to investigate the distribution and phylogeny of phospholipid biosynthesis enzymes across the two domains and evaluate the evidence for inter-domain horizontal gene transfer. Our analysis focused on the core enzymes that establish membrane lipid stereochemistry and attach the two carbon chains to the glycerol-phosphate backbone (Figure 1), as the histories of these enzymes are key to understanding the evolution of membrane stereochemistry and biosynthesis.

## Results and Discussion

*Extensive inter-domain lateral transfer of core phospholipid biosynthesis genes*

We performed BLASTp searches for the enzymes of the canonical archaeal and bacterial lipid biosynthesis pathways (Figure 1) against all archaeal and bacterial genomes in the NCBI nr database. Our BLAST searches revealed homologues for all of the core phospholipid biosynthesis genes of both pathways in both prokaryotic domains, with the exception of bacterial enzymes PlsB and PlsX, which we did not find in Archaea. Orthologues of the canonical archaeal genes are particularly widespread in many bacterial lineages (Table 1). Of 48 bacterial phyla, 6 have at least one sequence identified as an orthologue of each of the three archaeal genes (Table 1, highlighted in yellow). Of these phyla, Firmicutes (genera *Bacillus*, *Halanaerobium*), Actinobacteria (genera *Strepomyces*) and Fibrobacteres (genera *Chitinispirillum* and *Chitinivibrio*) contain species which have all three genes in their genomes (Table 1, highlighted yellow with asterisk). Based on the presence of all three core biosynthetic genes, and given their recognised role in lipid biosynthesis in *B. subtilis* (Guldan et al. 2008; Guldan et al. 2011), we predict that these members of the Firmicutes, Actinobacteria and Fibrobacteres lineages of Bacteria are capable of making archaeal-type, glycerol-1-linked phospholipids. Of the seven FBC (Fibrobacteres, Bacteroidetes and Chlorobi) phyla we surveyed, all of them have GGGPS and DGGGPS orthologues, but only Fibrobacteres have G1PDH orthologues (see Figure 1 for overview of pathway). In these species lacking G1PDH, it is unclear whether GGGPS and DGGGPS are active and if so, what they are used for; one possibility is that they catalyse the reverse reaction, catabolising archaeal lipids as an energy source. Just 10 of the 48 phyla had no orthologues of archaeal genes (Table 1, indicated by †).

Orthologues of the canonical bacterial genes are less widespread in Archaea (Table 1). Of the 20 phyla surveyed, none contained all homologues, although Lokiarchaeota contained Glp, GlpK, PlsC and PlsY. Of those 20, more than half (11) had no bacterial homologues. Orthologues of GpsA, Gpl and Gpk are found in all of the major archaeal groups (Euryarchaeota, TACK, Asgardarchaeota and DPANN (Williams et al. 2017)). However, they appear sporadically. Within Euryarchaeota, of the seven classes surveyed, GpsA and Glpk appear in four and Glp in five. With the TACK superphylum, Glp and GlpK appear in Crenarchaeota and Korarchaeota, while GpsA appears only in a single crenarchaeote (*Thermofilum*). GpsA is also found in two of the 11 DPANN phyla surveyed (Woesarchaeota and GW2011), while GlpK is found in one phylum (Woesearchaeota) and Glp is found in none. Within the Asgardarchaeota superphylum, no orthologues for GpsA are found, and only one of the four phyla (Lokiarchaeota) has Glp or GlpK. PlsC and PlsY are more restricted, being found

186 **Table 1 – Distribution of Lipid Biosynthesis gene**

187

| Domain | Super-phylum | Phylum | Class | G1PDH | GGGPS | DGGGPS | GpsA | Glp | GlpK | PlsC | PlsY |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Archaea | | Euryarchaeota | Archaeoglobi | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | |
| | | | Halobacteria | ✓ | ✓ | ✓ | | ✓ | ✓ | | |
| | | | Methanobacteria | ✓ | ✓ | ✓ | ✓ | | | | |
| | | | Methanococci | ✓ | ✓ | ✓ | | | | | |
| | | | Methanomicrobia | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| | | | Thermococci | ✓ | ✓ | ✓ | | ✓ | ✓ | | |
| | | | Thermoplasmatales | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| | TACK | Aigarchaeota | | ✓ | ✓ | ✓ | | | | | |
| | | Crenarchaeota | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | |
| | | Korarchaeota | | ✓ | ✓ | ✓ | | ✓ | ✓ | | |
| | | Thaumarchaeta | | ✓ | ✓ | ✓ | | | | | |
| | Asgard | Heimdallarchaeota | | ✓ | ✓ | ✓ | | | | ✓ | ✓ |
| | | Lokiarchaeota | | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ |
| | | Odinarchaeota | | ✓ | ✓ | ✓ | | | | | |
| | | Thorarchaeota | | ✓ | ✓ | ✓ | | | | | ✓ |
| | DPANN | Aenigmarchaeota | | ✓ | ✓ | ✓ | | | | | |
| | | Diapherotrites (GW2011_AR10/DUSEL3) | | | ✓ | ✓ | ✓ | | | | ✓ |
| | | Micrarchaeota (incl. Macid) | | ✓ | ✓ | ✓ | | | | | |
| | | Nanoarchaeota | | | | | | | | | |
| | | Nanohaloarchaeota | | | | | | | | | |
| | | Pacearchaeota | | | | | | | | | ✓ |
| | | Parvarchaeota | | | | | | ✓ | ✓ | | |
| | | Woesearchaeota | | | | | ✓ | | ✓ | ✓ | ✓ |
| Bacteria | | Acidobacteria | | ✓ | | | ✓ | | ✓ | ✓ | ✓ |
| | | Actinobacteria* | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | | Aminicenantes | | ✓ | | | ✓ | ✓ | ✓ | | ✓ |
| | | Aquificae† | | | | | ✓ | ✓ | ✓ | ✓ | ✓ |
| | | Armatimonadetes | | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ |
| | | Caldithrix | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ |
| | | Candidate division KSB1 | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | | Candidate division NC10† | | | | | ✓ | ✓ | ✓ | ✓ | ✓ |
| | | Candidate division TA06 | | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ |
| | | Candidate division WOR-3 | | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ |
| | | Candidatus Edwardsbacteria | | | ✓ | ✓ | ✓ | | | ✓ | ✓ |
| | | Candidatus Handelsmanbacteria | | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ |
| | | Candidatus Kerfeldbacteria | | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ |
| | | Candidatus Magnetoovum | | ✓ | ✓ | | ✓ | | | ✓ | ✓ |
| | | Candidatus Marinimicrobia | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | | Candidatus Raymondbacteria | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | | Chloroflexi | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | | Chrysiogenetes | | ✓ | ✓ | | ✓ | | | ✓ | ✓ |
| | | Cloacimonetes | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | | Cyanobacteria | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Deferribacterales† | | | | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Deinococcus-Thermus | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Dictyoglomi | ✓ | | | ✓ | | ✓ | ✓ | ✓ |
| | Elusimicrobia | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Firmicutes* | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Fusobacteria | | ✓ | | ✓ | ✓ | ✓ | | ✓ |
| | Gemmatimonadetes | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Latescibacteria | | ✓ | ✓ | ✓ | | ✓ | | ✓ |
| | Lentisphaerae | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Melainabacteria | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Nitrospinae† | | | | ✓ | ✓ | | ✓ | ✓ |
| | Nitrospirae | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Parcubacteria | | ✓ | ✓ | ✓ | | | ✓ | ✓ |
| | Proteobacteria | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Rhodothermaeota | | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ |
| | Spirochaetes | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Synergistetes | ✓ | | | ✓ | | ✓ | ✓ | ✓ |
| | Tenericutes† | | | | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Thermobaculum† | | | | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Thermodesulfobacteria† | | | | ✓ | | | ✓ | ✓ |
| | Thermotogae | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ |
| | TMED | | ✓ | | | | | ✓ | ✓ |
| FBC | Chlorobi | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Bacteroidetes | | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ |
| | Fibrobacteres* | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ |
| | Ignavibacteria | | ✓ | ✓ | ✓ | | | ✓ | ✓ |
| | Candidatus Kryptonium | | ✓ | ✓ | ✓ | | | ✓ | ✓ |
| | Candidatus Kryptobacter | | ✓ | ✓ | ✓ | | | ✓ | ✓ |
| | Candidate division Zixibacteria | | ✓ | ✓ | ✓ | | | ✓ | ✓ |
| PVC | Chlamydiae† | | | | ✓ | | ✓ | ✓ | ✓ |
| | Planctomycetes | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Verrumicrobia† | | | | ✓ | | ✓ | ✓ | ✓ |

**Table 1. Table detailing the distribution of lipid biosynthesis genes in bacteria and archaeal phyla.** Ticks represent phyla (class level for Euryarchaeota) with at least one genome which has a sequence for the corresponding gene. Bacterial phyla where all three archaeal genes are found are highlighted in yellow. Those bacterial phyla where all three archaeal genes are found within the same genome in at least one case, are marked with an asterisk (*). Those bacterial phyla with no archaeal genes are found are marked with †. It should be noted that in the case of environmental lineages, the lack of a tick may not represent absence of genes, given that these represent metagenomics bins, and the lack of said genes may be due to missing data. FBC are Fibrobacteres, Bacteroidetes and Chlorobi. PVC are Planctomycetes, Verrucomicrobia and Chlamydiae. TACK are Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota. DPANN include Diaphoretrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota, as well as several other lineages.

203 mainly in environmental lineages within Euryarchaeota (Marine Groups II/III, all in
204 class Thermoplasmatales), DPANN and Asgardarchaeota (Table 1).
205
206
207 *Early origins of archaeal-type membrane lipid biosynthesis genes in Bacteria*
208
209 To investigate the evolutionary histories of these genes, we inferred Bayesian single-
210 gene phylogenies from the amino acid sequences using PhyloBayes 4.1 (Lartillot and
211 Philippe 2004; Lartillot et al. 2007). We selected the best-fitting substitution model for
212 each gene according to its Bayesian Information Criterion (BIC) score using the model
213 selection tool in IQ-Tree (Nguyen et al. 2015). We used two complementary
214 approaches to root these single-gene trees: a lognormal uncorrelated molecular clock
215 in BEAST (Drummond and Rambaut 2007; Drummond et al. 2012), and the recently-
216 described minimal ancestor deviation (MAD) rooting method of Tria et al. (2017). The
217 MAD algorithm finds the root position that minimises pairwise evolutionary rate
218 variation, averaged over all pairs of taxa in the tree. Many of our single gene trees
219 were poorly resolved; since the existing implementation of the MAD algorithm (Tria et
220 al. (2017)) does not explicitly incorporate topological uncertainty, we used MAD to root
221 all of the trees sampled during the MCMC run, summarising posterior root support in
222 the same way as for the molecular clock analyses; in the discussion below, we use
223 the maximum *a posteriori* root as a point estimate for comparison between the two
224 methods. For the genes for which a suitable outgroup was available (G1PDH, GpsA
225 and Glp, following Yokobori et al. 2016), we compared our results to traditional
226 outgroup rooting. For more details, see Materials and Methods.
227
228 Glycerol-1-phosphate dehydrogenase (G1PDH) is the enzyme that establishes
229 phospholipid stereochemistry in Archaea. The archaeal G1PDH sequences form a
230 well-supported clan (Wilkinson et al. 2007)(PP = 0.99) in the tree (Figure 2a), although
231 the relationships within the group are poorly resolved. Interestingly, the majority of the
232 bacterial G1PDH orthologues do not appear to be recent horizontal acquisitions from
233 Archaea, but instead form a deep-branching clan resolved as sister to the archaeal
234 lineage. The root position that receives the highest posterior support in the relaxed
235 molecular clock analysis is that between the archaeal and bacterial clans, with a
236 marginal posterior probability of 0.68 (Supplementary Table 1). This is substantially
237 higher than the next most probable position, which places the root within the Bacteria
238 with a posterior of 0.1. When rooted using MAD, the same root position is recovered
239 with a marginal posterior probability of 0.62, also substantially higher than the next
240 most probable root of 0.1. Rooting single genes trees can prove difficult, and this
241 uncertainty is captured in the low root probabilities inferred using both the molecular
242 clock and MAD methods. However, these analyses can be used to exclude the root
243 from some regions of the trees with a degree of certainty. In the case of G1PDH, a
244 post-LUCA origin of the gene would predict a root on the archaeal stem or within the
245 archaea. In our analyses, no such root position has a significant probability (i.e.
246 PP>0.05), and therefore the root is highly unlikely to be within the archaea. The
247 bacterial clan mainly comprises sequences from Firmicutes and Actinobacteria, with
248 most of the other Bacteria grouping together in a single, maximally supported (PP =
249 1) lineage suggestive of recent horizontal acquisition from the
250 Firmicutes/Actinobacteria clade, followed by further HGT.
251

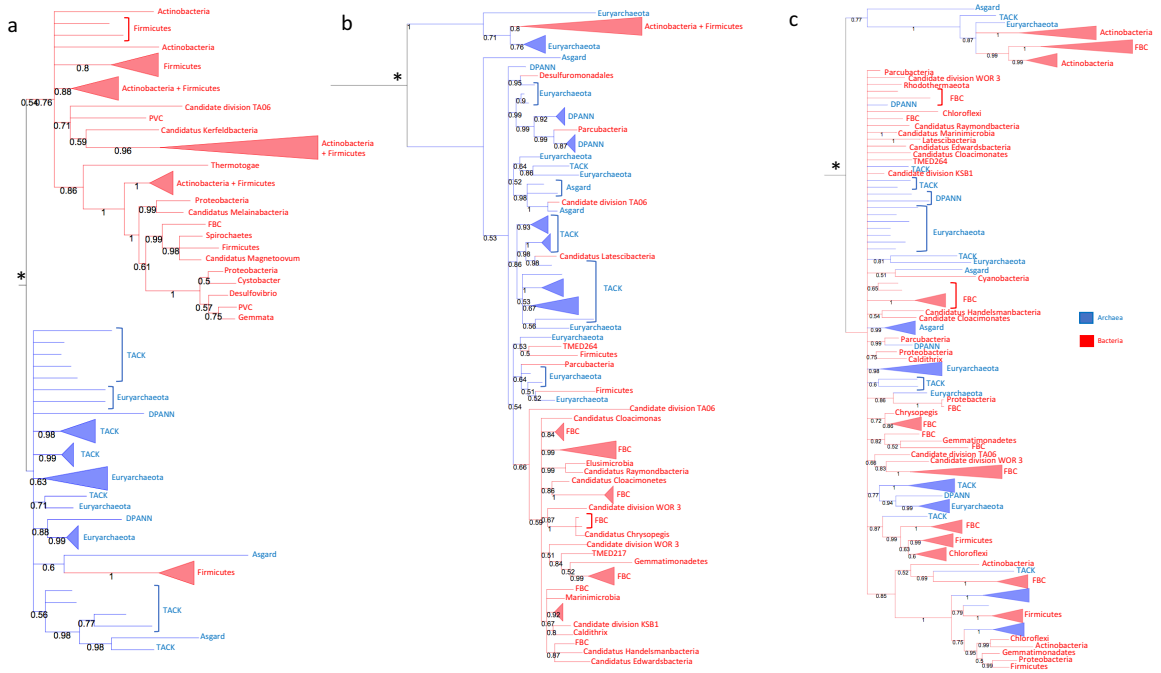


**Figure 2. Bayesian consensus trees of archaeal enzymes, rooted using the uncorrelated lognormal molecular clock.** Support values are Bayesian posterior probabilities, and the asterisk denotes the modal root position obtained using the MAD approach. Archaea in blue and Bacteria in red **a)** G1PDH tree (111 sequences, 190 positions) inferred under the best-fitting LG+C60 model. **b)** GGGPS tree (133 sequences, 129 positions) inferred under the best-fitting LG+C40 model. **c)** DGGGPS tree (177 sequences, 97 positions) inferred under the best-fitting LG+C50 model. FBC are Fibrobateres, Bacteroidetes and Chlorobi. PVC are Planctomycetes, Verrucomicrobia and Chlamyidiae. TACK are Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota. DPANN include Diaphoretrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota, as well as several other lineages. For full trees, see Supplementary Figures 1-4. For full unrooted trees see Supplementary Figures 16-18.

This root position is consistent with two scenarios that we cannot distinguish based on the available data. One possibility is an early transfer of G1PDH from stem Archaea into Bacteria, either into the bacterial stem lineage with subsequent loss in later lineages, or into the ancestor of Actinobacteria and Firmicutes, with subsequent transfers to other Bacteria. Alternatively, G1PDH could have already been present in LUCA, and was subsequently inherited vertically in both Archaea and Bacteria, followed by loss in later bacterial lineages. The Firmicute sequences within the archaeal clade are a later transfer into those Firmicutes, apparently from a thorarchaeote archaeon or related lineage.

Geranylgeranylglyceryl phosphate synthase (GGGPS) attaches the first isoprenoid chain to G1P. Phylogenetic analysis of GGGPS (Fig 2b) revealed two deeply divergent paralogues, with the tree confidently rooted between them using both the relaxed molecular clock (PP = 0.99) and MAD methods (PP = 1) (Supplementary Table 1); resolution within each of the paralogues was poor. One of these paralogues comprises sequences from some Euryarchaeota (including members of the Haloarchaea, Methanomicrobia and Archaeoglobi), along with Firmicutes and Actinobacteria. The other paralogue comprises sequences from the rest of the Archaea – including other Euryarchaeota - and a monophyletic bacterial clade largely consisting of members of

286 the FBC lineage. Taken with the root position between the two paralogues, the tree
287 topology implies an ancestral duplication followed by sorting out of the paralogues and
288 multiple transfers into Bacteria. To improve resolution among the deeper branches of
289 the tree, we inferred an additional phylogeny focusing just on the larger of the two
290 paralogues (Supplementary Figure 3). The root within this paralogous sub-tree fell
291 between reciprocally monophyletic archaeal and bacterial clades (PP = 0.8, much
292 higher than the next most likely root, within the bacteria, with PP = 0.07), suggesting
293 that the gene duplication at the base of the GGGPS tree pre-dates LUCA. The
294 recovery of two distinct paralogues has been noted in several previous studies
295 (Nemoto et al. 2003; Boucher et al. 2004; Peterhoff et al. 2014). Since genes from
296 both GGGPS paralogues have been experimentally characterised as
297 geranylgeranylglyceryl phosphate synthases (Nemoto et al. 2003; Boucher et al 2004),
298 it appears that this activity was already present in LUCA before the radiation of the
299 bacterial and archaeal domains.
300
301 Digeranylgeranylglyceryl phosphate synthase (DGGGPS) attaches the second
302 isoprenoid chain to G1P. DGGGPS is present in all sampled Archaea, with the
303 exception of three of the DPANN genomes. Although the DGGGPS tree is poorly
304 resolved (Fig. 2c), both the molecular clock and MAD root the tree between two clades,
305 with a diversity of Archaea and Bacteria on either side (PP = 0.41 and 0.79
306 respectively) (Supplementary Table 1). The wide distribution of this enzyme across
307 both Archaea and Bacteria, and the occurrence of a diversity of both domains on either
308 side of the root, for both rooting methods, suggest either multiple transfers into
309 Bacteria from Archaea, or that DGGGPS was present in LUCA and inherited in various
310 archaeal and bacterial lineages, followed by many later losses in and transfers
311 between various lineages.
312
313 In sum, our phylogenetic analyses of archaeal lipid biosynthesis genes suggest that
314 GGGPS and DGGGPS were already present in LUCA, with G1PDH either present in
315 LUCA or evolving along the archaeal stem. They also provide evidence for repeated,
316 independent inter-domain transfer of these genes from archaea to bacteria throughout
317 the evolutionary history of life.
318
319 *Transfers of bacterial membrane lipid genes into Archaea*
320
321 In contrast to our analyses of proteins of the classical archaeal pathway, phylogenies
322 of proteins of bacterial-type lipid biosynthesis pathways suggested that their
323 orthologues in Archaea were the result of relatively recent horizontal acquisitions. The
324 root positions for each of the trees using both molecular clock and MAD have low
325 posterior probabilities (Supplementary Table 1), so that the exact root positions are
326 unclear; yet, support for root positions outside of the Bacteria was never obtained. This
327 is consistent with the hypothesis that the core bacterial pathway first evolved after the
328 bacterial lineage diverged from LUCA.
329
330 GpsA and glp are two genes that code for glycerol-3-phosphate (G3PDH), which
331 establishes phospholipid stereochemistry in Bacteria. The deep relationships between
332 the archaeal and bacteria sequences in the GpsA tree are poorly resolved (Fig. 3a),
333 while being better resolved for Glp (Fig. 3b). The root position in both trees is poorly
334 resolved for both rooting methods (Supplementary Table 1). The highest marginal
335 posterior probability for the root positions recovered in the GpsA tree are 0.31 and

0.59 and for the molecular clock and MAD respectively, and 0.5 and 0.44 respectively for Glp. The tree inferred for GlpK (the gene that codes for glycerol synthase, which can synthesise G3P from glycerol) (Fig. 4a) shows a similar pattern like the phylogenies of GpsA and Glp. Again, the root positions have low posterior support (0.47 and 0.34 for the molecular clock and MAD respectively). However, in each case, there is evidence of multiple recent transfers from Bacteria to Archaea, as we recover several distinct bacterial and archaeal clades with moderate to high support (0.8-1). The main archaeal recipients of these genes are Euryarchaeota which is consistent with reports of bacterial-like fatty acid esters in this group (Gattinger et al. 2002) , and which may suggest the occurrence of an earlier transfer into the stem lineage of this clade. The tree topology also supports a number of more recent transfer events into various archaeal lineages.

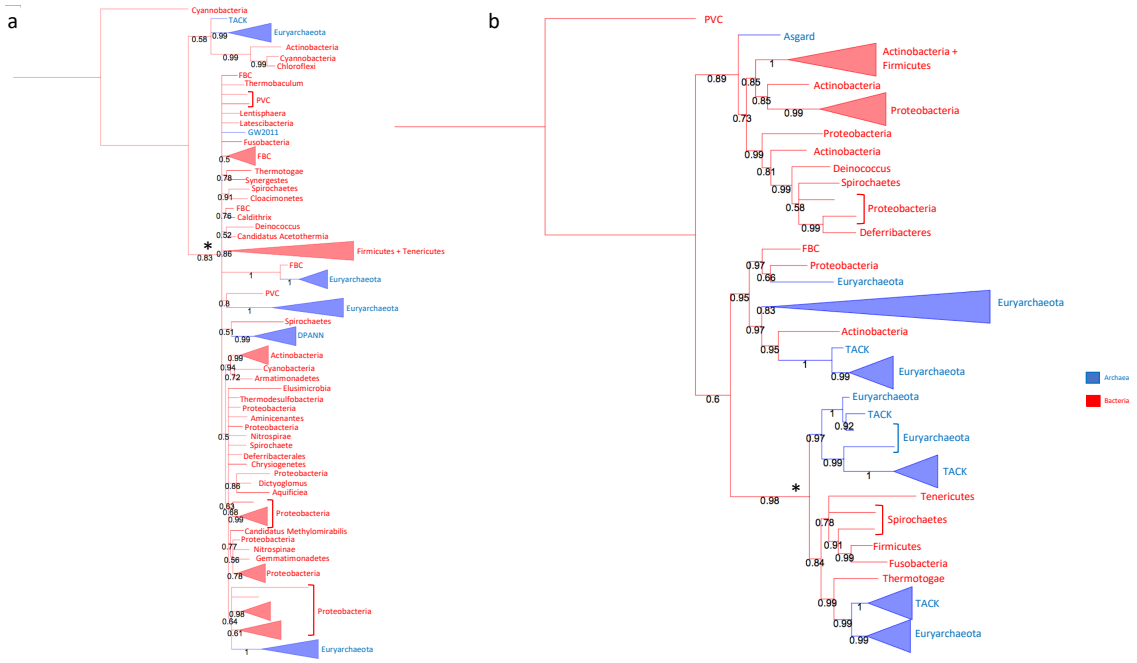


**Figure 3. Bayesian consensus trees of both G3PDH enzymes, rooted using the uncorrelated lognormal molecular clock**. Support values are Bayesian posterior probabilities, and the asterisk denotes the modal root position obtained using the MAD approach. Archaea in blue and Bacteria in red. **a)** gpsA tree (84 sequences, 169 positions) inferred under the best-fitting LG+C60 model **b)** glp tree (51 sequences, 199 positions) inferred under the best-fitting LG+C40 model. FBC are Fibrobateres, Bacteroidetes and Chlorobi. PVC are Planctomycetes, Verrucomicrobia and Chlamyidiae. TACK are Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota. DPANN include Diaphoretrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota, as well as several other lineages. For full trees, see Supplementary Figures 5-6. For full unrooted trees see Supplementary Figures 19-20.

PlsC and PlsY (which attach fatty acids to G3P) both have many fewer orthologues among archaeal genomes, all of which are derived from environmental sa(Embley and Martin 2006; Martin et al. 2015; Eme et al. 2017)mples. Both trees are poorly resolved (Fig 4b). Both are rooted within the Bacteria, with PlsC having the low posterior of 0.28 (with the next most likely, also within the Bacteria, being 0.1). The PlsY (Fig 4c) has a more certain root position, with a posterior of 0.57, and the next most probable being

0.1. For PlsY, MAD recovers the same root as the molecular clock, with a high posterior probability (0.85). When the PlsC tree is rooted using MAD, the root is resolved between two clades, which are not recovered in the inferred tree topology and has a low posterior probability of 0.03. All of the archaeal homologues seem to be derived from transfers from Bacteria to Archaea.

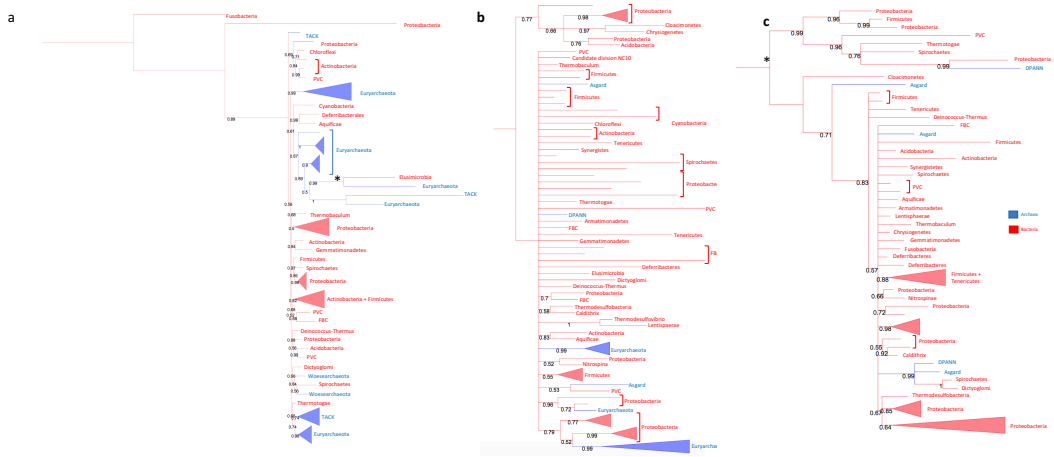


**Figure 4. Bayesian consensus trees of glpK, PlsC and PlsY enzymes, rooted using the uncorrelated lognormal molecular clock**. Support values are Bayesian posterior probabilities, and the asterisk denotes the modal root position obtained using the MAD approach. Archaea in blue and Bacteria in red. **a)** glpK tree (77 sequences, 363 positions) inferred under the best-fitting LG+C60 model **b)** PlsC tree (74 sequences, 57 positions) inferred under the best-fitting LG+C50 model **c)** PlsY tree (60 sequences, 104 positions) inferred under the best-fitting LG+C50 model. FBC are Fibrobacteres, Bacteroidetes and Chlorobi. PVC are Planctomycetes, Verrucomicrobia and Chlamydiae. TACK are Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota. DPANN include Diaphoretrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota, as well as several other lineages. For full trees, see Supplementary Figures 7-9. For full unrooted trees see Supplementary Figures 20-23.

*Comparisons with outgroup rooting*

The most widely-used approach for rooting trees is to place the root on the branch leading to a pre-defined outgroup (Penny 1976), but this can be challenging for ancient genes for which closely-related outgroups are lacking (Gouy et al. 2015). However, several lipid biosynthesis genes are members of larger protein families whose other members, although distantly related, have conserved structures and related functions (Peretó et al. 2004). Recently, Yokobori et al. (2016) used other members of these families as outgroups to root trees for G1PDH, G3PDH (both GpsA and Glp) and GlpK. Their root inferences differed from ours in that they found that bacterial G1PDH sequences formed a monophyletic group that branched from within Archaea, suggesting more recent horizontal transfer from Archaea to Bacteria, as opposed to transfer from stem Archaea or vertical inheritance from LUCA. On the other hand, their analysis of Glp recovered Bacteria on one side of the root, and a clade of Bacteria and Archaea on the other. They interpreted this as evidence for the presence of Glp in LUCA, and therefore that LUCA would have had G3P phospholipids.

A potential concern when using distantly-related sequences to root a tree is that the long branch leading to the outgroup can induce errors in the in-group topologies due

410 to long branch attraction (LBA) artefacts. Further, single-matrix phylogenetic models,
411 such as those used by Yokobori et al. (2016), have been shown to be more susceptible
412 to artefacts of this type than the profile mixture models used here (Lartillot et al. 2007).
413 To investigate whether the differences in root inference between our analyses and
414 those of Yokobori et al. (2016) might be the result of LBA, we performed outgroup
415 rooting analysis on G1PDH, GpsA and Glp, augmenting our datasets with a
416 subsample of the outgroups used by Yokobori et al. and using the same models used
417 to infer the unrooted trees (LG+C60 in each case). The resulting trees (Supplementary
418 Figures 10-12) show different topologies when compared to the unrooted trees
419 (Supplementary Figures 16, 19-20). This suggests that the long branch outgroup may
420 be distorting the ingroup topology.
421
422 We also performed model testing in IQ-Tree and compared the fit of the chosen
423 models to the models used by Yokobori et al. (see Material and Methods below).
424 LG+C60 was selected for both G1PDH and Glp, while LG+C50 was selected for Gpsa
425 (Supplementary figure 24). The results of these analyses indicate that the empirical
426 profile mixture models which we have used here fits each of these alignments
427 significantly better than the single-matrix models of Yokobori et al. (Supplementary
428 Table 2). However, even analyses under the best-fitting available models show
429 distortion of the ingroup topology upon addition of the outgroup (Supplementary
430 Figures 10-12, 24), when compared to the unrooted topologies (Supplementary
431 Figures 16, 19-20). In each case, we found the root in a different place to those
432 recovered by Yokobori et al. In the G1PDH tree, we find Bacteria, specifically
433 Firmicutes to be most basal, rather the Crenarchaeota found by Yokobori. In the case
434 of GpsA, Yokobori et al. did not find compelling support for an origin in LUCA, but they
435 did recover one archaeal lineage (the Euryarchaeota) at the base of the in-group tree
436 with low (bootstrap 48) support. While our GpsA tree is also poorly resolved, we do
437 not find evidence to support the basal position of the archaeal lineages, and therefore
438 for the presence of GpsA in LUCA. For glp, which Yokobori et al. trace back to LUCA
439 due to the basal position of the archaeal sequences, the outgroup sequences did not
440 form a monophyletic group, and were instead distributed throughout the tree
441 (Supplementary Figure 11). Thus, analyses under the best-fitting available models did
442 not support the presence of bacterial lipid biosynthesis genes in LUCA. Further, the
443 distortion of the ingroup topologies suggests that these outgroups may not be suitable
444 for root inference, at least given current data and methods.
445
446
447 *Origin of eukaryotic lipid biosynthesis genes*
448
449 Phylogenetics and comparative genomics suggest that eukaryotes arose from a
450 symbiosis between an archaeal host cell and a bacterial endosymbiont that evolved
451 into the mitochondrion (reviewed, from a variety of perspectives, in Embley and Martin
452 2006; Martin et al. 2015; Eme et al. 2017; Roger et al. 2017). Specifically, genomic
453 and phylogenetic evidence indicates that the host lineage belonged to the
454 Asgardarchaeota superphylum, although the exact position of eukaryotes within
455 Asgardarchaeota is unclear (Spang et al. 2015; Zaremba-Niedzwiezka et al. 2017).
456 The origin of bacterial-type membranes in eukaryotes is therefore an important
457 evolutionary question that needs explanation. As noted above, multiple explanations
458 have been proposed for the origin of eukaryotic membrane lipids (Woese et al. 1990;
459 Kandler 1995; Lopez-Garcia and Moreira 2006; Baum and Baum 2014; Gould et al.

2015). An additional possibility, raised by our analyses, is that eukaryotes may have inherited their bacterial lipids vertically from the archaeal host cell.

Indeed, evidence from both our study and Villanueva et al. (2016) points to the presence of orthologues for bacterial lipid genes in asgardarchaeote lineages. Both our study and Viallanueva et al. find Glp, PlsC and PlsY orthologues in Lokiarchaeota. We additionally find a PlsC orthologue in Heimadallarchaeota, and PlsC and PlsY orthologues in Heimdallarchaeota and Thorarchaeota (Table 1).

To evaluate this hypothesis, we expanded our datasets for GpsA, Glp and PlsC with a representative set of eukaryotic homologues. The resulting trees are poorly resolved (Supplementary Figures 13-15), but do not support a specific relationship between the eukaryotic sequences and any archaeal lineages, and so do not provide any compelling support for an origin of eukaryotic lipids via the archaeal host cell.

*Conclusions*

Our phylogenetic analyses of lipid biosynthesis genes support two main conclusions about prokaryotic cell physiology and early cell evolution. First, there is clear evidence for extensive horizontal transfer of lipid genes between Archaea and Bacteria, from potentially very early to more recent evolutionary times. In at least one recognised case – the synthesis of archaeal lipids by the bacterium *B. subtilis* (Guldan et al. 2008; Guldan et al. 2011) – these horizontally acquired genes are involved in membrane lipid synthesis; by induction, they could have similar functions in many of the other organisms that possess them. Therefore, the capability to synthesise both types of membranes may be more widespread than has been appreciated hitherto. However, gene presence is not sufficient to establish membrane composition, as these genes might be involved in other cellular processes. As in *B. subtilis*, experiments would be needed to test these predictions in any particular case. Crucially, the evidence that these genes undergo horizontal transfer, both early in evolution and more recently, provides a potential mechanism for the remarkable diversity of membrane lipids, and especially ether lipids, in diverse environmental settings (Schouten et al. 2001). We also note that it is intriguing that bacterial lipids with archaeal features are particularly abundant in settings characterised by high archaeal abundances, including cold seeps, wetlands and geothermal settings (Schouten et al. 2013b), potentially providing ecological opportunity for gene transfer.

A second, and more tentative, result of our study relates to the antiquity of the canonical archaeal and bacterial pathways. Our analyses suggest that the enzymes for making G1P lipids were already present in the common ancestor of Archaea and Bacteria. Under the consensus view that the root of the tree of life lies between Bacteria and Archaea, this would imply that LUCA could have made archaeal type membranes. This finding is intriguing in light of previous work suggesting the presence of isoprenoids produced by the mevalonate pathway in LUCA (Lombard and Moreira 2011; Castelle and Banfield 2018). By contrast, the roots for the bacterial genes were weakly resolved within the bacterial domain. There is therefore no positive evidence from our trees to suggest that the bacterial pathway was present in LUCA, although we cannot exclude this possibility.

The consensus universal root between Bacteria and Archaea is supported by analyses of ancient gene duplications (Gogarten et al. 1989; Iwabe et al. 1989; Zhaxybayeva et al. 2005) and genome networks (Dagan et al. 2010), but some analyses have supported an alternative placement of the root within Bacteria (Cavalier-Smith 2006; Lake et al. 2009; Williams et al. 2015). Our trees do not exclude a within-Bacteria root, in which case LUCA would have possessed the bacterial pathway, and the archaeal pathway would have evolved along the archaeal stem, or in a common ancestor of Archaea and Firmicutes (Cavalier-Smith 2006; Lake et al. 2009).

If one membrane lipid pathway evolved before the other, this would imply that one of the two prokaryotic lineages changed its membrane lipid composition during early evolution. The evolutionary processes that drive such changes remain unclear, in part because we still do not fully understand the functional differences between modern archaeal and bacterial membranes. Compared to bacterial-type membranes, archaeal-type membranes maintain their physiochemical properties over a broader range of temperatures, and may be more robust to other environmental extremes (Van de Vossenberg et al. 1998; Koga 2012). If the archaeal pathway is older than the bacterial pathway, then that could reflect a LUCA adapted to such extreme settings. It is then intriguing to speculate on the evolutionary drivers for subsequent adoption of bacterial-type membranes, especially since the bacteria appear to be more successful than the archaea in terms of abundance and genetic diversity (Danovaro et al. 2016; Hug et al. 2016; Castelle and Banfield 2018). Moreover, an analogous change has happened at least once in evolutionary history, during the origin of eukaryotic cells (Martin et al. 2015).

It is possible that a transition to the bacterial type was driven by the lower energetic cost of making and repurposing fatty acid ester lipids, although we know of no published experimental data on these relative biosynthetic costs. Alternatively, the bacterial-type membrane lipids comprise a variety of fatty acyl moieties, varying in chain length, unsaturation, degree of branching and cyclisation, and these could impart a degree of flexibility and adaptability that provides a marginal benefit in dynamic mesophilic environments. If so, that advantage could translate to bacterial ether lipids that are also widespread in non-extreme settings and also characterised by a variety of alkyl forms (Pancost et al. 2001). Conversely, if bacterial-type membranes were ancestral, the transition to archaeal-type membranes could have been driven by adaptation to high environmental temperatures: ether bonds are more thermostable than esters (Van de Vossenberg et al. 1998; Koga 2012), and are also found in the membranes of thermophilic Bacteria (Kaur et al. 2015). In any case, the widespread occurrence of bacterial-type, archaeal-type and mixed-type membrane lipids in a range of environments, as well as the widespread occurrence of the associated biosynthetic pathways across both domains, suggests that except for high temperature and low pH settings, the advantages of either membrane type is marginal.

Our results indicate that inter-domain transfer of membrane lipid biosynthesis genes appears to be widespread, providing a potential mechanism for understanding the variety of lipids with mixed characteristics that occur in the environment. Unfortunately, very little is currently known about the stereochemical diversity of environmental lipids; we are aware of only one study (Weijers et al., 2006) that has investigated this for a class of lipids of mixed character, the brGDGTs, which exhibit bacterial-type stereochemistry. Our work suggests that stereochemical diversity, just like other

putative features of the lipid divide, should also be re-investigated. Overall, and taken together with evidence from natural and experimental settings for the stability of mixed membranes (Guldan et al. 2008; Guldan et al. 2011; Caforio et al. 2018), our analyses suggest that membrane lipid composition is not an immutable hallmark of cellular lineages but, like other features of prokaryote physiology (Jain et al. 1999) can change over time.

## Materials and Methods

### *Sequence selection*

For Archaea, we selected 43 archaeal genomes, sampled evenly across the archaeal tree. We took corresponding archaeal G1PDH, GGGPS and DGGPS amino acid sequences from the data set of Villanueva et al. (2016) and performed BLASTp searches the find these sequences in genomes not included in that dataset. For Bacteria, we selected 64 bacterial genomes, sampled so as to represent the known genomic diversity of bacterial phyla (Hug et al. 2016). We used G3PDH *gpsa*, G3PDH *glp* and GlpK sequences from Yokobori et al. (2016) and performed BLASTp searches to find those sequences in bacterial species not in their data set. For PlsC and PlsY, we took the corresponding sequences form Vilanueva et al. 2016, and performed BLASTp searches to find these sequences in the remaining genomes. For PlsB and PlsX, we searched for the respective terms in the gene database on the NCBI website, and upon finding well-verified occurrences, performed BLASTp searches to find the corresponding amino acid sequences in the remaining genomes. We then used BLASTp to look for bacterial orthologues of the archaeal enzymes and vice versa. We selected sequences that had an E-value of less the e-7 and at least 50% coverage. Accession numbers for sequences used are provided in Supplementary Table 3.

### *Phylogenetics*

The sequences were aligned in mafft (Katoh et al. 2002) using the --auto option and trimmed in BMGE (Criscuolo and Gribaldo 2010) using the BLOSUM30 model, which is most suitable for anciently-diverged genes. To construct gene trees form our amino acid sequences, we first selected the best-fitting substitution model for each gene according to its BIC score using the model selection tool in IQ-Tree (Nguyen et al. 2015). For all of the genes we analysed, the best-fitting model was a mixture model combining the LG exchangeability matrix (Le and Gascuel 2008) with site-specific composition profiles (the C40, C50 and C60 models (Lartillot and Philippe 2004; Le et al. 2008)) to accommodate across-site variation in the substitution process. LG+C60 was used for G1PDH, GpsA, Glp and GlpK. LG+50 was used for DDDGPS, PlsC and PlsY. LG+C40 was used for GGGPS. A discretised Gamma distribution (Yang 1994) with 4 rate categories was used to model across-site rate variation. The trees were run with their respective models in PhyloBayes (Lartillot and Philippe 2004, 2006; Lartillot et al. 2007); convergence was assessed using the bpcomp and tracecomp programs (maxdiff < 0.1; effective sample sizes > 100), as recommended by the authors.

609 The trees were rooted with a lognormal uncorrelated molecular clock, using the LG
610 model with a discretised Gamma distribution (Yang 1994) with 4 rate categories, and
611 a Yule tree prior (Stadler 2009; Hartmann et al. 2010) in BEAST (Drummond 2007;
612 Drummond et al. 2012). We also rooted the trees using minimal ancestral deviation
613 (MAD) rooting (Tria et al 2017). MAD rooting requires a fully-resolved, bifurcating tree;
614 since some parts of the consensus phylogenies were poorly resolved, we integrated
615 over this topological uncertainty by computing the optimum MAD root position for each
616 tree sampled during the MCMC analysis, and obtained marginal posterior probabilities
617 for these root positions using RootAnnotator (Calvignac-Spencer et al. 2014).
618
619 For G1PDH, GpsA and Glp, we also rooted the trees using a subsample of the
620 outgroup sequences used by Yokobori et al. (2016). The outgroups used were two 3-
621 dehydroquinate synthase (DHQS), five glucerol dehydrogenase (GDH) and five
622 alcohol dehydrogenase (ALDH) sequences for G1PDH; six hydroxyacyl-CoA
623 dehydrogenase (HACDH) and six UDP-glucose 6-dehydrogenase (UDPGDH)
624 sequences for GpsA; and 12 FAD-dependent oxidoreductase sequences for Glp. All
625 of three of these trees were inferred under the LG+C60 model to directly compare to
626 the unrooted trees. Tress were also inferred from best fit models selected in IQTree
627 (LG+C60 for G1PDH and Glp, and LG+C50 for GpsA). To construct trees for GpsA,
628 Glp and PlsC including eukaryotic sequences, we performed BLAST searches for
629 these sequences on 35 eukaryotic genomes from across the eukaryotic tree and
630 followed the above steps to infer trees. LG+C50 was the best-fitting model for all three
631 trees.
632
633 Eukaryotic orthologues of prokaryotic lipid biosynthesis genes were identified by
634 performing BLASTp searches on 35 eukaryotic genomes from across eukaryotic
635 diversity using *Homo sapiens* query as the sequence in each case, selecting
636 sequences with an E-value of e-7 or less, and at least 50% coverage. We then
637 performed model testing in IQTree and inferred trees in PhyloBayes using the selected
638 substitution model (LG+C50 for all three).
639
640 All sequences, alignments and trees referred to in this study can be obtained from
641 10.6084/m9.figshare.6210137.
642
643
644 **Acknowledgements**
645
646 GAC, RDP, and TAW conceived the project. GAC and TAW designed and performed
647 the analyses. GAC, RDP and TAW interpreted the results and wrote the manuscript.
648 GAC is supported by a Royal Society Research Grant to TAW. TAW is supported by
649 a Royal Society University Research Fellowship. We thank George S. Attard, Anja
650 Spang and T. Martin Embley for helpful feedback on the manuscript.
651
652
653 **References**
654 Adam PS, Borrel G, Gribaldo S. 2018. Evolutionary history of carbon monoxide
655 dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. Proc.
656 Natl. Acad. Sci.:201716667. doi:10.1073/pnas.1716667115.
657
658 Baum DA, Baum B. 2014. An inside-out origin for the eukaryotic cell. BMC Biol.

12:1–22. doi:10.1186/s12915-014-0076-2.

Bell SD, Jackson SP. 1998. Transcription and translation in Archaea: A mosaic of eukaryal and bacterial features. Trends Microbiol. 6:222–227. doi:10.1016/S0966-842X(98)01281-5.

Boucher Y, Kamekura M, Doolittle WF. 2004. Origins and evolution of isoprenoid lipid biosynthesis in archaea. Mol. Microbiol. 52:515–527. doi:10.1111/j.1365-2958.2004.03992.x.

Caforio A, Siliakus MF, Exterkate M, Jain S, Jumde VR. 2018. Converting Escherichia coli into an archaebacterium with a hybrid heterochiral membrane. PNAS 115:1–6. doi:10.1073/pnas.1721604115.

Calvignac-Spencer S, Schulze JM, Zickmann F, Renard BY. 2014. Clock Rooting Further Demonstrates that Guinea 2014 EBOV is a Member of the Zaïre Lineage. PLoS Curr. 6:1–8. doi:10.1371/currents.outbreaks.c0e035c86d721668a6ad7353f7f6fe86.

Castelle CJ, Banfield JF. 2018. Perspective Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. Cell 172:1181–1197. doi:10.1016/j.cell.2018.02.016.

Cavalier-Smith T. 2006. Rooting the tree of life by transition analyses. Biol. Direct 1:19. doi:10.1186/1745-6150-1-19.

Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM. 2008. The archaebacterial origin of eukaryotes. Proc. Natl. Acad. Sci. 105:20356–20361. doi:10.1073/pnas.0810647105.

Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol. Biol. 10. doi:10.1186/1471-2148-10-210.

Dagan T, Roettger M, Bryant D, Martin W. 2010. Genome networks root the tree of life between prokaryotic domains. Genome Biol. Evol. 2:379–392. doi:10.1093/gbe/evq025.

Danovaro R, Molari M, Corinaldesi C, Dell'Anno A. 2016. Macroecological drivers of archaea and bacteria in benthic deep-sea ecosystems. Sci. Adv. 2:1–12. doi:10.1126/sciadv.1500961.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:1–8. doi:10.1186/1471-2148-7-214.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29:1969–1973. doi:10.1093/molbev/mss075.

Embley TM, Martin W. 2006. Eukaryotic evolution, changes and challenges. Nature.

440:623–630. doi:10.1038/nature04546.

Eme L, Spang A, Lombard J, Stairs CW, Ettema TJG. 2017. Archaea and the origin of eukaryotes. Nat. Rev. Microbiol. 15:711–723. doi:10.1038/nrmicro.2017.133.

Fan Q, Relini A, Cassinadri D, Gambacorta A, Gliozzi A. 1995. Stability against temperature and external agents of vesicles composed of archaeal bolaform lipids and egg PC. BBA - Biomembr. 1240:83–88. doi:10.1016/0005-2736(95)00157-X.
Garcia-Vallvé S, Romeu a, Palau J. 2000. Horizontal gene transfer in bacterial and archaeal complete genomes. Genome Res. 10:1719–1725. doi:10.1101/gr.130000.particular.

Gattinger A, Schloter M, Munch JC. 2002. Phospholipid etherlipid and phospholipid fatty acid fingerprints in selected euryarchaeotal monocultures for taxonomic profiling. FEMS Microbiol. Lett. 213:133–139. doi:10.1016/S0378-1097(02)00794-2.

Gogarten JP, Rausch T, Bernasconi P, Kibak H, Taiz L. 1989. Molecular Evolution of H+-ATPases. I. Methanococcus and Sulfolobus are Monophyletic with Respect to Eukaryotes and Eubacteria. Zeitschrift fur Naturforsch. - Sect. C J. Biosci. 44:641–650. doi:10.1515/znc-1989-7-816.

Goldfine H. 2010. The appearance, disappearance and reappearance of plasmalogens in evolution. Prog. Lipid Res. 49:493–498. doi:10.1016/j.plipres.2010.07.003.

Gould SB, Garg SG, Martin WF. 2016. Bacterial Vesicle Secretion and the Evolutionary Origin of the Eukaryotic Endomembrane System. Trends Microbiol. 24:525–534. doi:10.1016/j.tim.2016.03.005.

Gouy R, Baurain D, Philippe H. 2015. Rooting the tree of life: the phylogenetic jury is still out. Philos. Trans. R. Soc. B Biol. Sci. 370:20140329. doi:10.1098/rstb.2014.0329.

Guldan H, Matysik FM, Bocola M, Sterner R, Babinger P. 2011. Functional assignment of an enzyme that catalyzes the synthesis of an archaea-type ether lipid in bacteria. Angew. Chemie - Int. Ed. 50:8188–8191. doi:10.1002/anie.201101832.

Guldan H, Sterner R, Babinger P. 2008. Identification and Characterization of a Bacterial Glycerol-1-phosphate Dehydrogenase: $Ni^{2+}$ -Dependent AraM from *Bacillus subtilis*. Biochemistry. 47:7376–7384. doi:10.1021/bi8005779.

Guy L, Ettema TJG. 2011. The archaeal “TACK” superphylum and the origin of eukaryotes. Trends Microbiol. 19:580–587. doi:10.1016/j.tim.2011.09.002.

Hartmann K, Wong D, Stadler T. 2010. Sampling trees from evolutionary models. Syst. Biol. 59:465–476. doi:10.1093/sysbio/syq026.

Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Kotaro I, et al. 2016. A new view of the tree and life’s diversity. Nautre. doi:10.1038/nmicrobiol.2016.48.

759
760 Iwabe N, Kuma K-I, Hasegawa M, Osawa S, Source TM, Hasegawat M, Osawat S,
761 Miyata T. 1989. Evolutionary Relationship of Archaebacteria, Eubacteria, and
762 Eukaryotes Inferred from Phylogenetic Trees of Duplicated Genes. Proc. Natl. Acad.
763 Sci. United States Am. Evol. 86:9355–9359. doi:10.1073/pnas.86.23.9355.
764
765 Jain R, Rivera MC, Lake JA. 1999. Horizontal gene transfer among genomes: The
766 complexity hypothesis. Proc. Natl. Acad. Sci. 96:3801–3806.
767 doi:10.1073/pnas.96.7.3801.
768
769 Kandler O. 1995. Cell wall biochemistry in Archaea and its phylogenetic implications.
770 J. Biol. Phys. 20:165–169. doi:10.1007/BF00700433.
771
772 Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid
773 multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res.
774 30:3059–3066. doi:10.1093/nar/gkf436.
775
776 Kaur G, Mountain BW, Stott MB, Hopmans EC, Pancost RD. 2015. Temperature and
777 pH control on lipid composition of silica sinters from diverse hot springs in the Taupo
778 Volcanic Zone, New Zealand. Extremophiles. 19:327–44. doi:10.1007/s00792-014-
779 0719-9.
780
781 Kelman LM, Kelman Z. 2014. Archaeal DNA Replication. Annu. Rev. Genet. 48:71–
782 97. doi:10.1146/annurev-genet-120213-092148.
783
784 Koga Y. 2011. Early evolution of membrane lipids: How did the lipid divide occur? J.
785 Mol. Evol. 72:274–282. doi:10.1007/s00239-011-9428-5.
786
787 Koga Y. 2012. Thermal adaptation of the archaeal and bacterial lipid membranes.
788 Archaea. 2012. doi:10.1155/2012/789652.
789
790 Koga Y, Kyuragi T, Nishihara M, Sone N. 1998. Did archaeal and bacterial cells arise
791 independently from noncellular precursors? A hypothesis stating that the advent of
792 membrane phospholipid with enantiomeric glycerophosphate backbones caused the
793 separation of the two lines of descent. J. Mol. Evol. 46:54–63.
794 doi:10.1007/PL00006283.
795
796 Lake J a, Skophammer RG, Herbold CW, Servin J a. 2009. Genome beginnings:
797 rooting the tree of life. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 364:2177–2185.
798 doi:10.1098/rstb.2009.0035.
799
800 Lartillot N, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction
801 artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol. Biol.
802 7:1–14. doi:10.1186/1471-2148-7-S1-S4.
803
804 Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site
805 heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. 21:1095–
806 1109. doi:10.1093/molbev/msh112.
807
808 Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. Mol.

809 Biol. Evol. 25:1307–1320. doi:10.1093/molbev/msn067.
810
811 Le SQ, Lartillot N, Gascuel O. 2008. Phylogenetic mixture models for proteins.
812 Philos. Trans. R. Soc. B Biol. Sci. 363:3965–3976. doi:10.1098/rstb.2008.0180.
813
814 Lombard J, López-García P, Moreira D. 2012. The early evolution of lipid
815 membranes and the three domains of life. Nat. Rev. Microbiol. 10:507–515.
816 doi:10.1038/nrmicro2815.
817
818 Lombard J, Moreira D. 2011. Origins and early evolution of the mevalonate pathway
819 of isoprenoid biosynthesis in the three domains of life. Mol. Biol. Evol. 28:87–99.
820 doi:10.1093/molbev/msq177.
821
822 López-García P, Moreira D. 2006. Selective forces for the origin of the eukaryotic
823 nucleus. BioEssays. 28:525–533. doi:10.1002/bies.20413.
824
825 Martin W, Muller M. 1998. The hydrogen hypothesis for the origin of the first
826 eukaryote. Nature. 392:37–41.
827
828 Martin W, Russell MJ. 2003. On the origins of cells: a hypothesis for the evolutionary
829 transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from
830 prokaryotes to nucleated cells. Philos Trans R Soc L. B Biol Sci 358:55–59.
831 doi:10.1098/rstb.2002.1183.
832
833 Martin WF, Garg S, Zimorski V. 2015. Endosymbiotic theories for eukaryote origin.
834 Philos. Trans. R. Soc. B Biol. Sci. 370:20140330. doi:10.1098/rstb.2014.0330.
835
836 Nemoto N, Oshima T, Yamagishi A. 2003. Purification and characterization of
837 geranylgeranylglyceryl phosphate synthase from a thermoacidophilic archaeon,
838 Thermoplasma acidophilum. J. Biochem. 133:651–657. doi:10.1093/jb/mvg083.
839
840 Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and
841 effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol.
842 Biol. Evol. 32:268–274. doi:10.1093/molbev/msu300.
843
844 Pancost RD, Bouloubassi I, Aloisi G, Sinninghe Damsté JS. 2001. Three series of
845 non-isoprenoidal dialkyl glycerol diethers in cold-seep carbonate crusts. Org.
846 Geochem. 32:695–707. doi:10.1016/S0146-6380(01)00015-8.
847
848 Penny D. 1976. Criteria for optimising phylogenetic trees and the problem of
849 determining the root of a tree. J. Mol. Evol. 8:95–116. doi:10.1007/BF01739097.
850
851 Peretó J, López-García P, Moreira D. 2004. Ancestral lipid biosynthesis and early
852 membrane evolution. Trends Biochem. Sci. 29:469–477.
853 doi:10.1016/j.tibs.2004.07.002.
854
855 Peterhoff D, Beer B, Rajendran C, Kumpula EP, Kapetaniou E, Guldan H, Wierenga
856 RK, Sterner R, Babinger P. 2014. A comprehensive analysis of the
857 geranylgeranylglyceryl phosphate synthase enzyme family identifies novel members
858 and reveals mechanisms of substrate specificity and quaternary structure

organization. Mol. Microbiol. 92:885–899. doi:10.1111/mmi.12596.

Reeve JN, Sandman K, Daniels CJ. 1997. Archaeal Histones, Nucleosomes, and Transcription Initiation. 89:999–1002.

Roger AJ, Muñoz-Gómez SA, Kamikawa R. 2017. The Origin and Diversification of Mitochondria. Curr. Biol. 27:R1177–R1192. doi:10.1016/j.cub.2017.09.015.

Schouten S, Hopmans EC, Sinninghe Damsté JS. 2013a. The organic geochemistry of glycerol dialkyl glycerol tetraether lipids: A review. Org. Geochem. 54:19–61. doi:10.1016/j.orggeochem.2012.09.006.

Schouten S, Hopmans EC, Sinninghe Damsté JS. 2013b. The organic geochemistry of glycerol dialkyl glycerol tetraether lipids: A review. Org. Geochem. 54:19–61. doi:10.1016/j.orggeochem.2012.09.006.

Schouten S, Wakeham SG, Sinninghe Damsté JS. 2001. Evidence for anaerobic methane oxidation by archea in euxinic waters of the Black Sea. Org. Geochem. 32:1277–1281.

Shimada H, Yamagishi A. 2011. Stability of heterochiral hybrid membrane made of bacterial sn -G3P lipids and archaeal sn -G1P lipids. Biochemistry. 50:4114–4120. doi:10.1021/bi200172d.

Sinninghe Damsté JS, Rijpstra WIC, Hopmans EC, Foesel BU, Wüst PK, Overmann J, Tank M, Bryant DA, Dunfield PF, Houghton K, et al. 2014. Ether- and ester-bound iso-diabolic acid and other lipids in members of Acidobacteria subdivision 4. Appl. Environ. Microbiol. 80:5207–5218. doi:10.1128/AEM.01066-14.

Sinninghe Damsté JS, Rijpstra WIC, Hopmans EC, Schouten S, Balk M, Stams AJM. 2007. Structural characterization of diabolic acid-based tetraester, tetraether and mixed ether/ester, membrane-spanning lipids of bacteria from the order Thermotogales. Arch. Microbiol. 188:629–641. doi:10.1007/s00203-007-0284-z.

Sinninghe Damsté JS, Schouten S, Hopmans EC, van Duin ACT, Geenevasen JAJ. 2002. Crenarchaeol:the characteristic core glycerol dibiphytanyl glycerol tetraether membrane lipid of cosmopolitan pelagic crenarchaeota. J. Lipid Res. 43:1641–1651. doi:10.1194/jlr.M200148-JLR200.

Sinninghe Damsté JS, Strous M, Rijpstra WIC, Hopmans EC, Geenevasen JAJ, van Duin ACT, van Niftrik LA, Jetten MSM. 2002. Linearly concatenated cyclobutane lipids form a dense bacterial membrane. Nature. 419:708–712. doi:10.1038/nature01067.

Sojo V, Pomiankowski A, Lane N. 2014. A Bioenergetic Basis for Membrane Divergence in Archaea and Bacteria. PLoS Biol. 12. doi:10.1371/journal.pbio.1001926.

Sousa FL, Martin WF. 2014. Biochemical fossils of the ancient transition from

geoenergetics to bioenergetics in prokaryotic one carbon compound metabolism. Biochim. Biophys. Acta - Bioenerg. 1837:964–981. doi:10.1016/j.bbabio.2014.02.001.

Sousa FL, Thiergart T, Landan G, Nelson-Sathi S, Pereira I a C, Allen JF, Lane N, Martin WF. 2013. Early bioenergetic evolution. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 368:20130088. doi:10.1098/rstb.2013.0088.

Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, Eijk R Van, Schleper C, Guy L, Ettema TJG. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature. doi:10.1038/nature14447.

Stadler T. 2009. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. J. Theor. Biol. 261:58–66. doi:10.1016/j.jtbi.2009.07.018.

Tan HH, Makino A, Sudesh K, Greimel P, Kobayashi T. 2012. Spectroscopic evidence for the unusual stereochemical configuration of an endosome-specific lipid. Angew. Chemie - Int. Ed. 51:533–535. doi:10.1002/anie.201106470.

Tria FDK, Landan G, Dagan T. 2017. Phylogenetic rooting using minimal ancestor deviation. Nat. Ecol. Evol. 1:193. doi:10.1038/s41559-017-0193.

Villanueva L, Schouten S, Sinninghe Damsté JS. 2016. Phylogenomic analysis of lipid biosynthetic genes of Archaea shed light on the “lipid divide.” Environ. Microbiol. 19:54–69. doi:10.1111/1462-2920.13361.

Van de Vossenberg JLCM, Driessen AJM, Konings WN. 1998. The essence of being extremophilic: The role of the unique archaeal membrane lipids. Extremophiles. 2:163–170. doi:10.1007/s007920050056.

Wächtershäuser G. 1988. Pyrite Formation, the First Energy Source for Life: a Hypothesis. Syst. Appl. Microbiol. 10:207–210. doi:10.1016/S0723-2020(88)80001-8.

Wächtershäuser G. 1992. Groundworks for an evolutionary biochemistry: The iron-sulphur world. Prog. Biophys. Mol. Biol. 58:85–201. doi:10.1016/0079-6107(92)90022-X.

Wächtershäuser G. 2003. From pre-cells to Eukarya--a tale of two lipids. Mol. Microbiol. 47:13–22. doi:3267 [pii].

Weijers JWH, Schouten S, Hopmans EC, Geenevasen JAJ, David ORP, Coleman JM, Pancost RD, Sinninghe Damsté JS. 2006. Membrane lipids of mesophilic anaerobic bacteria thriving in peats have typical archaeal traits. Environ. Microbiol. 8:648–657. doi:10.1111/j.1462-2920.2005.00941.x.

Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-sathi S, Martin WF. 2016. The physiology and habitat of the last universal common ancestor. Nat. Microbiol.1:1–8. doi:10.1038/nmicrobiol.2016.116.

Wilkinson M, McInerney JO, Hirt RP, Foster PG, Embley TM. 2007. Of clades and clans: terms for phylogenetic relationships in unrooted trees. Trends Ecol. Evol. 22:114–115. doi:10.1016/j.tree.2007.01.002.

Williams T a, Foster PG, Cox CJ, Embley TM. 2013. An archaeal origin of eukaryotes supports only two primary domains of life. Nature. 504:231–6. doi:10.1038/nature12779.

Williams TA, Embley TM. 2015. Changing ideas about eukaryotic origins. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 370:20140318. doi:10.1098/rstb.2014.0318.
Williams TA, Heaps SE, Cherlin S, Nye TMW, Boys RJ, Embley TM. 2015. New substitution models for rooting phylogenetic trees. Philos. Trans. R. Soc. B Biol. Sci. 370:20140336. doi:10.1098/rstb.2014.0336.

Williams TA, Szöllősi GJ, Spang A, Foster PG, Heaps SE, Boussau B, Ettema TJG, Embley TM. 2017. Integrative modeling of gene and genome evolution roots the archaeal tree of life. Proc. Natl. Acad. Sci.:201618463. doi:10.1073/pnas.1618463114.

Woese CR, Kandler O, Wheelis ML. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. 87:4576–4579. doi:10.1073/pnas.87.12.4576.

Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. J. Mol. Evol. 39:306–314. doi:10.1007/BF00160154.

Yokobori S, Nakajima Y, Akanuma S, Yamagishi A. 2016. Birth of Archaeal Cells: Molecular Phylogenetic Analyses of G1P Dehydrogenase, G3P Dehydrogenases, and Glycerol Kinase Suggest Derived Features of Archaeal Membranes Having G1P Polar Lipids. Archaea. 2016:1–16. doi:10.1155/2016/1802675.

Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, Anantharaman K, Starnawski P, Kjeldsen KU, Stott MB, Vancaester E, et al. 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nat. Publ. Gr. doi:10.1038/nature21031.

Zhaxybayeva O, Lapierre P, Gogarten JP. 2005. Ancient gene duplications and the root(s) of the tree of life. Protoplasma. 227:53–64. doi:10.1007/s00709-005-0135-1.