Phylogenomic analysis of lipid biosynthetic genes of Archaea shed light on the 'lipid divide'

Laura Villanueva[1*], Stefan Schouten[1,2], and Jaap S. Sinninghe Damsté[1,2]

[1]NIOZ, Royal Netherlands Institute for Sea Research. Department of Marine Microbiology and Biogeochemistry, and Utrecht University, P.O. Box 59, NL-1790 AB Den Burg, Texel, The Netherlands.

[2]Utrecht University, Faculty of Geosciences, P.O. Box 80.021, 3508 TA Utrecht, The Netherlands

*Correspondence to L.V. e-mail: laura.villanueva@nioz.nl

Resubmitted to Environmental Microbiology

Runing title: Archaeal lipid biosynthesis and the lipid divide

Keywords: 'lipid divide', Archaea, membrane lipids, glycerol stereochemistry, fatty acid synthesis, Lokiarchaeum, uncultured marine euryarchaeota group II and III.

Summary

The lipid membrane is one of the most characteristic traits distinguishing the three domains of life. Membrane lipids of Bacteria and Eukarya are composed of fatty acids linked to glycerol-3-phosphate (G3P) via ester bonds, while those of Archaea possess isoprene-based alkyl chains linked by ether linkages to glycerol-1-phosphate (G1P), resulting in the opposite stereochemistry of the glycerol phosphate backbone. This 'lipid divide' has raised questions on the evolution of microbial life since eukaryotes are thought to have evolved from the Archaea, requiring a radical change in membrane composition. Here, we searched for homologs of enzymes involved in membrane lipid and fatty acid synthesis in a wide variety of archaeal genomes and performed phylogenomic analyses. We found that two uncultured archaeal groups, i.e. marine euryarchaeota group II/III and 'Lokiarchaeota', recently discovered descendants of the archaeal ancestor leading to eukaryotes, lack the gene to synthesize G1P and, consequently, the capacity to synthesize archaeal membrane lipids. However, our analyses reveal their genetic capacity to synthesize G3P-based 'chimeric lipids' with either two ether-bound isoprenoidal chains or with an ester-bound fatty acid instead of an ether-bound isoprenoid. These archaea may reflect the 'archaea-to-eukaryote' membrane transition stage which have led to the current 'lipid divide'.

Introduction

Membrane lipids are essential building blocks for the cell since membranes define the 'inside' and the 'outside' of the cell. Membranes are involved in many biological processes such as establishing and maintaining trans-membrane gradients, compartmentalizing biochemical reactions into distinct functional domains, controlling transport into and out of cells, and inter- and intra-cellular communication. The membrane lipid composition is also one of the most remarkable traits distinguishing the three domains of life, Archaea, Bacteria, and Eukarya (Woese and Fox, 1977). Membrane lipids of Bacteria and Eukarya share a large number of structural similarities as they are typically composed of two fatty acid chains that are linked to a glycerol moiety via ester bonds and are organized in a bilayer structure (Lombard *et al*., 2012a). On the other hand, membrane lipids of Archaea are characterized by ether linkages between the glycerol moiety and isoprene-based alkyl chains in either a bilayer or monolayer (Koga and Morii, 2007; Lombard *et al*., 2012a). These traits are not fully exclusive to these groups since membrane-spanning lipids have also been reported in some members of the Bacteria (Sinninghe Damsté *et al*., 2002, 2007; Weijers *et al*., 2006), and fatty acids in some archaeal species (Gattinger *et al*., 2002). An exclusive distinction in the structures of membrane lipids of archaea and bacteria/eukaryotes is the opposite stereochemistry of the glycerol phosphate backbone, being *sn*-glycerol-1-phosphate (G1P) in archaea, and *sn*-glycerol-3-phosphate (G3P) in bacteria and eukaryotes (Kates, 1993). The biosynthesis of G3P and G1P is catalyzed by two entirely different enzymes (i.e. glycerol-1- and glycerol-3-phosphate dehydrogenase, G1PDH, and G3PDH, respectively) that, based on differences in the catalytic reaction and protein sequence (Koga *et al*., 2003; Han *et al*., 2005), are not evolutionary related (Koga *et al*., 1998). This differentiation of lipid structures between Archaea, on the one hand, and Bacteria and Eukarya, on the other, has been coined as the 'the lipid divide'.

This 'lipid divide' has posed some fundamental questions on microbial evolution. Since Archaea and Bacteria are believed to stem from a common ancestor (the cenancestor or last universal common ancestor, LUCA), their completely different membrane lipid structures represent a conundrum. Koga *et al*. (Koga *et al*., 1998) proposed that the cenancestor lacked a membrane and

that the specific archaeal and bacterial membrane lipid biosynthetic pathways emerged later and independently in the lineages leading to Archaea and Bacteria. Martin & Russell (2003) hypothesized that the cenancestor had mineral monosulfide compartments instead of lipids. Wächtershäuser (2003) suggested that the cenancestor had a lipid heterochiral membrane containing both stereochemical forms of the glycerol phosphate backbone, which progressively diverged into a more stable homochiral membrane leading to the differentiation between archaea and bacterial membranes. However, experiments with liposomes containing both archaeal and bacterial membrane lipids showed that heterochiral membranes are also stable (Fan *et al.*, 1995; Shimada *et a*l., 2011), suggesting that there exists no evolutionary pressure to select for organisms with a homochiral membrane. A recent study by Sojo *et al*. (2014) based on modeling of membrane bioenergetics suggested that LUCA did not have membranes with glycerol phosphate headgroups, which would have reduced proton permeability, but rather a lipid bilayer composed of both fatty acids and isoprenes, and that modern membranes in Bacteria and Archaea arose later and independently.

Another conundrum is the similarity of membrane lipids of the Eukarya with those of Bacteria rather than with those of the Archaea, which are believed to be the predecessors of the Eukarya (Pereto *et al*., 2004). According to the classical Woesian three-domain phylogeny, the last common ancestor of archaea and eukaryotes would have had an archaeal membrane that was later replaced by a bacterial-like membrane in eukaryotes, or alternatively that an ancestral mixed membrane with G1P- and G3P-based membrane lipids evolved to an archaeal membrane in archaea and to a bacterial-like membrane in eukaryotes. However, both options are difficult to reconcile as they would involve an intensive horizontal gene transfer of the genes required, while the mixed membrane model would imply that bacterial-like membranes evolved twice from the cenancestor in bacteria and in eukaryote. Currently, the most accepted early life evolutionary theory considers Archaea and Bacteria as primary branches derived directly from the cenancestor, while Eukarya would have evolved secondarily as a chimeric organism derived from the endosymbiosis of one bacterium (the ancestor of mitochondria) within a host cell (Gray and Doolittle, 1982; Golding and

Gupta, 1995; among others). However, the origin of the host cell is still under debate. In any case, most models of the origin of eukaryotic cells require a transition from an archaeal-like membrane to a bacterial membrane, including a reversal of the glycerol stereochemistry of membrane lipids Nonetheless, no evidence for this kind of transition has ever been found in bacteria or archaea, casting some doubts on this mechanism.

The growing availability of genomes could shed light on the evolutionary processes leading to the 'lipid divide'. Recent studies based on environmental metagenomics have defined several new archaeal lineages (Castelle *et al.*, 2015). Currently, the domain Archaea is represented by two superphyla (DPANN and TACK) and the phylum Euryarchaeota (Guy and Ettema, 2011; Fig. S1). The TACK-superphylum is comprised of Thaumarchaeota, Aigarchaeota, Crenarchaeota, Korarchaeota, and some other phyla (e.g. Guy and Ettema, 2011; Williams *et al*., 2012; Martijn and Ettema, 2013). Some phylogenomic studies have provided evidence that the archaeal 'ancestor' of the eukaryotic cell emerged from the TACK superphylum (Guy and Ettema, 2011). Furthermore, a recent study suggests that the novel candidate archaeal phylum 'Lokiarchaeota' (Deep-Sea Archaeal Group/Marine Benthic Group B, DSAG/MBG-B; Spang *et al*., 2015; Fig. S1), a deep branching clade of the TACK superphylum, forms a monophyletic group with the eukaryotes. Indeed, the 'Lokiarchaeum' composite genome codes a remarkable number of eukaryotic signature proteins, supporting the hypothesis that the eukaryotic cell evolved from an archaeal ancestor of this group (Spang *et al*., 2015).

The recent discovery of Lokiarchaeum, which potentially shares a common ancestor with eukaryotes, prompted us to re-examine the 'lipid divide' conundrum. We investigated two key aspects of the lipid divide: the specific stereoconfiguration of archaeal lipids and the capacity for fatty acid synthesis in archaea. We searched for homologs of genes encoding for enzymes involved in membrane lipid and fatty acid biosynthetic pathways in archaeal genomes and performed phylogenomic analyses with the annotated homologs. The results reveal differences in the lipid biosynthetic pathway, especially concerning the stereochemistry of the glycerol phosphate

backbone, in certain uncultured archaeal groups at key evolutionary phylogenetic positions with substantial implications for our understanding of the 'lipid divide'.

Results and Discussion

Enzymes involved in the glycerol phosphate stereospecific biosynthesis in Archaea

The stereoconfiguration of archaeal lipids is established by the enzyme G1PDH. This enzyme is thought to be restricted to archaea (e.g. Pereto *et al*., 2004; Koga and Morii, 2007; Matsumi *et al*., 2011) although a G1PDH homolog (AraM) has been found in *Bacillus* sp., and some related bacterial species (Guldan *et al*., 2008). Our survey of archaeal genomes revealed that the gene coding for G1PDH (*egs*A, Fig.1) is present in almost all examined archaea but, interestingly, is absent in marine group II euryarchaeota (MGII; Iverson *et al*., 2012), in fosmid sequences of the MGII and marine group III euryarchaeota (MGIII) (Deschamps *et al.,* 2014), 'Lokiarchaeum' (Spang *et al*., 2015), and all examined species of the DPANN superphylum (Castelle *et al*., 2015) (Table 1, Table S1). This suggests that these uncultivated archaea may not have the ability to synthesize the G1P backbone of archaeal lipids. For the members of the DPANN this is not surprising because they have simplified genomes of reduced size and are thought to rely on host cells or cell debris for the synthesis of their lipids (Waters *et al*., 2003; Jahn *et al*., 2004). In this respect, it is notable that in some of the DPANN genomes some homologs of enzymes involved in the archaeal membrane lipid biosynthesis are present, although they lack the MVK gene encoding for mevalonate kinase (Table 1; Table S1), an essential enzyme for isoprenoid biosynthesis. This situation may represent an intermediate stage of progressively losing those genes. However, truly exceptional is the lack of G1PDH in MGII and MGIII euryarchaeota and in 'Lokiarchaeum', as we found that genomes of these groups of archaea still harbor all the other known genes coding for the enzymes of the archaeal lipid biosynthetic pathway (i.e. geranylgeranylglyceryl phosphate synthase, GGGP; digeranylgeranylglyceryl phosphate synthase, DGGGP; geranyl reductase, GR, among others; Table 1; Table S1, S2). It should be noted that the current genome assembly of 'Lokiarchaeum' (*Lokiarchaeaum* sp. GC14_75), is 92% complete (Spang *et al*., 2015), and thus

there is a small chance that the gene coding for G1PDH would be located in the unsequenced part. However, we also did not detect any homologs of the G1PDH coding gene in the larger metagenome dataset (LCGC14AMP, 56.6 Gbp) reported in the same study (Spang *et al*., 2015). The lack of G1PDH homologs in the fosmid sequences of MGIII (Deschamps *et al.,* 2014) could potentially be due to their lack of completeness, however this can be ruled out for the genome of MGII reported by Iverson *et al*., 2012, which is closed. Our analysis strongly suggests that MGII and MGIII euryarchaeota and 'Lokiarchaeum' are capable of synthesizing isoprenoid-based ether lipids for their membranes but not with G1P as the glycerol building block.

G3PDH (encoded by the *gps* gene) catalyzes the conversion of dihydroxyacetone phosphate (DHAP) into G3P and this enzyme is responsible in Bacteria and Eukarya for the stereochemistry of the glycerol units of their membrane lipids. *Gps*-coded G3PDH homologs were previously detected in the euryarchaeota *Archaeoglobus fulgidus* and *Methanothermobacter themoautotrophicum,* in addition to G1PDH, but their metabolic function is unknown (Pereto *et al*., 2004). Some (heterotrophic) archaea have been reported to synthesize a 'G3PDH' enzyme encoded by the *glp* gene (Pereto *et al*., 2004; Koga and Morii, 2007) but this enzyme catalyzes the conversion of G3P into DHAP, the reverse of the reaction catalyzed by *gps*-coded G3PDH (Fig. 1). It is believed that this enables heterotrophic archaea to feed glycerol into the glycolysis pathway (Nishihara *et al*., 1999), as it is also observed for bacteria able to metabolize glycerol. These heterotrophic archaea still biosynthesize membrane lipids with the archaeal stereochemistry as they also harbor G1PDH (Fig. 1).

In our survey of archaeal genomes, *gps*-encoded G3PDH homologs were detected in MGII/III euaryarchaeota, and in some archaea of the orders Archaeoglobales and Methanobacteriales (*Methanobrevibacter* sp.), as well as in species of the DPANN superphylum (AR9 and AR11 genomes of the Woesearchaeota), but not in 'Lokiarchaeum' (Table 1; Table S1). Homologs of *glp*-encoded G3PDH were found in several archaeal genomes including members of the Thermococcales, Archaeoglobales, Halobacteriales, Thermoplasmatales, Korarchaeota, some Crenarchaeota genomes, as well as in the 'Lokiarchaeum' genome, which contains three putative

150 homologs of the enzyme (Table 1; Table S1, S2). To infer the evolutionary history of the annotated
151 archaeal G3PDH homologs, we constructed a phylogenetic tree including the archaeal *glp*- and *gps*-
152 coded G3PDH homologs, as well as bacterial and eukaryotic homologs (Fig. 2). The *glp*-coded
153 G3PDH archaeal homologs were mainly grouped in two main clusters with cluster 1 including
154 homologs of the Thermococcales, Thermoproteales, Desulfurococcales, Archaeoglobales, among
155 others, and cluster 2 comprising homologs of the Halobacteriales (Fig. 2A). Interestingly, the three
156 putative *glp*-coded G3PDH detected in the 'Lokiarchaeum' genome were closely related to bacterial
157 and eukaryotic G3PDH homologs (Fig. 2A). In addition, the *gps*-coded G3PDH archaeal homologs
158 do not form a monophyletic group (Fig. 2B), which may indicate horizontal gene transfer (HGT)
159 from bacteria to archaea in different independent events (Pereto *et al*., 2004).
160 G3P is not only formed from DHAP by *gps*-coded G3PDH but also by phosphorylation of
161 glycerol catalyzed by glycerol kinase, encoded by the *glp*K gene (Fig. 1). We detected homologs of
162 the *glp*K gene in genomes of the euryarchaeota Thermococcales, Archaeoglobales, Halobacteriales,
163 and Thermoplasmatales, as well as in the genomes of the *Aciduliprofundum* and MGII/III groups,
164 and in genomes of the Korarchaeota and Crenarchaeota phyla (Table 1). Two putative homologs of
165 the gene encoding for the glycerol kinase were detected in the 'Lokiarchaeum' genome (Table 1;
166 Table S1, S2). The two putative homologs of 'Lokiarchaeum' annotated as glycerol kinases display
167 a XylB pentulose or hexulose kinase region. In order to confirm the identity of these homologs, we
168 constructed a phylogenetic tree including the glycerol kinase proteins previously described in
169 archaeal genomes and carbohydrate kinase proteins closely related to the annotated 'Lokiarchaeum'
170 *glp*K (Fig. 3). The 'Lokiarchaeum' *glp*K homologs were closely related to carbohydrate kinases of
171 the euryarchaeon *Archaeoglobus fulgidus* and also to xylulose kinases of Bacteria, while the
172 glycerol kinases of archaeal genomes were grouped in another cluster (Fig. 3). Considering this
173 analysis, we cannot confirm the identity of the putative *glp*K coding genes annotated in the
174 'Lokiarchaeum' genomes as true glycerol kinases based on their divergence with previously
175 characterized archaeal glycerol kinases.

G3P can also be synthesized by degradation of glycerophosphodiesters by a glycerophosphodiester phosphodiesterase (GDPD) producing the corresponding alcohols and G3P (Larson *et al*., 1983; Fig. 1). Glycerophosphodiesters are enzymatically produced by phospholipases A_1 and A_2 from membrane phospholipids (e.g. Istivan *et al*., 2006). Glycerophosphodiester phosphodiesterase activities have been characterized in bacteria as well as in eukaryotes (Tommassen *et al*., 1991; Fisher *et al*., 2005; van der Rest *et al*., 2002), and genomic analyses have revealed a wide distribution of this protein family from bacteria and Archaea to metazoans, plants, and fungi (Santelli *et al*., 2004). In the bacterium *Escherichia coli*, the transformation of glycerophosphodiesters into G3P is thought to be catalyzed by two homologous enzymes, a periplasmic GDPD GlpQ, and a cytosolic GDPD UgpQ with a broad substrate specificity toward various glycerophosphodiesters (e.g. Tommassen *et al*., 1991). Here we detected archaeal UgpQ homologs in genomes of the Crenarchaeota, and in euryarchaeotal genomes of the Methanobacteriales, Thermoccoccales, Methanomicrobiales, Halobacteriales and Thermoplasmatales (Table 1; Table S1). Two putative UgpQ GDPD homologs were also detected in two genomes of the uncultured marine euryarchaeota group II/III, which were in turn closely related to putative UgpQ GDPD of the euryarchaeota Halobacteriales (Fig. 4). In addition, two putative homologs of UgpQ GDPD were detected in the 'Lokiarchaeum' genome (Table 1; Table S1-S2), which were closely related to another putative UgpQ GDPD annotated in the DPANN Woesearchaeota genome GW2011_AR3 as well as UgpQ detected in bacterial genomes of the Thermotogae (Fig. 4). The detection of two putative GDPD UgpQ in the 'Lokiarchaeum' genome opens the possibility of the formation of a G3P backbone by degradation of glycerophosphodiesters, as indicated in Fig. 1.

Putting this genetic evidence together, the data indicate that the unique archaeal groups that lack the gene encoding for G1PDH, i.e. the MGII/III euryarchaeota and 'Lokiarchaeum', harbor homologs of genes involved in the synthesis of the G3P backbone in Bacteria and Eukarya, either through the catalysis of DHAP to G3P in the case of MGII/III euryarchaeota (G3PDH encoded by *gps*), or in 'Lokiarchaeum' by degradation of glycerophosphodiesters by glycerophosphodiester

phosphodiesterase (UgpQ GDPD) or by metabolism of glycerol into G3P via glycerol kinase (encoded by the *glp*K gene). Consequently, it is tempting to speculate that both MGII and MGIII euryarchaeota and 'Lokiarchaeum' synthesize archaeal ether-linked membrane lipids, since they have the complete set of the genes coding for the enzymes of the archaeal lipid biosynthetic pathway (Table 1; Table S1, S2), but with the G3P stereochemistry typical for Bacteria and Eukarya, as they lack genes for G1PDH but possess genes encoding for enzymes for G3P formation. However, the synthesis of these kinds of lipids would be only possible if enzymes similar to GGGP and DGGGP synthases would catalyze the formation of ether bonds between isoprenoid chains and the G3P backbone as specified in the hypothetical pathway in Fig. 1.

Enzymes involved in the fatty acid biosynthetic pathway in Archaea

In bacterial fatty acid biosynthesis, the acetyl-CoA carboxylase (ACC) converts acetyl-coenzyme A (CoA) into malonyl-CoA (Fig. 5). Secondly, the peptide cofactor acyl carrier protein (ACP) has to be activated by an ACP synthase. Finally, the malonyl-CoA:ACP transacylase (MCAT) charges the malonyl-CoA to holo-ACP, resulting in malonyl-ACP building blocks needed by fatty acid synthases (Fig. 5). Although the occurrence of diacyl glycerols is generally believed to be restricted to Bacteria and Eukarya, the presence of minor amounts of free fatty acids have been previously reported in some archaea (Kates *et al*., 1968; Langworthy *et al*., 1974; Gattinger *et al*., 2002). In line with this, homologs of several of bacterial enzymes of the fatty acid biosynthetic pathway (i.e. ACC; beta-ketoacyl synthase, KAS, FabH; beta-ketoacyl reductase, KR, FabG; DH, beta-hydroxyacyl dehydratase; and enoyl reductase, ER) have been detected in some archaeal genomes (Pereto *et al*., 2004; Fig. 5). In addition, a study by Iverson *et al*. (2012) annotated homologs of the genes coding for ACC, KAS, KR, and ER in the genome of a MGII euryarchaeota. Lombard *et al*. (2012b) observed that, except for a few unrelated species that probably acquired ACP by independent HGT (i.e. some species of the Methanomicrobiales and Halobacteriales), no homologs of the ACP-processing machinery (ACP synthase and MCAT) could be detected in archaeal genomes. More recently, Dibrova *et al*. (2014) suggested a hypothetical archaeal fatty acid

pathway based on the presence of the gene coding for archaeal acetyl-CoA-acetyltransferase (also known as acetyl-CoA C-acyl transferase) in all archaeal genomes up to date with the exception of *Nanoarchaeum equitans* (Matsumi *et al*., 2011). This enzyme catalyzes the first condensation reaction in the mevalonate pathway producing acetoacyl-CoA, which then would be further reduced and dehydrated by bacterial-type enzymes involved in the β-oxidation of fatty acids (acyl-CoA dehydrogenase, FadE; enoyl-CoA hydratase, FadB1; 3-hydroxyacyl-CoA dehydrogenase, FadB2; Fig. 5), operating in reverse direction.

To shed further light on the potential capacity of archaea to synthesize fatty acids we performed an extensive search for archaeal homologs of the enzymes involved in the ACP-processing machinery (i.e. ACP synthase and MCAT), which are generally lacking in archaeal genomes and could potentially prohibit a complete the archaeal fatty acid biosynthesis. We also searched for archaeal homologs of the beta-ketoacyl synthase (KAS), which have been previously annotated in several archaeal genomes. Moreover, we also performed a more extensive search for the homologs potentially involved in the hypothetical archaeal fatty acid synthesis pathway as proposed by Dibrova *et al*., (2014) in available archaeal genomes.

Putative homologs of ACP synthase were detected in some species of Methanomicrobiales, Archaeoglobales, Halobacteriales, Thermoplasmatales, in some Crenarchaeota species, and in a genome of a species in the DPANN superphylum (*i.e.* AR5 Aenigmarchaeota; Table 1; Table S1). In our survey, we also detected MCAT (FabD) homologs in MGII and MGIII genomes (Table 1; Table S1), as well as in some Halobacteriales (cf. Lombard *et al*., 2012b). In the case of the MGII, we detected the MCAT (FabD) gene as part of the previously annotated polyketide synthase, which also contained a beta-ketoacyl synthase (KAS) and a polyketide synthase dehydratase (FabA/Z dehydratase) domains. We have also observed this distribution of the MCAT (FabD) coding region for several of the MGII/III genomes released by Deschamps *et al*. (2014) (Table 1; Table S1). In addition, a putative MCAT homolog was also detected in the 'Lokiarchaeum' genome within the protein previously annotated as phenol phthiocerol synthesis polyketide synthase type I (Table S1, S2). This protein harbors a MCAT (FabD; comprised between residues 14604−15485; Table S2),

and KAS regions (FabBI and FabFII; residues 13674−14525). We performed phylogenetic analyses of the MCAT (FabD) proteins detected in archaeal genomes to determine their evolutionary relationships between each other and with bacterial homologs (Fig. 6). MCAT homologs of the MGII and MGIII, and 'Lokiarchaeum' were related to bacterial MCAT homologs of the Acidobacteria, Chloroflexi, and Firmicutes, while the MCAT homologs detected in Halobacteriales genomes were quite different from the rest of the archaeal homologs as well as the bacterial ones, suggesting a different evolutionary origin for these MCAT proteins.

Archaeal homologs of the genes coding for the acyl-CoA dehydrogenase FadE, enoyl-CoA hydratase FadB1, and 3-hydroxyacyl-CoA dehydrogenase FadB2 were detected in some of the genomes of the euryarchaeota Archaeoglobales, Halobacteriales, Thermoplasmatales, uncultured marine euryarchaeota group II/III, MBG-D, and most of the Crenarchaeotal and Thaumarchaeotal genomes (Table 1; Table S1). One putative homolog of FadE was detected in the 'Lokiarchaeum' genome, as well as multiple copies of putative homologs of FadB1 and FadB2 (Table 1; Table S1, S2), which would also suggest that 'Lokiarchaeum' harbors the potential for fatty acid synthesis with the hypothetical pathway proposed by Dibrova *et al*. (2014).

Our survey of the occurrence of genes coding for key fatty acid biosynthetic enzymes (i.e. ACP synthase, MCAT and KAS) in archaeal genomes (Table 1; Table S1) suggests that only species of the Halobacteria, MGII/III euryarchaeota and 'Lokiarchaeum' have all the key genes required to potentially synthesize fatty acids. Species of the Halobacteria possess all three key genes (Table 1; Table S1). Although MGII/III and 'Lokiarchaeum' lack annotated homologs of ACP synthase (Table 1; Table S1), these archaeal groups could potentially still synthesize bacterial-like fatty acids by an ACP-independent pathway as proposed by Lombard *et al*. (2012b). Furthermore, the extensive search we performed for the genes coding for the FadE, FadB1 and FadB2, also showed that most of the archaeal genomes, including MGII/III euryarchaeota and 'Lokiarchaeum', harbor the potential for hypothetical fatty acid biosynthesis pathway proposed by Dibrova *et al*. (2014) (Fig. 5).

Based on the genetic potential of some archaea to produce fatty acids, we further investigated archaeal genomes for genes coding for enzymes catalyzing the esterification of fatty acids and G3P required for the formation of glycerol ester lipids (Fig. 5). There are two families of enzymes responsible for the acylation of the 1-position of the G3P. The PlsB acyltransferase, found in the bacteria *Escherichia coli* and in many eukaryotes, primarily uses acyl-acyl carrier protein (ACP) end products of fatty acid biosynthesis (acyl-ACP) as acyl donors but may also use acyl-CoA derived from exogenous fatty acids (Fig. 5). The other family concerns the PlsY acyltransferase which is more widely distributed in Bacteria and uses as donor acyl-phosphate produced from acyl-ACP by the PlsX, an acyl-ACP:PO_4 transacylase enzyme (Fig. 5). The acylation in the 2-position of the G3P is carried out by the 1-acylglycerol-3-phosphate *O*-acyltransferase (PlsC) (Fig. 5).

We performed genomic searches of putative archaeal homologs of the bacterial acyl-ACP transferases involved in the formation of ester bonds between fatty acids and the G3P backbone in the phospholipid synthesis (Fig. 5). No homologs of the PlsB acyltransferase were detected in any of the archaeal genomes with the exception of one species of the DPANN superphylum (AR1 Pacearchaeota; Table 1; Table S1). Archaeal homologs to the PlsY acyltransferase were only found in the 'Lokiarchaeum' genome (i.e. one putative PlsY homolog; Table 1; Table S1, S2). Molecular phylogeny (Fig. 7) indicated that the 'Lokiarchaeum' PlsY homolog was closely related to PlsY enzymes of bacterial groups such as Thermotogae, Spirochaetales and Dictioglomales which suggest that this enzyme was acquired by lateral gene transfer from Bacteria. A wide variety of putative homologs of the PlsC 1-acylglycerol-3-phosphate *O*-acyltransferase were found in MGII and MGIII genomes, a DPANN genome (AR11 Woesearchaeota), and two putative PlsC were found in the 'Lokiarchaeum' genome (Table 1; Table S1, S2). Moreover, PlsC protein homologs of the MGII/III euryarchaeota, AR11 DPANN, and 'Lokiarchaeum' were closely related to those of the α-, β-Proteobacteria or Actinobacteria, the newly proposed Parcubacteria group (Brown *et al*., 2015), and a species of the Firmicutes, respectively (Fig. 7).

Our data suggest that, in addition to the apparent ability to synthesize 'bacterial' G3P, MGII/III euryarchaeota and 'Lokiarchaeum' also possess a putative fatty acid synthetic pathway.

Furthermore, the detection of homologs of the acyl transferase PlsY in the 'Lokiarchaeum' genome and PlsC homologs in both MGII/III euryarchaeota and 'Lokiarchaeum' genomes suggests that their biosynthetic machinery would be able to form ester-bonded fatty acid membrane lipids with G3P stereochemistry. Based on the enzyme inventory (Table 1; Table S1), these archaea are predicted to produce chimeric membrane lipids, such as di- or tetraether-linked isoprenoidal membrane lipids with a bacterial/eukaryote G3P stereochemistry, or lipids with one ether-linked isoprenoidal chain at position *sn*-1 of a G3P backbone and one ester-bound fatty acid at position *sn*-2 (see "hypothetical part" of Fig. 1). Mixed ether/ester membrane lipids have been previously detected in aerobic and anaerobic bacteria such as anammox bacteria, sulfate-reducing bacteria, members of the bacterial order Thermotogales and Acidobacteria (Rütters *et al*., 2001; Sinninghe Damsté *et al*., 2002, 2007, 2011, 2014). In fact, the presence of these types of lipids in the order Thermotogales, an early-branching clade of the Bacteria, was interpreted as an indication that the ability to produce both ether and ester-linked membrane lipids developed relatively early during microbial evolution (Sinninghe Damsté *et al*., 2007). However, the early branching in the tree of life of Thermotogales and Aquificales has been questioned and it has even been proposed that the majority of the genes of these groups shows affinities to Archaea and Firmicutes (Zhaxybayeva *et al*., 2009, among others).

In addition, 'chimeric' tetraether lipids containing both n-alkyl and isoprenoidal chains have been previously detected in the environment (Schouten *et al*., 2000; Liu *et al*., 2012). In the case of 'Lokiarchaeum', it is also possible that they synthesize bacterial-like fatty acids ester-bound at the *sn*-1 and *sn*-2 positions of the G3P, as we have detected both putative acyltranferases (PlsY and PlsC) in its genome. However, we did not detect a putative homolog of the PlsX protein involved in the transformation of acyl-ACP to acylphosphate needed for the catalysis mediated by PlsY. Therefore it remains unknown if 'Lokiarchaeum' is able to mediate the formation of bacterial-like ester-bond fatty acid in the *sn*-1 position. The formation of the hypothetical chimeric tetraether lipids (see "hypothetical part" of Fig. 1) would follow a biosynthetic pathway in which enzymes similar to GGGP and DGGGP synthases would catalyze the formation of ether bonds between

isoprenoid chains and G3P backbone in *sn*-2 position instead of with the expected G1P. Since GGGP synthase has proven to be selective for G1P (Peterhoff *et al*., 2014), it would suggest that this step is mediated by a totally different enzyme

Implications for the 'lipid divide'

The potential capacity of synthesis of isoprenoidal ether and fatty acid ester lipids with a G3P backbone within a single organism, i.e. MGII/III euryarchaeota species and 'Lokiarchaeum', sheds new light on the current 'lipid divide'. This is especially relevant for 'Lokiarchaeum' as its genome codes a remarkable number of eukaryotic signature proteins, which has been used as an argument to support the hypothesis that the eukaryotic cell evolved from an archaeal ancestor of this group (Spang *et al*., 2015). Our genome mining study suggests that 'Lokiarchaeum' has the biosynthetic capacity to synthesize archaeal ether-linked and fatty acid ester linked membrane lipids with the bacterial/eukaryotic G3P stereochemistry further supports the hypothesis of Spang *et al*. (2015). If 'Lokiarchaeum' was indeed a descendant of the archaeal ancestor leading to the eukaryotic cell, then this ancestor may have possessed the capacity for both isoprenoidal ether and fatty acid ester lipids with a G3P backbone. After the endosymbiosis of the archaeal ancestor with a bacterium, the capacity for isoprenoid ether lipid synthesis may have been lost, leaving the fatty acids ester lipids with a G3P backbone as the main membrane lipid.

It is not clear why phylogenetically distant archaeal groups such as MGII/III euryarchaeota and 'Lokiarchaeum' both harbor these particular lipid biosynthetic capacities. Extensive bacteria-to-archaea gene transfer has occurred in MGII/III euryarchaeota, Thaumarchaeota, Halobacteria and mesophilic methanogens (López-García *et al*., 2004; Brochier-Armanet *et al*., 2011; Nelson-Sathi *et al*., 2012; Deschamps *et al*., 2014). It has been proposed that this has promoted their adaptation to a mesophilic lifestyle (López-García *et al*., 2015). The Lokiarchaeum genome also contains a relatively high fraction of genes that display a high similarity to genes of bacterial origin (i.e. 29% of all genes; Spang *et al.,* 2015), which is comparable to that in MGII/III euryarchaeota (Deschamps *et al*., 2014). This high level inter-domain gene exchange between Bacteria and

Archaea may have substantially impacted the membrane lipid biosynthetic pathway in both MGII/III euryarchaeota and 'Lokiarchaeum' to such an extent that they produce 'chimeric' membrane lipids. Acquisition of only one more bacterial gene (PlsB) by an ancestor of 'Lokiarchaeum' would result in a 'full' bacterial/eukaryotic lipid membrane pathway, paving the road leading to the development of a 'truly' eukaryotic cell membrane. Our study suggests that the 'lipid divide' between the domain Archaea, on the one hand, and those of the Bacteria and Eukarya, on the other, is less clear cut as previously thought.

The required next step following our phylogenomic study is to provide confirmation of our hypothesis by identification of the 'chimeric' lipids predicted here. However, such an endeavor is strongly hindered by the lack of cultivated representatives of MGII/III euryarchaeota and 'Lokiarchaeota'. Future studies should focus on determining 'unusual' membrane lipids in natural environments with high abundances of these uncultured archaeal groups, in particular by determining the stereochemistry of their glycerol membrane lipids (cf. Weijers *et al.,* 2006). This may also provide further insight in the evolution of Eukarya from the prokaryotes.

Experimental procedures

Computational analysis

Putative homologs of the enzymes mentioned in the text (Table 1; Table S1, S2) were detected by tblastn (search translated nucleotide databases using a protein query) and blastp (protein query against protein databases) searches using annotated enzymes as query sequences and with a minimum e-value of $1e^{-25}$. The identity of the putative homologs was further investigated by visual inspection of the alignment.

Phylogenetic analyses

Putative and annotated partial homologs aligned by Muscle (Edgar, 2004) in Mega6 software (Tamura *et al*., 2013) and edited manually. Phylogenetic reconstruction was performed by maximum likelihood in PhyML v3.0 (Guindon *et al*., 2010) using the best model according to AIC indicated by ProtTest 2.4 (Abascal *et al*., 2005) as indicated in the figure legends.

References

Abascal, F., Zardoya, R., and Posada, D. (2005) ProtTest: Selection of best-fit models of protein evolution. Bioinformatics 21: 2104-2105.

Boucher, Y., Kamekura, M, and Doolittle, W.F. (2004) Origins and evolution of isoprenoid lipid biosynthesis in archaea. Mol Microbiol 52: 515-527.

Brochier-Armanet, C., Forterre, P., and Gribaldo, S. (2011) Phylogeny and evolution of the Archaea: one hundred genomes later. Curr Opin Microbiol 3: 274-281.

Brown, C.T., Hug, L.A., Thomas, B.C., Sharon, I., Castelle, C.J., Singh, A., Wilkins, M.J., Wrighton, K.C., Williams, K.H., and Banfield, J.F. (2015) Unusual biology across a group comprising more than 15% of domain Bacteria. Nature 523: 208-211.

Castelle, C.J., Wrighton, K.C., Thomas, B.C., Hug, L.A., Brown, C.T., Wilkins, M.J., Frischkorn, K.R., Tringe, S.G., Singh, A., Markillie, L.M., Taylor, R.C., Williams, K.H., and Banfield, J.F. (2015) Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. Curr Biol 25: 690-701.

Deschamps, P., Zivanovic, Y., Moreira, D., Rodriguez-Valera, F., and López-García, P. (2014) Pangenome evidence for extensive interdomain horizontal transfer affecting lineage core and shell genes in uncultured planktonic thaumarchaeota and euryarchaeota. Genome Biol Evol 6: 1549-1563.

Dibrova, D.V., Galperin, M.Y., and Mulkidjanian, A.Y. (2014) Phylogenomic reconstruction of archaeal fatty acid metabolism. Environ Microbiol 16: 907-918.

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113.

Fan, Q., Relini A., Cassinadri, D., Gambacorta, A., and Gliozzi, A. (1995) Stability against temperature and external agents of vesicles composed of archael bolaform lipids and egg PC. Biochim Biophys Acta 1240: 83-88.

Fisher, E., Almaguer, C., Holic, R., Griac, P., and Patton-Vogt, J. (2005) Glycerophosphocholine-dependent growth requires Gde1p (YPL110c) and Git1p in Saccharomyces cerevisiae. J Biol Chem 280: 36110-36117.

Gattinger, A., Schloter, M., and Munch, J. C. (2002) Phospholipid etherlipid and phospholipid fatty acid fingerprints in selected euryarchaeotal monocultures for taxonomic profiling. FEMS Microbiol Lett 213: 133-139.

Golding, G.B., and Gupta, R.S. (1995) Protein-based phylogenies support a chimeric origin for the eukaryotic genome. Mol Biol Evol 12: 1-6.

Gray, M., and Doolittle, W.F (1982) Has the endosymbiont hypothesis being proven? Microbiol Rev 46: 1-42.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59: 307-321.

Guldan, H., Sterner, R., and Babinger, P. (2008) Identification and characterization of a bacterial glycerol-1-phosphate dehydrogenase:Ni^{2+}-dependent AraM from Bacillus subtilis. Biochem 47: 7376-7384.

Guy, L., and Ettema, T.J. (2011)The archaeal 'TACK' superphylum and the origin of eukaryotes. Trends Microbiol 19: 580-587.

Han, J-S., and Ishikawa, K (2005) Active site of $Zn^{(2+)}$-dependent sn-glycerol-1-dehydrogenase from Aeropyrum pernix K1. Archaea 1: 311-317.

Istivan, T.S., and Coloe, P.J. (2006) Phospholipase A in Gram-negative bacteria and its role in pathogenesis. Microbiology 152: 1263-1274.

Iverson, V., Morris, R.M., Frazar, C.D., Berthiaume, C.T., Morales, R.L., and Armbrust, E.V. (2012) Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. Science 335: 587-590.

Jahn, U., Summons, R., Sturt, H., Grosjean, E., and Huber, H. (2004) Composition of the lipids of Nanoarchaeum equitans and their origin from its host Ignicoccus sp. strain KIN4/I . Archiv Microbiol 182: 404-413.

Jung, M.Y., Park, S.J., Kim, S.J., Kim, J.G., Sinninghe Damsté, J.S., Jeon, C.O., and Rhee, S.K. (2014) A mesophilic, autotrophic, ammonia-oxidizing archaeon of thaumarchaeal group I.1a cultivated from a deep oligotrophic soil horizon. Appl Environ Microbiol 80: 3645-3655.

Kates, M. (1993) in Membrane lipids of archaea, The biochemistry of Archaea (Archaebacteria). D. J. Kushner, D.J., and Matheson, A.T. eds. (Elsevier Science Publishers B.V., Amsterdam, The Netherlands), p. 261-295.

Kates, M., Wassef, M.K., and Kushner, D.J. (1968) Radioisotopic studies on the biosynthesis of the glyceryl diether lipids of Halobacterium cutirubrum. Can J Biochem 46: 971-977.

Koga, Y., and Morii, H. (2007) Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations. Microbiol Mol Biol Rev 71: 97-120.

Koga, Y., Kyuragi, T., Nishihara, M., and Sone, N. (1998) Did archaeal and bacterial cells arise independently from non-cellular precursors? A hypothesis stating that the advent of membrane phospholipid with enantiomeric glycerol phosphate backbones caused the separation of the two lines of descent. J Mol Evol 47: 631.

Koga, Y., Sone, N., Noguchi, S., and Morii, H. (2003) Transfer of pro-R hydrogen from NADH to dihydroxyacetonephosphate by sn-glycerol-1-phosphate dehy- drogenase from the archaeon Methanothermobacter thermautotrophicus. Biosci Biotechnol Biochem 67: 1605-1608.

Langworthy, T.A., Mayberry, W.R., and Smith, P.F. (1974) Long-chain glycerol diether and polyol dialkyl glycerol triether lipids of Sulfolobus acidocaldarius. J Bacteriol 119: 106-116.

Larson, T.J., Ehrmann, M., and Boos, W. (1983) Periplasmic glycerophosphodiester phosphodiesterase of Escherichia coli, a new enzyme of the glp regulon. J Biol Chem 258: 5428-5432.

Lebedeva, E.V., Hatzenpichler, R., Pelletier, E., Schuster, N., Hauzmayer, S., Bulaev, A., Grigor'eva, N.V., Galushko, A., Schmid, M., Palatinszky, M., Le Paslier, D., Daims, H. and Wagner, M. (2013) Enrichment and genome sequence of the group I.1a ammonia-oxidizing Archaeon "Ca. Nitrosotenuis uzonensis" representing a clade globally distributed in thermal habitats. PLoS One 8: e80835.

Liu, X.L., Summons, R.E., and Hinrichs, H.U. (2012) Extending the known range of glycerol ether lipids in the environment: structural assignments based on tandem mass spectral fragmentation patterns. Rapid Commun Mass Spectrom 26: 2295-2302.

Lloyd, K.G., Schreiber, D.G., Petersen, D.G., Kjeldsen, K.U., Lever, M.A., Steen, A.D., Stepanauskas, R., Richter, M., Kleindienst, S., Lenk, S., Schramm, A., and Jørgensen, B.B. (2013) Predominant archaea in marine sediments degrade detrital proteins. Nature 496: 215-218.

Lombard, J., López-García, P., and Moreira, D. (2012a) The early evolution of lipid membranes and the three domains of life. Nat Rev Microbiol 10: 507-515.

Lombard, J., López-García, P., and Moreira, D. (2012b) Phylogenomic investigation of phospholipid synthesis in archaea. Archaea 2012: 630910.

López-García, P., Brochier, C., Moreira, D., and Rodríguez-Valera, F. (2004) Comparative analysis of a genome fragment of an uncultivated mesopelagic crenarchaeote reveals multiple horizontal gene transfers. Environ Microbiol 6: 19-34.

López-García, P., Zivanovic, Y., Deschamps, P., and Moreira, D. (2015) Bacterial gene import and mesophilic adaptation in archaea. Nat Rev Microbiol 13: 447-456.

Martijn, J., and Ettema, T.J.G. (2013) From archaeon to eukaryote: the evolutionary dark ages of the eukaryotic cell. Biochem Soc Trans 41: 451-457.

Martin, W., and Russell, M.J. (2003) On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. Philos Trans R Soc Lond B Biol Sci 358: 59-83.

Matsumi, R., Atomi, H., Driessen, A.J., and van der Oost, J. (2011) Isoprenoid biosynthesis in Archaea--biochemical and evolutionary implications. Res Microbiol 162: 39-52.

Nelson-Sathi, S., Dagan, T., Janssen, A., Steel, M., McInerney, J.O., Deppenmeier, U., and Martin, W.F. Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. Proc Natl Acad Sci USA 109: 20537-20542.

Nishihara, M., Yamazaki, T., Oshima, T., and Koga, Y. (1999) sn-Glycerol-1-phosphate-forming activities in archaea: separation of archaeal phospholipid biosynthesis and glycerol catabolism by glycerophosphate enantiomers. J Bacteriol 181: 1330-1333.

Pereto, J., Lopez-Garcia, P., and Moreira, D. (2004) Ancestral lipid biosynthesis and early membrane evolution. Trends Biochem Sci 29: 469-477.

Peterhoff, D., Beer, B., Rajendran, C., Kumpula, E.P., Kapetaniou, E., Guldan, H., Wierenga, R.K., Sterner, R., and Babinger, P. (2014) A comprehensive analysis of the geranylgeranylglyceryl phosphate synthase enzyme family identifies novel members and reveals mechanisms of substrate specificity and quaternary structure organization. Mol Microbiol 92: 885-899.

Rütters, H., Sass, H., Cypionka, H., and Rullkotter, J. (2001) Monoalkylether phospholipids in the sulfate-reducing bacteria Desulfosarcina variabilis and Desulforhabdus amnigenus. Arch Microbiol 176: 435-442.

Santelli, E., Schwarzenbacher, R., McMullan, D. et al. (2004) Crystal structure of a glycerophosphodiester phosphodiesterase (GDPD) from Thermotoga maritima (TM1621) at 1.60Å resolution. Proteins 56: 167-170.

Santoro, A.E., Dupont, C.L., Richter, R.A., Craig, M.T., Carini, P., McIlvin, M.R., Yang, Y., Orsi, W.D., Moran, D.M., and Saito, M.A. (2015) Genomic and proteomic characterization of "Candidatus Nitrosopelagicus brevis": an ammonia-oxidizing archaeon from the open ocean. Proc Natl Acad Sci USA 112: 1173-1178.

Schouten, S., Hopmans, E.C., Pancost, R.D., and Damsté, J.S. (2000) Widespread occurrence of structurally diverse tetraether membrane lipids: evidence for the ubiquitous presence of low-temperature relatives of hyperthermophiles. Proc Natl Acad Sci USA 97: 14421-14426.

Shimada, H., and Yamagishi, A. (2011) Stability of heterochiral hybrid membrane made of bacterial sn-G3P lipids and archaeal sn-G1P lipids. Biochem 50: 4114-4120.

Sinninghe Damsté, J. S., Rijpstra, W.I., Hopmans, E.C., Schouten, S., Balk, M., and Stams, A.J. (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.

Sinninghe Damsté, J.S., Rijpstra. W.I., Hopmans, E.C., Foesel, B.U., Wüst, P.K., Overmann, J., Tank, M., Bryant, D.A., Dunfield, P.J., Houghton, K., and Stott, M.B. (2014) Ether- and ester-bound iso-diabolic acid and other lipids in members of acidobacteria subdivision 4. Appl Environ Microbiol 80: 5207-5218.

Sinninghe Damsté, J.S., Rijpstra. W.I., Hopmans, E.C., Weijers, J.W., Foesel, B.U., Overmann, J.. and Dedysh, S.N. (2011) 13,16-Dimethyl octacosanedioic acid (iso-diabolic acid), a common membrane-spanning lipid of Acidobacteria subdivisions 1 and 3. Appl Environ Microbiol 77: 4147-4154.

Sinninghe Damsté, J.S., Strous, M., Rijpstra, W.I., Hopmans, E.C., Geenevasen, J.A., van Duin, A.C., van Niftrik, L.A., and Jetten, M. (2002) Linearly concatenated cyclobutane lipids form a dense bacterial membrane. Nature 419: 708-712.

Sojo, V., Pomiankowski, A., and Lane, N. (2014) A bioenergetic basis for membrane divergence in archaea and bacteria. PLoS Biol. 12: e1001926.

Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., van Eijk, R., Schleper, C., Guy, L., and Ettema, T.J. (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521: 173-179.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30: 2725-2729.

Tommassen, J., Eiglmeier, K., Cole, S.T., Overduin, P., Larson, T.J., and Boos, W. (1991) Characterization of two genes, glpQ and ugpQ, encoding glycerophosphoryl diester phosphodiesterases of Escherichia coli. Mol Gen Genet 226: 321-327.
van der Rest, B., Boisson, A.M., Gout, E., Bligny, R., and Douce, R. (2002) Glycerophosphocholine metabolism in higher plant cells. Evidence of a new glyceryl-phosphodiester phosphodiesterase. Plant Physiol 130: 244-255.

Villanueva, L., Schouten, S., and Sinninghe Damsté, J.S. (2014) A Re-evaluation of the Archaeal Membrane Lipid Biosynthetic Pathway. Nat Rev Microbiol 12: 438-448.

Wächtershäuser, G. (2003) From pre-cells to Eukarya–a tale of two lipids. Mol Microbiol 47: 13-22.

Waters, E., Hohn, M.J., Ahel, I., Graham, D.E., Adams, M.D., Barnstead, M., Beeson, K.Y., Bibbs, L., Bolanos, R., Keller, M., Kretz, K., Lin, X., Mathur, E., Ni, J., Podar, M., Richardson, T., Sutton, G.G., Simon, M., Soll, D., Stetter, K.O., Short, J.M., and Noordewier, M. (2003) The genome of Nanoarchaeum equitans: insights into early archaeal evolution and derived parasitism. Proc Natl Acad Sci USA 100: 12984-12988.

Weijers, J. W., Schouten, S., Hopmans, E.C., Geenevasen, J.A., David, O.R., Coleman, J. M., Pancost, R.D., and Sinninghe Damsté, J.S. (2006) Membrane lipids of mesophilic anaerobic bacteria thriving in peats have typical archaeal traits. Environ Microbiol 8: 648-655.

Williams, T.A., Foster, P.G., Nye, T.M.W., Cox, C.J., and Embley, T.M. (2012) A congruent phylogenomic signal places eukaryotes within the Archaea. Proc Royal Soc B 279: 4870-4879.

Woese, C.R, and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc Nat Acad Sci USA 74: 5088-5090.

Zhaxybayeva, O., Swithers, K.S., Lapierre, P., Fournier, G.P., Bickhart, D.M., DeBoy, R.T., Nelson, K.E., Nesbø, C.L., Doolittle, W.F., Gogarten, J.P., and Noll, K.M. (2009) On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. Proc Natl Acad Sci U S A. 106: 5865-5870.

Acknowledgments

This research was supported by the SIAM Gravitation Grant 024.002.002 from the Dutch Ministry of education, Culture, and Science (OCW. Author contributions: L.V. and J.S.S.D. conceived the study. L.V. was responsible for data processing, and analysis. All authors contributed to data interpretation, formulation of hypotheses, and the writing of the paper. The authors do not have any conflict of interest to declare.

Table 1. Presence (√ in green) and absence (× in red) of archaeal homologs of enzymes related to membrane lipid biosynthesis, glycerol catabolism, and enzymes involved in fatty acid and mono- and diacyl glycerol biosynthesis (see Figure 1 and 5). Enzymes studied: MVK, Mevalonate kinase; G1PDH, Glycerol-1-phosphate dehydrogenase; G3PDH, Glycerol-3-phosphate dehydrogenase; *glp*K, glycerol kinase; GDPD, glycerophosphodiester phosphodiesterase; GGGP, geranylgeranylglyceryl phosphate synthase (see archaeal GGGP synthase phylogenetic tree in Fig. S2), DGGGP, digeranylgeranylglyceryl phosphate synthase (phylogenetic tree in Fig. S3); GR, geranylgeranyl reductase; FabD, MCAT Malonyl-coA:ACP-transacylase; KAS, beta-ketoacyl synthase; FadE, Acyl-CoA dehydrogenase; FadB1, Enoyl-CoA hydratase; FadB2, 3-hydroxyacyl-CoA dehydrogenase; PlsB, glycerol-3-phosphate *O*-acyltransferase; PlsY, glycerol-3-phosphate acyltransferase; PlsC, 1-acylglycerol-3-phosphate *O*-acyltranferase. For a complete overview see Table S1.

		Isoprenoid biosythesis	**Glycerol backbone biosynthesis**		**Glycerol & glycerol phosphodiester catabolism**			**Ether lipid production** (ether bond formation & saturation of isoprenoids)			**Fatty acid biosynthesis**						**Ester–bond formation glycerol and fatty acids**		
Phylogenetic classification	**%****	**MVK**	**G1PDH**	**G3PDH *gps***	**G3PDH *glp***	***glp*K**	**GDPD**	**GGGP**	**DGGGP**	**GR**	**ACP synthase**	**FabD**	**KAS**	**FadE**	**FadB1**	**FadB2**	**PlsB**	**PlsY**	**PlsC**
EURYARCHAEOTA																			
Methanococcales	100	√	√	×	×	×	×	√	√	√	×	×	×	×	×	×	×	×	×
Methanobacteriales	100	√	√	√	×	×	√	√	√	√	×	×	×	×	×	×	×	×	×
Thermococcales	100	√	√	×	√	√	√	√	√	√	×	×	×	×	×	×	×	×	×
Methanosarcinales	100	√	√	×	×	×	×	√	√	√	×	×	×	×	√	×	×	×	×
Methanomicrobiales	100	√	√	×	×	×	√	√	√	√	√	×	×	×	×	×	×	×	×
Archaeoglobales	100	√	√	√	√	√	×	√	√	√	√	×	√	√	√	√	×	×	×
Halobacteriales	100	√	√	×	√	√	√	√	√	√	√	√	×	√	√	√	×	×	×
Thermoplasmatales																			
Thermoplasmata	100	√	√	×	√	√	√	√	√	√	√	×	×	√	×	√	×	×	×
Unclassified																			
Aciduliprofundum	100	√	√	×	√	√	√	√	√	√	×	×	×	×	×	×	×	×	×
Marine group II/III†	100	√	×	√	×	√	√	√	√	√	×	√	√	√	√	√	×	×	√
MBG-D‡	70	×	√	×	×	×	×	√	√	√	×	×	×	√	√	√	×	×	×
AIGARCHAEOTA																			
Ca. Caldiarchaeum subterraneum	100	√	√	×	×	×	×	√	√	√	×	×	×	√	√	√	×	×	×
KORARCHAEOTA																			
Ca. Korarchaeum cryptofilum	100	√	√	×	√	√	×	√	√	√	×	×	×	√	×	√	×	×	×

CRENARCHAEOTA																		
Sulfolobales	100	√	√	×	×	√	√	√	√	√	√	×	×	√	√	√	×	×	×
Desulfurococcales	100	√	√	×	√	√	√	√	√	√	√	×	×	√	√	√	×	×	×
Acidilobales	100	√	√	×	×	×	×	√	√	√	×	×	×	√	×	√	×	×	×
Thermoproteales	100	√	√	×	√	√	√	√	√	√	√	×	√	√	√	√	×	×	×
THAUMARCHAEOTA																			
Cenarchaeales	100	√	√	×	×	×	×	√	?*	√	×	×	×	×	√	√	×	×	×
Nitrosopumilales	100	√	√	×	×	×	×	√	?	√	×	×	×	√	√	√	×	×	×
Nitrososphaerales	100	√	√	×	×	×	×	√	?	√	×	×	×	√	√	√	×	×	×
Unclassified§	100	√	√	×	×	×	×	√	?	√	×	×	×	√	√	√	×	×	×
DSAG/MBG-B#																			
Lokiarchaeum	92	√	×	×	√	√	√	√	√	√	×	√	√	√	√	√	×	√	√
DPANN¶																			
Diapherotrites																			
Ca. Iainarchaeum andersonii	88.5	×	×	×	×	×	×	√	×	√	×	×	×	×	×	×	×	×	×
AR10	100	×	×	×	×	×	×	√	×	√	×	×	×	×	×	×	×	×	×
Woesearchaeota																			
AR20	100	×	×	×	×	×	×	×	×	√	×	×	×	×	×	×	×	×	×
AR3	63	×	×	×	×	×	√	×	×	×	×	×	×	×	×	×	×	×	×
AR9	76	×	×	√	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
AR11	76	×	×	√	×	×	×	×	×	×	×	×	×	×	×	×	×	×	√
Pacearchaeota																			
AR19	91	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
AR1	89	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	√	×	×
Aenigmarchaeota																			
AR5	93	×	×	×	×	×	×	×	×	√	√	×	×	×	×	×	×	×	×
Micrarchaeota																			
Ca. Micrarchaeum acidiphilum	100	×	×	×	×	×	×	×	×	×	×	×	×	×	×	√	×	×	×
Nanoarchaeota																			
Nanoarchaeum equitans	100	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×

*Refers to the apparent lack of DGGGP synthases in the genomes of Thaumarchaeota; Villanueva *et al*., 2014; **Percentage of completeness of the (meta)genome; † Marine group II and III euryarchaeota genomes including MGII amplified from surface water (CM001443.1; Iverson *et al*., 2012), marine group II euryarchaeote SCGC AB-629-J06 (NZ_AQVM00000000.1), Marine Group III euryarchaeote SCGC AAA288-E19 (AQTX00000000.1), and sequences obtained by Deschamps *et al.* (2014). Only the MGII genome reported by Iverson *et al*., 2012 is closed; ‡MBG-D, Marine Benthic Group D, SCGC AB-539-N05 (ALXL00000000) by Lloyd *et al*. (2013); §Unclassified Thaumarchaeota include *Ca*. Nitrosopelagicus brevis (GCA_000812185.1; Santoro *et al*., 2015); *Ca.* Nitrosoterreus chungbukensis (AVSQ01000000; Jung *et al*., 2014), and *Ca.* Nitrosotenuis uzonensis (CBTY000000000; Lebedeva *et al*., 2013); #DSAG/MBG-B, Deep-Sea Archaeal Group/Marine Benthic Group-B, composite genome 'Lokiarchaeaum' by Spang *et al*. (2015); ¶ DPANN superphylum including the metagenomes described in Castelle *et al*. (2015).

Figure legends

FIGURE. 1. Overview of the known and hypothetical diether and ether/ester lipid biosynthetic pathway in Archaea and the related pathway of glycerol metabolism. Autotrophic Archaea produce G1P, which is subsequently incorporated into archaeal membrane lipids, from GAP via DHAP (orange arrows). The glycerol metabolism pathway of heterotrophic Archaea, which feeds glycerol into the glycolysis pathway, is indicated with blue arrows. The green arrows indicate the here-proposed formation of G3P from DHAP by *gps*-coded G3PDH, commonly only found in Bacteria and Eukarya, and the formation of G3P from the degradation of glycerophosphodiesters by a glycerophosphodiester phosphodiesterase (GDPD) (see text for details). Some other reactions are also performed by these organisms as indicated (B= Bacteria; E= Eukarya). Arrows in gradient orange/blue indicate steps performed by both auto- and heterotrophic archaea. Purple triangles indicate putative homologs in the 'Lokiarchaeum' and/or MGII/III genomes (Table 1). Steps included in the dashed line box are hypothetical and based on the predicted occurrence of specific enzymes (Table 1; Table S1) as discussed in the text. The names of enzymes are underlined and in case where the genes encoding for these enzymes have specific names they are given in italics. Abbreviations used: GAP, d-Glyceraldehyde-3-phosphate; G1P, Glycerol-1-phosphate; G3P, Glycerol-3-phosphate; GGGP, geranylgeranylglyceryl phosphate; DGGGP, digeranylgeranylglyceryl phosphate; DHAP, Dihydroxyacetone phosphate; G1PDH, glycerol-1-phosphate dehydrogenase; G3PDH, glycerol-3-phosphate dehydrogenase. "?" indicates an hypothetical enzyme similar to GGGP synthase but with a G3P stereo-chemistry as indicate din the text. 'DGGGP synthase' and 'PlsC' indicate hypothetical enzymes that are expected to perform a similar reaction to the original ones.

FIGURE 2. Phylogenetic tree of *sn*-glycerol-3-phosphate dehydrogenases (G3PDH). The tree clearly reveals two main clusters of G3PDH encoded by the *glp* (A) and *gps* (B) genes. These different forms of G3PDH catalyze the conversion of G3P into DHAP and the reverse reaction, respectively (see Fig. 1). *Glp*-coded G3PDH is common in Bacteria, Eukarya and Archaea, where it is one of the enzymes involved in feeding glycerol into the glycolysis pathway. The putative *glp*-G3PDH homologs found in the composite 'Lokiarchaeum' genome (Spang *et al*., 2015) are indicated in bold. *Gps*-coded G3PDH is common in Bacteria and Eukarya but the tree reveals that also quite some archaea possess the *gps* gene. The *gps*-G3PDH homologs found in uncultured MG II/III euryarchaeota genomes are indicated in bold. This tree was constructed using the maximum likelihood method with a WAG model plus gamma distribution and invariant site (WAG+G+I+F). The analysis included 1064 positions in the final dataset. Homologous proteins of the closely related family of the UDP (Uridine diphosphate)-glucose 6-dehydrogenases (UDPG-DH) were used as outgroup to construct the tree. The scale bar represents number of amino acid substitutions per site. Branch support was calculated with the approximate likelihood ratio test (aLRT) and values ≥50% are indicated on the branches.

FIGURE 3. Phylogenetic tree of putative archaeal glycerol kinases (*glp*K) and homologous proteins of the xylulose/carbohydrate kinase family. The archaeal *glp*K previously described were grouped in a distinctive cluster, while annotated *glp*K in the 'Lokiarchaeum' genome (indicated in bold; Spang *et al*., 2015) were closely related to carbohydrate kinases of the euryarchaeon *Archaeoglobus fulgidus* and also to xylulose kinases of Bacteria. This tree was constructed using the maximum likelihood method with a LG model plus gamma distribution and invariant site (LG+G+I). The analysis included 585 positions in the final dataset. The scale bar represents number of amino acid substitutions per site. Branch support was calculated with the approximate likelihood ratio test (aLRT) and values ≥50% are indicated on the branches.

FIGURE 4. Phylogenetic tree of putative archaeal UgpQ glycerophosphodiester phosphodiesterases (GDPD) and close relatives within the Bacteria. The putative UgpQ GDPDs detected in the 'Lokiarchaeum' genome (indicated in bold; Spang *et al*., 2015) were closely related to a putative UgpQ GDPD in one genome of the DPANN Woesearchaeota, as well as UgpQ detected in bacterial genomes of

the Thermotogae. Homologous proteins of the closely related family of periplasmic GlpQ GDPDs in Bacteria were used as outgroup. This tree was constructed using the maximum likelihood method with a LG model plus gamma distribution and invariant site (LG+G+I). The analysis included 553 positions in the final dataset. The scale bar represents number of amino acid substitutions per site. Branch support was calculated with the approximate likelihood ratio test (aLRT) and values ≥50% are indicated on the branches.

FIGURE 5. Bacterial biosynthetic pathway resulting in glycerol diester phospholipids formation. Fatty acids are synthesized from acyl-ACP via the FAS-II pathway and coupled with G3P to form phospholipids. The hypothetical archaeal fatty acid biosynthetic pathway proposed by Dibrova *et al*. (2014), based on the archaeal acetyl-CoA-acetyltransferase (acetyl-CoA C-acyl transferase; indicated in orange) and bacterial-type enzymes of the β-oxidation of fatty acids (acyl-CoA dehydrogenase, FadE; enoyl-CoA hydratase, FadB1; 3-hydroxyacyl-CoA dehydrogenase, FadB2) operating in reverse direction is indicated in the dashed line box. Enzymes indicated in red have been previously reported to be present as homologs in archaeal genomes, while enzymes in green indicate bacterial enzymes previously concluded to be absent in archaea. Enzymes discussed in the text are indicated with an asterisk. Purple triangles indicate putative homologs in the 'Lokiarchaeum' and/or MGII/III genomes (Table 1). Abbreviations: ACP, acyl-carrier protein; MCAT, malonyl-CoA:ACP-transacylase, FabD; KAS, beta-ketoacyl synthase (KAS I FabB; KAS II, FabF; KAS III, FabH); KR, beta-ketoacyl reductase, FabG; DH, beta-hydroxyacyl dehydratase, FabA/Z; ER, enoyl reductase, FabI; PlsB, glycerol-3-phosphate *O*-acyltransferase; PlsX, acyl-ACP:PO_4 transacylase; PlsY, G3P acyltransferase; PlsC, 1-acylglycerol-3-phosphate *O*-acyltranferase; G3P, glycerol-3-phosphate.

FIGURE 6. Phylogenetic tree of Malonyl-CoA:ACP-transacylase (MCAT, FabD domain), a key enzyme in the pathway of fatty acid synthesis (see Fig. 5 for details). MCAT homologs in archaeal genomes (in bold) and their closest bacterial sequences are shown. This phylogenetic tree was constructed using the maximum likelihood method and the LG model plus gamma distribution and invariant site (LG+G+I). The analysis included 486 positions in the final dataset. The scale bar represents number of amino acid substitutions per site. Branch support was calculated with the approximate likelihood ratio test (aLRT) and values ≥50% are indicated on the branches. MGII/III: uncultured marine group II/III euryarchaeota.

FIGURE 7. Phylogenetic tree of the putative archaeal homologs of the PlsY glycerol-3-phosphate acyltransferase and the PlsC 1-acylglycerol-3-phosphate *O*-acyltransferase (see Fig. 5 for details). PlsY and PlsC homologs in archaeal genomes (in bold) and their closest bacterial sequences are shown. This tree was constructed using the maximum likelihood method and the LG model plus gamma distribution and invariant site (LG+G+I). The scale bar represents number of substitutions per site. The analysis included 455 positions in the final dataset. Branch support was calculated with the approximate likelihood ratio test (aLRT) and values ≥50% are indicated on the branches. MGII/III: uncultured marine group II/III euryarchaeota.

Supporting information

Fig. S1. Archaeal 16S rRNA gene-based phylogeny modified from Spang *et al*. (2015). TACK, Thaumarchaeota-Aigarchaeota-Crenarchaeota-Korarchaeota superphylum; Bathyarchaeota (Miscellaneous Crenarchaeota Group, MCG and group C3); DSAG, Deep-Sea Archaeal Group/Marine Benthic Group B (including 'Lokiarchaeum'; Spang *et al*., 2015); MHVG, Marine Hydrothermal Vent Group; Euryarchaeota superphylum includes the uncultured group II (MGII) and group III (MGIII) euryarchaeota, among others. DPANN superphylum includes Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, among others (Castelle *et al*., 2015).

Fig. S2. Phylogenetic tree of geranylgeranylglyceryl phosphate (GGGP) synthase homologs in archaeal genomes. Tree was constructed using the maximum likelihood method and the LG model plus gamma

distribution and invariant site (LG+G+I). The scale bar represents number of substitutions per site. The analysis included 364 positions in the final dataset. Branch support was calculated with the approximate likelihood ratio test (aLRT) and values ≥50% are indicated on the branches. This tree showed the previously reported divergence of GGGP synthases in two different clusters (cluster 1 including the Halobacteriales, Archaeoglobales and Methanomicrobiales, Fig. S2B; and cluster 2 including Thaumarchaeota, Crenarchaeota and GGGP synthases of the rest of euryarchaeotal orders; Fig. S2A; Boucher *et al*., 2004; Villanueva *et al*., 2014). The putative GGGP synthase annotated in the 'Lokiarchaeum' genome. This sequence is closely related to GGGP synthases of the Thermoplasmatales, including uncultured marine group II and III euryarchaeota (MGII/III). Fig. S2C indicates the phylogenetic position of the Thaumarchaeota single cell genomes within the tree.

Fig. S3. Phylogenetic tree of putative digeranylgeranylglyceryl phosphate (DGGGP) synthase homologs in archaeal genomes. This tree is based on the putative archaeal DGGGP synthase phylogenetic tree by Villanueva *et al*. (2014). Tree was constructed using the maximum likelihood method and the LG model plus gamma distribution and invariant site (LG+G+I). The scale bar represents number of amino acid substitutions per site. The analysis included 422 positions in the final dataset. Branch support was calculated with the approximate likelihood ratio test (aLRT) and values ≥50% are indicated on the branches. Fig. S3A indicates the phylogenetic relationship between the thaumarchaeotal prenyltransferases and the archaeal DGGGP synthase. Fig. S3B indicates the distribution of the putative DGGGP synthases in the different archaeal groups. The putative DGGGP synthase annotated in the 'Lokiarchaeum ' genome (indicated in bold) was closely related to the DGGGP synthases of the euryarchaeaotal group Archaeoglobales. Putative DGGGP synthases annotated in the genomes of the uncultured marine group II and III euryarchaeota are not clustered with the rest of the putative DGGGP synthases.

Table S1. Compilation of NCBI accession numbers of the enzymes included in Table 1 for the different archaeal genomes analyzed in this study.

Table S2. Presence (√ in green) and absence (× in red) putative homologs of enzymes involved in the archaeal membrane lipid and fatty acid biosynthetic pathways in the 'Lokiarchaeum' genome (Spang *et al*., 2015) and NCBI accession numbers (see Fig. 1, 5, and text for details).

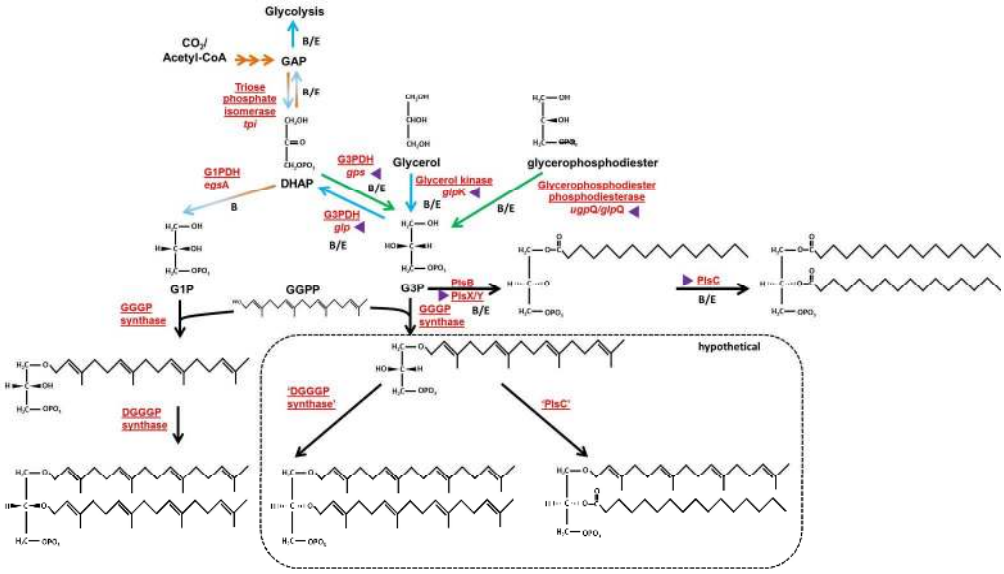


FIGURE. 1. Overview of the known and hypothetical diether and ether/ester lipid biosynthetic pathway in Archaea and the related pathway of glycerol metabolism. Autotrophic Archaea produce G1P, which is subsequently incorporated into archaeal membrane lipids, from GAP via DHAP (orange arrows). The glycerol metabolism pathway of heterotrophic Archaea, which feeds glycerol into the glycolysis pathway, is indicated with blue arrows. The green arrows indicate the here-proposed formation of G3P from DHAP by gps-coded G3PDH, commonly only found in Bacteria and Eukarya, and the formation of G3P from the degradation of glycerophosphodiesters by a glycerophosphodiester phosphodiesterase (GDPD) (see text for details). Some other reactions are also performed by these organisms as indicated (B= Bacteria; E= Eukarya). Arrows in gradient orange/blue indicate steps performed by both auto- and heterotrophic archaea. Purple triangles indicate putative homologs in the 'Lokiarchaeum' and/or MGII/III genomes (Table 1). Steps included in the dashed line box are hypothetical and based on the predicted occurrence of specific enzymes (Table 1; Table S1) as discussed in the text. The names of enzymes are underlined and in case where the genes encoding for these enzymes have specific names they are given in italics. Abbreviations used: GAP, d-Glyceraldehyde-3-phosphate; G1P, Glycerol-1-phosphate; G3P, Glycerol-3-phosphate; GGGP, geranylgeranylglyceryl phosphate; DGGGP, digeranylgeranylglyceryl phosphate; DHAP, Dihydroxyacetone phosphate; G1PDH, glycerol-1-phosphate dehydrogenase; G3PDH, glycerol-3-phosphate dehydrogenase. "?" indicates an hypothetical enzyme similar to GGGP synthase but with a G3P stereo-chemistry as indicate din the text. 'DGGGP synthase' and 'PlsC' indicate hypothetical enzymes that are expected to perform a similar reaction to the original ones.
470x267mm (150 x 150 DPI)

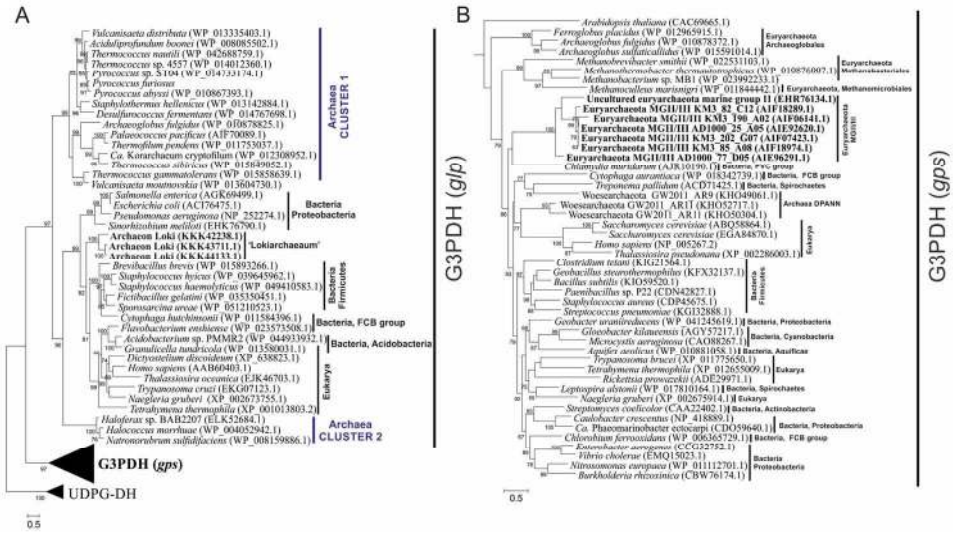


FIGURE 2. Phylogenetic tree of sn-glycerol-3-phosphate dehydrogenases (G3PDH). The tree clearly reveals two main clusters of G3PDH encoded by the glp (A) and gps (B) genes. These different forms of G3PDH catalyze the conversion of G3P into DHAP and the reverse reaction, respectively (see Fig. 1). Glp-coded G3PDH is common in Bacteria, Eukarya and Archaea, where it is one of the enzymes involved in feeding glycerol into the glycolysis pathway. The putative glp-G3PDH homologs found in the composite 'Lokiarchaeum' genome (Spang et al., 2015) are indicated in bold. Gps-coded G3PDH is common in Bacteria and Eukarya but the tree reveals that also quite some archaea possess the gps gene. The gps-G3PDH homologs found in uncultured MG II/III euryarchaeota genomes are indicated in bold. This tree was constructed using the maximum likelihood method with a WAG model plus gamma distribution and invariant site (WAG+G+I+F). The analysis included 1064 positions in the final dataset. Homologous proteins of the closely related family of the UDP (Uridine diphosphate)-glucose 6-dehydrogenases (UDPG-DH) were used as outgroup to construct the tree. The scale bar represents number of amino acid substitutions per site. Branch support was calculated with the approximate likelihood ratio test (aLRT) and values ≥50% are indicated on the branches.

306x172mm (150 x 150 DPI)

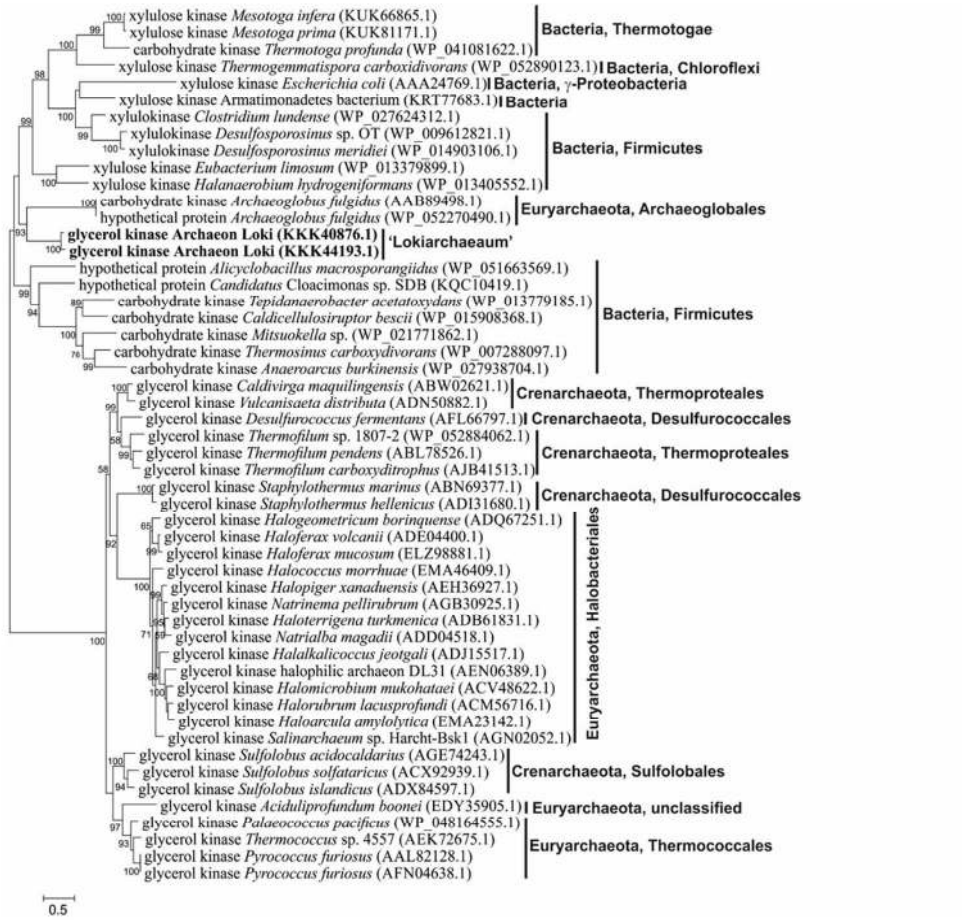


FIGURE 3. Phylogenetic tree of putative archaeal glycerol kinases (glpK) and homologous proteins of the xylulose/carbohydrate kinase family. The archaeal glpK previously described were grouped in a distinctive cluster, while annotated glpK in the 'Lokiarchaeum' genome (Spang et al., 2015; indicated in bold) were closely related to carbohydrate kinases of the euryarchaeon Archaeoglobus fulgidus and also to xylulose kinases of Bacteria. This tree was constructed using the maximum likelihood method with a LG model plus gamma distribution and invariant site (LG+G+I). The analysis included 585 positions in the final dataset. The scale bar represents number of amino acid substitutions per site. Branch support was calculated with the approximate likelihood ratio test (aLRT) and values ≥50% are indicated on the branches.

198x184mm (150 x 150 DPI)

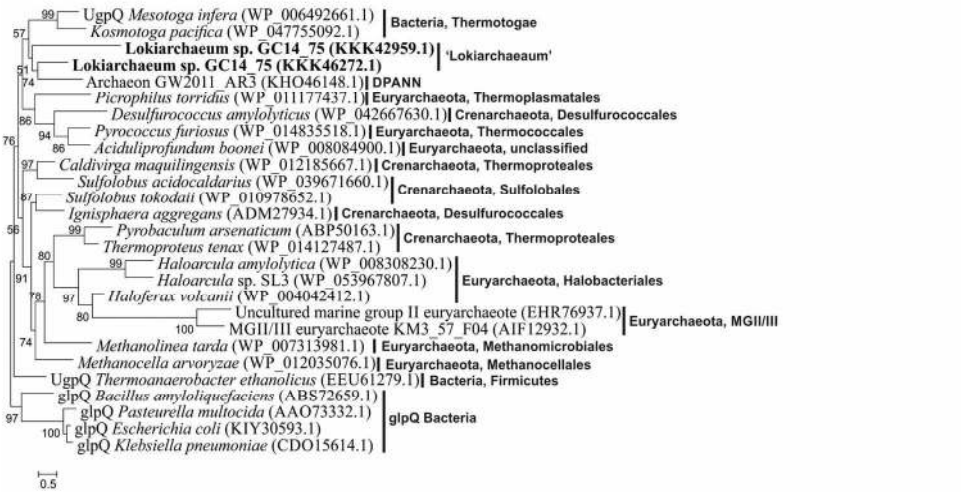


FIGURE 4. Phylogenetic tree of putative archaeal UgpQ glycerophosphodiester phosphodiesterases (GDPD) and close relatives within the Bacteria. The putative UgpQ GDPDs detected in the 'Lokiarchaeum' genome (Spang et al., 2015; indicated in bold) were closely related to a putative UgpQ GDPD in one genome of the DPANN Woesearchaeota, as well as UgpQ detected in bacterial genomes of the Thermotogae. Homologous proteins of the closely related family of periplasmic GlpQ GDPDs in Bacteria were used as outgroup. This tree was constructed using the maximum likelihood method with a LG model plus gamma distribution and invariant site (LG+G+I). The analysis included 553 positions in the final dataset. The scale bar represents number of amino acid substitutions per site. Branch support was calculated with the approximate likelihood ratio test (aLRT) and values ≥50% are indicated on the branches.

301x153mm (150 x 150 DPI)

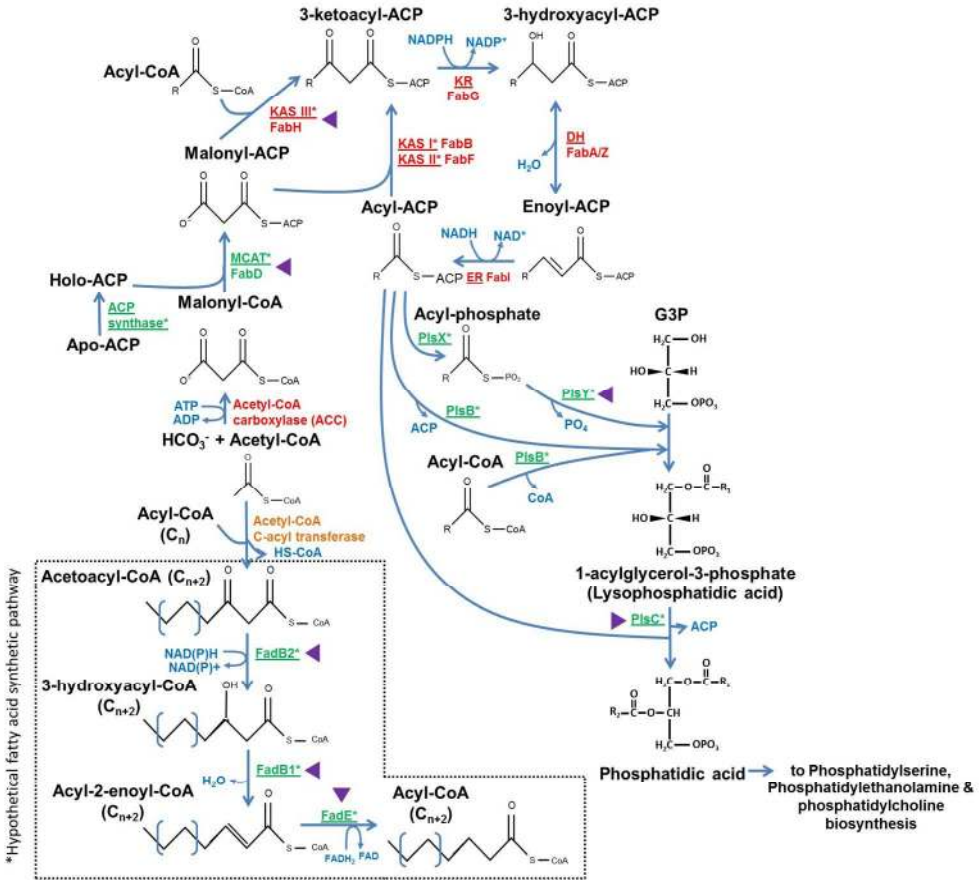


FIGURE 5. Bacterial biosynthetic pathway resulting in glycerol diester phospholipids formation. Fatty acids are synthesized from acyl-ACP via the FAS-II pathway and coupled with G3P to form phospholipids. The hypothetical archaeal fatty acid biosynthetic pathway proposed by Dibrova et al. (2014), based on the archaeal acetyl-CoA-acetyltransferase (acetyl-CoA C-acyl transferase; indicated in orange) and bacterial-type enzymes of the β-oxidation of fatty acids (acyl-CoA dehydrogenase, FadE; enoyl-CoA hydratase, FadB1; 3-hydroxyacyl-CoA dehydrogenase, FadB2) operating in reverse direction is indicated in the dashed line box. Enzymes indicated in red have been previously reported to be present as homologs in archaeal genomes, while enzymes in green indicate bacterial enzymes previously concluded to be absent in archaea. Enzymes discussed in the text are indicated with an asterisk. Purple triangles indicate putative homologs in the 'Lokiarchaeum' and/or MGII/III genomes (Table 1). Abbreviations: ACP, acyl-carrier protein; MCAT, malonyl-CoA:ACP-transacylase, FabD; KAS, beta-ketoacyl synthase (KAS I FabB; KAS II, FabF; KAS III, FabH); KR, beta-ketoacyl reductase, FabG; DH, beta-hydroxyacyl dehydratase, FabA/Z; ER, enoyl reductase, FabI; PlsB, glycerol-3-phosphate O-acyltransferase; PlsX, acyl-ACP:PO4 transacylase; PlsY, G3P acyltransferase; PlsC, 1-acylglycerol-3-phosphate O-acyltranferase; G3P, glycerol-3-phosphate.
327x288mm (150 x 150 DPI)

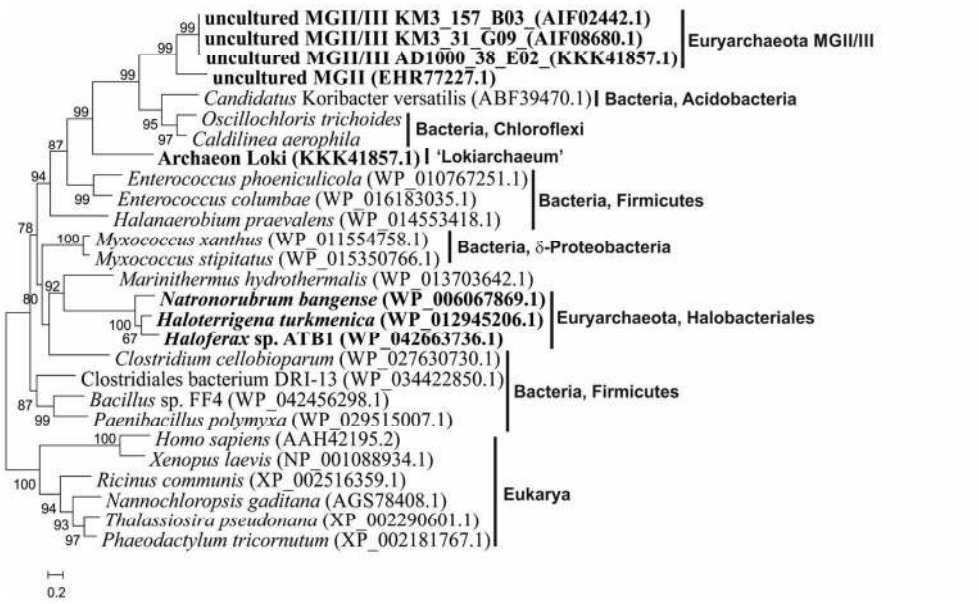


FIGURE 6. Phylogenetic tree of Malonyl-CoA:ACP-transacylase (MCAT, FabD domain), a key enzyme in the pathway of fatty acid synthesis (see Fig. 5 for details). MCAT homologs in archaeal genomes (in bold) and their closest bacterial sequences are shown. This phylogenetic tree was constructed using the maximum likelihood method and the LG model plus gamma distribution and invariant site (LG+G+I). The analysis included 486 positions in the final dataset. The scale bar represents number of amino acid substitutions per site. Branch support was calculated with the approximate likelihood ratio test (aLRT) and values ≥50% are indicated on the branches. MGII/III: uncultured marine group II/III euryarchaeota.

267x173mm (150 x 150 DPI)

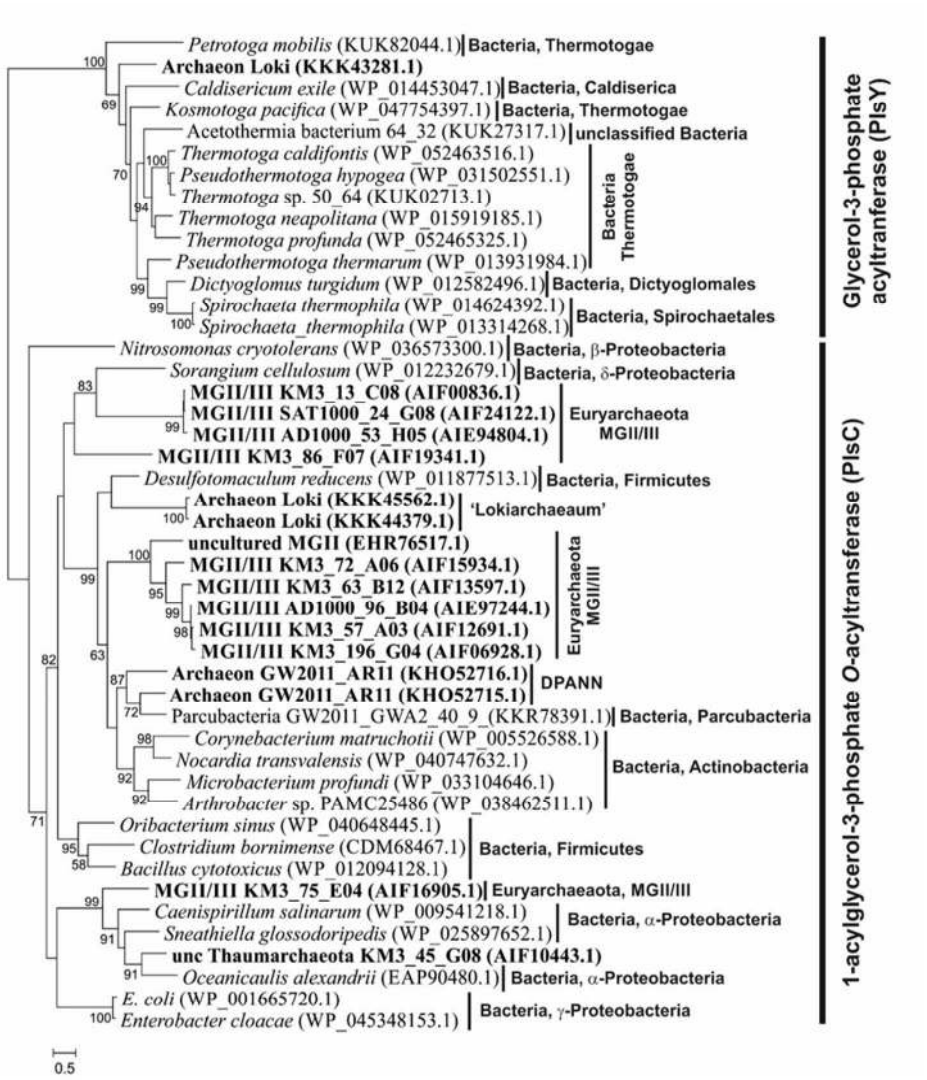


FIGURE 7. Phylogenetic tree of the putative archaeal homologs of the PlsY glycerol-3-phosphate acyltransferase and the PlsC 1-acylglycerol-3-phosphate O-acyltransferase (see Fig. 5 for details). PlsY and PlsC homologs in archaeal genomes (in bold) and their closest bacterial sequences are shown. This tree was constructed using the maximum likelihood method and the LG model plus gamma distribution and invariant site (LG+G+I). The scale bar represents number of substitutions per site. The analysis included 455 positions in the final dataset. Branch support was calculated with the approximate likelihood ratio test (aLRT) and values ≥50% are indicated on the branches. MGII/III: uncultured marine group II/III euryarchaeota.

170x186mm (150 x 150 DPI)

Figure 1

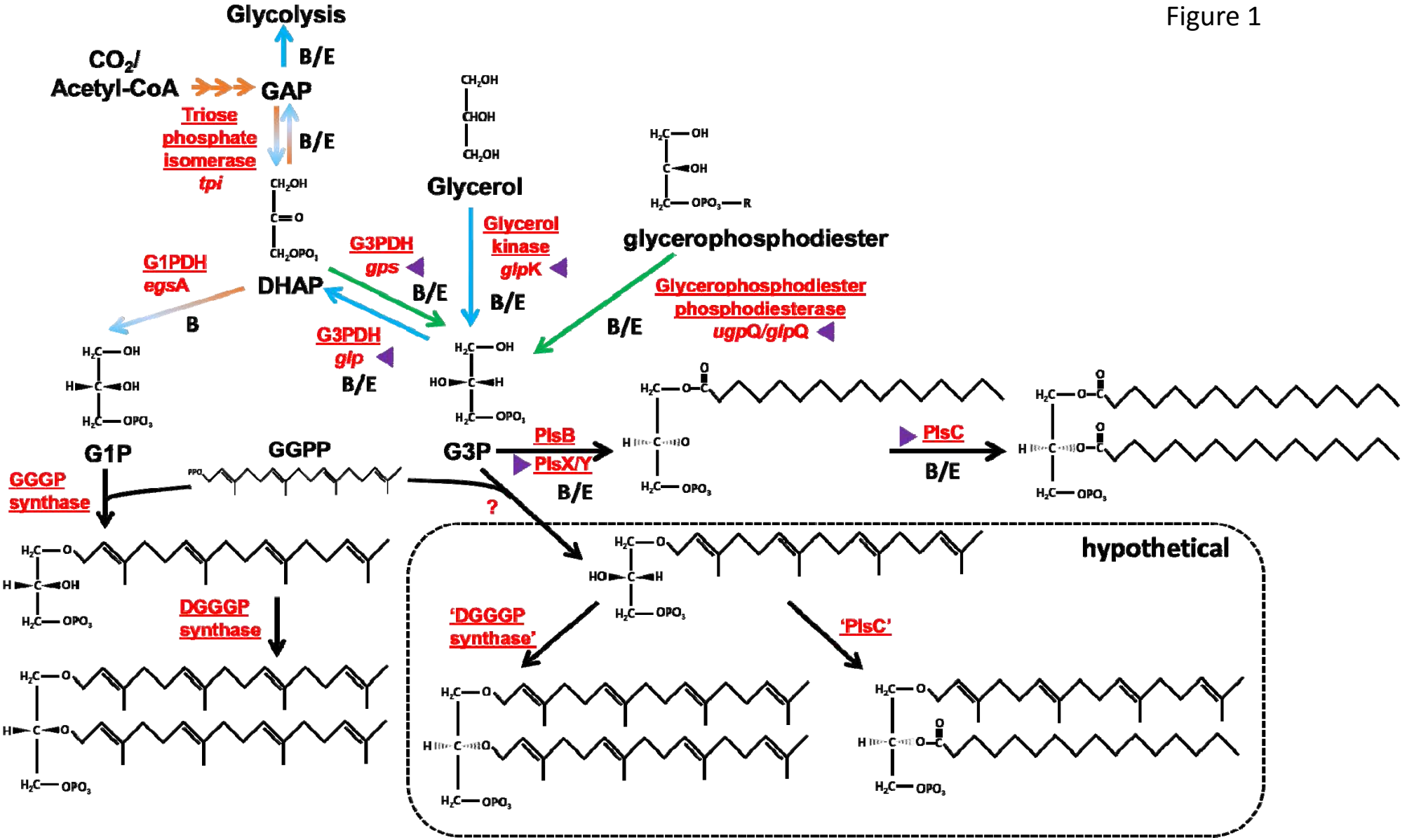
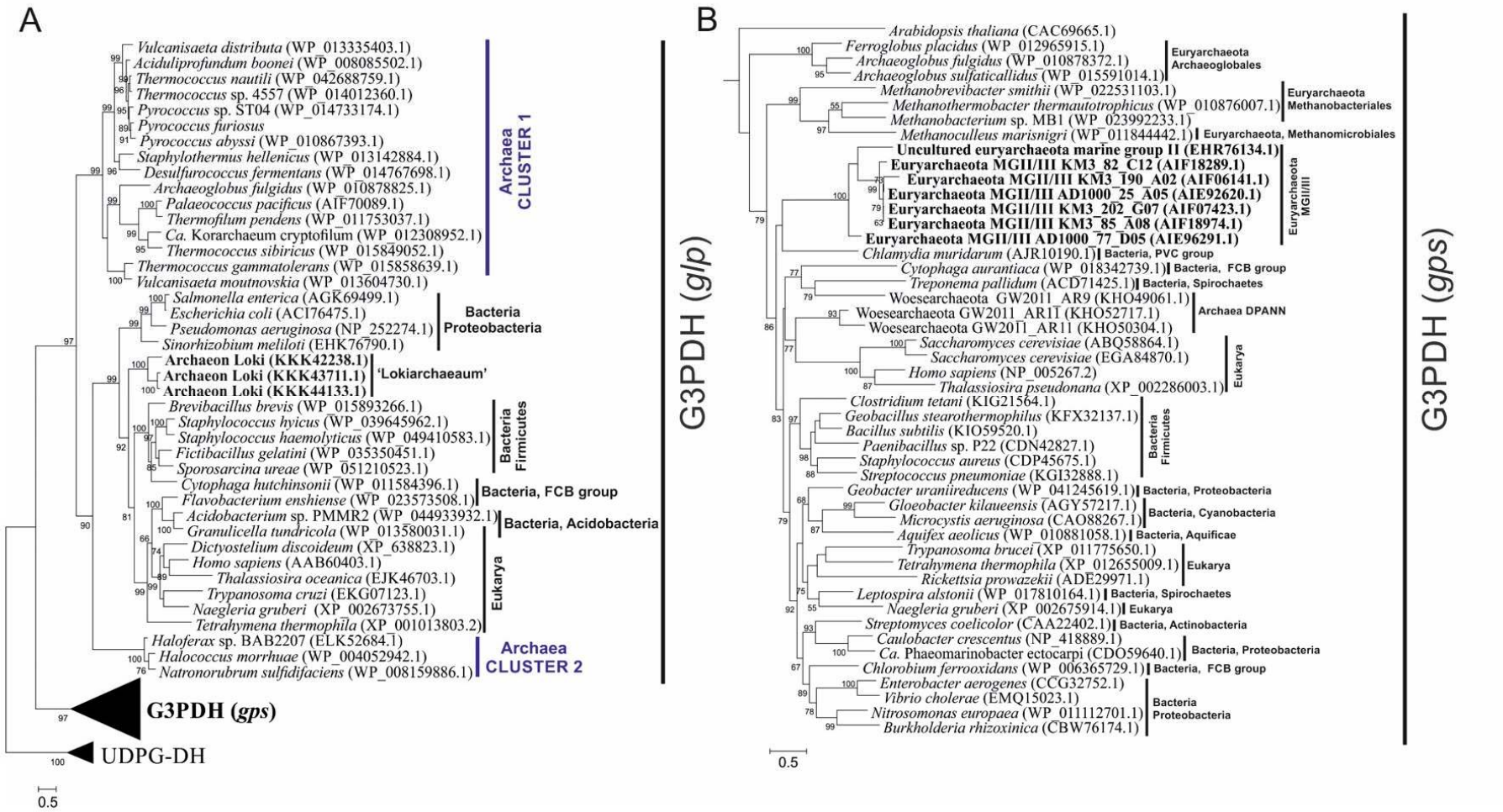

Glycolysis
B/E
CO2/
Acetyl-CoA
GAP
Triose
phosphate
isomerase
tpi
B/E
Glycerol
Glycerol
kinase
glpK
glycerophosphodiester
G3PDH
gps
G1PDH
egsA
DHAP
B/E
B
B/E
G3PDH
glp
B/E
Glycerophosphodiester
phosphodiesterase
ugpQ/glpQ
B/E
G1P
GGPP
G3P
PlsB
PlsX/Y
B/E
PlsC
B/E
GGGP
synthase
?
hypothetical
DGGGP
synthase
'DGGGP
synthase'
'PlsC'

Figure 2


A
Vulcanisaeta distributa (WP_013335403.1)
Aciduliprofundum boonei (WP_008085502.1)
Thermococcus nautili (WP_042688759.1)
Thermococcus sp. 4557 (WP_014012360.1)
Pyrococcus sp. ST04 (WP_014733174.1)
Pyrococcus furiosus
Pyrococcus abyssi (WP_010867393.1)
Staphylothermus hellenicus (WP_013142884.1)
Desulfurococcus fermentans (WP_014767698.1)
Archaeoglobus fulgidus (WP_010878825.1)
Palaeococcus pacificus (AIF70089.1)
Thermofilum pendens (WP_011753037.1)
Ca. Korarchaeum cryptofilum (WP_012308952.1)
Thermococcus sibiricus (WP_015849052.1)
Thermococcus gammatolerans (WP_015858639.1)
Vulcanisaeta moutnovskia (WP_013604730.1)
Archaea CLUSTER 1
Salmonella enterica (AGK69499.1)
Escherichia coli (ACI76475.1)
Pseudomonas aeruginosa (NP_252274.1)
Sinorhizobium meliloti (EHK76790.1)
Bacteria Proteobacteria
Archaeon Loki (KKK42238.1)
Archaeon Loki (KKK43711.1)
Archaeon Loki (KKK44133.1)
'Lokiarchaeaum'
Brevibacillus brevis (WP_015893266.1)
Staphylococcus hyicus (WP_039645962.1)
Staphylococcus haemolyticus (WP_049410583.1)
Fictibacillus gelatini (WP_035350451.1)
Sporosarcina ureae (WP_051210523.1)
Bacteria Firmicutes
Cytophaga hutchinsonii (WP_011584396.1)
Bacteria, FCB group
Flavobacterium enshiense (WP_023573508.1)
Acidobacterium sp. PMMR2 (WP_044933932.1)
Bacteria, Acidobacteria
Granulicella tundricola (WP_013580031.1)
Dictyostelium discoideum (XP_638823.1)
Homo sapiens (AAB60403.1)
Thalassiosira oceanica (EJK46703.1)
Trypanosoma cruzi (EKG07123.1)
Naegleria gruberi (XP_002673755.1)
Tetrahymena thermophila (XP_001013803.2)
Eukarya
Haloferax sp. BAB2207 (ELK52684.1)
Halococcus morrhuae (WP_004052942.1)
Natronorubrum sulfidifaciens (WP_008159886.1)
Archaea CLUSTER 2
G3PDH (gps)
UDPG-DH
0.5
G3PDH (g/p)
B
Arabidopsis thaliana (CAC69665.1)
Ferroglobus placidus (WP_012965915.1)
Archaeoglobus fulgidus (WP_010878372.1)
Archaeoglobus sulfaticallidus (WP_015591014.1)
Euryarchaeota Archaeoglobales
Methanobrevibacter smithii (WP_022531103.1)
Methanothermobacter thermautotrophicus (WP_010876007.1)
Methanobacterium sp. MB1 (WP_023992233.1)
Euryarchaeota Methanobacteriales
Methanoculleus marisnigri (WP_011844442.1)
Euryarchaeota, Methanomicrobiales
Uncultured euryarchaeota marine group II (EHR76134.1)
Euryarchaeota MGII/III KM3_82_C12 (AIF18289.1)
Euryarchaeota MGII/III KM3_190_A02 (AIF06141.1)
Euryarchaeota MGII/III AD1000_25_A05 (AIE92620.1)
Euryarchaeota MGII/III KM3_202_G07 (AIF07423.1)
Euryarchaeota MGII/III KM3_85_A08 (AIF18974.1)
Euryarchaeota MGII/III AD1000_77_D05 (AIE96291.1)
Euryarchaeota MGII/III
Chlamydia muridarum (AJR10190.1)
Bacteria, PVC group
Cytophaga aurantiaca (WP_018342739.1)
Bacteria, FCB group
Treponema pallidum (ACD71425.1)
Bacteria, Spirochaetes
Woesearchaeota GW2011_AR9 (KHO49061.1)
Woesearchaeota GW2011_AR10 (KHO52717.1)
Woesearchaeota GW2011_AR11 (KHO50304.1)
Archaea DPANN
Saccharomyces cerevisiae (ABQ58864.1)
Saccharomyces cerevisiae (EGA84870.1)
Homo sapiens (NP_005267.2)
Thalassiosira pseudonana (XP_002286003.1)
Eukarya
Clostridium tetani (KIG21564.1)
Geobacillus stearothermophilus (KFX32137.1)
Bacillus subtilis (KIO59520.1)
Paenibacillus sp. P22 (CDN42827.1)
Staphylococcus aureus (CDP45675.1)
Streptococcus pneumoniae (KGI32888.1)
Bacteria Firmicutes
Geobacter uraniireducens (WP_041245619.1)
Bacteria, Proteobacteria
Gloeobacter kilaueensis (AGY57217.1)
Microcystis aeruginosa (CAO88267.1)
Bacteria, Cyanobacteria
Aquifex aeolicus (WP_010881058.1)
Bacteria, Aquificae
Trypanosoma brucei (XP_011775650.1)
Tetrahymena thermophila (XP_012655009.1)
Rickettsia prowazekii (ADE29971.1)
Eukarya
Leptospira alstonii (WP_017810164.1)
Bacteria, Spirochaetes
Naegleria gruberi (XP_002675914.1)
Eukarya
Streptomyces coelicolor (CAA22402.1)
Bacteria, Actinobacteria
Caulobacter crescentus (NP_418889.1)
Ca. Phaeomarinobacter ectocarpi (CDO59640.1)
Bacteria, Proteobacteria
Chlorobium ferrooxidans (WP_006365729.1)
Bacteria, FCB group
Enterobacter aerogenes (CCG32752.1)
Vibrio cholerae (EMQ15023.1)
Nitrosomonas europaea (WP_011112701.1)
Burkholderia rhizoxinica (CBW76174.1)
Bacteria Proteobacteria
0.5
G3PDH (gps)

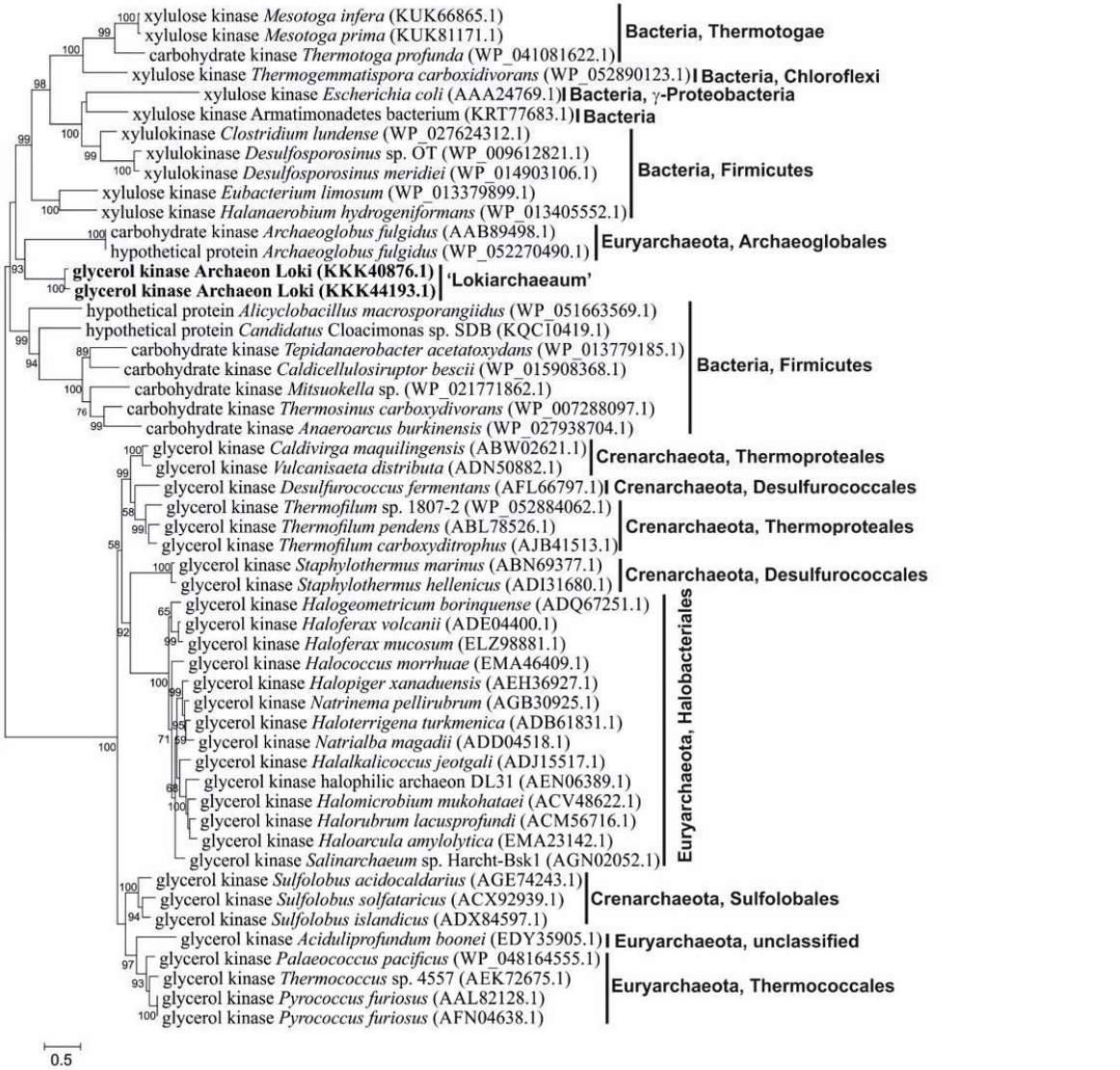

xylulose kinase *Mesotoga infera* (KUK66865.1)
xylulose kinase *Mesotoga prima* (KUK81171.1)
carbohydrate kinase *Thermotoga profunda* (WP_041081622.1)
Bacteria, Thermotogae
xylulose kinase *Thermogemmatispora carboxidivorans* (WP_052890123.1)
Bacteria, Chloroflexi
xylulose kinase *Escherichia coli* (AAA24769.1)
Bacteria, γ-Proteobacteria
xylulose kinase Armatimonadetes bacterium (KRT77683.1)
Bacteria
xylulokinase *Clostridium lundense* (WP_027624312.1)
xylulokinase *Desulfosporosinus* sp. OT (WP_009612821.1)
xylulokinase *Desulfosporosinus meridiei* (WP_014903106.1)
xylulose kinase *Eubacterium limosum* (WP_013379899.1)
xylulose kinase *Halanaerobium hydrogeniformans* (WP_013405552.1)
Bacteria, Firmicutes
carbohydrate kinase *Archaeoglobus fulgidus* (AAB89498.1)
hypothetical protein *Archaeoglobus fulgidus* (WP_052270490.1)
Euryarchaeota, Archaeoglobales
glycerol kinase Archaeon Loki (KKK40876.1)
glycerol kinase Archaeon Loki (KKK44193.1)
'Lokiarchaeaum'
hypothetical protein *Alicyclobacillus macrosporangiidus* (WP_051663569.1)
hypothetical protein *Candidatus* Cloacimonas sp. SDB (KQC10419.1)
carbohydrate kinase *Tepidanaerobacter acetatoxydans* (WP_013779185.1)
carbohydrate kinase *Caldicellulosiruptor bescii* (WP_015908368.1)
carbohydrate kinase *Mitsuokella* sp. (WP_021771862.1)
carbohydrate kinase *Thermosinus carboxydivorans* (WP_007288097.1)
carbohydrate kinase *Anaeroarcus burkinensis* (WP_027938704.1)
Bacteria, Firmicutes
glycerol kinase *Caldivirga maquilingensis* (ABW02621.1)
glycerol kinase *Vulcanisaeta distributa* (ADN50882.1)
Crenarchaeota, Thermoproteales
glycerol kinase *Desulfurococcus fermentans* (AFL66797.1)
Crenarchaeota, Desulfurococcales
glycerol kinase *Thermofilum* sp. 1807-2 (WP_052884062.1)
glycerol kinase *Thermofilum pendens* (ABL78526.1)
glycerol kinase *Thermofilum carboxyditrophus* (AJB41513.1)
Crenarchaeota, Thermoproteales
glycerol kinase *Staphylothermus marinus* (ABN69377.1)
glycerol kinase *Staphylothermus hellenicus* (ADI31680.1)
Crenarchaeota, Desulfurococcales
glycerol kinase *Halogeometricum borinquense* (ADQ67251.1)
glycerol kinase *Haloferax volcanii* (ADE04400.1)
glycerol kinase *Haloferax mucosum* (ELZ98881.1)
glycerol kinase *Halococcus morrhuae* (EMA46409.1)
glycerol kinase *Halopiger xanaduensis* (AEH36927.1)
glycerol kinase *Natrinema pellirubrum* (AGB30925.1)
glycerol kinase *Haloterrigena turkmenica* (ADB61831.1)
glycerol kinase *Natrialba magadii* (ADD04518.1)
glycerol kinase *Halalkalicoccus jeotgali* (ADJ15517.1)
glycerol kinase halophilic archaeon DL31 (AEN06389.1)
glycerol kinase *Halomicrobium mukohataei* (ACV48622.1)
glycerol kinase *Halorubrum lacusprofundi* (ACM56716.1)
glycerol kinase *Haloarcula amylolytica* (EMA23142.1)
glycerol kinase *Salinarchaeum* sp. Harcht-Bsk1 (AGN02052.1)
Euryarchaeota, Halobacteriales
glycerol kinase *Sulfolobus acidocaldarius* (AGE74243.1)
glycerol kinase *Sulfolobus solfataricus* (ACX92939.1)
glycerol kinase *Sulfolobus islandicus* (ADX84597.1)
Crenarchaeota, Sulfolobales
glycerol kinase *Aciduliprofundum boonei* (EDY35905.1)
Euryarchaeota, unclassified
glycerol kinase *Palaeococcus pacificus* (WP_048164555.1)
glycerol kinase *Thermococcus* sp. 4557 (AEK72675.1)
glycerol kinase *Pyrococcus furiosus* (AAL82128.1)
glycerol kinase *Pyrococcus furiosus* (AFN04638.1)
Euryarchaeota, Thermococcales
0.5

Figure 3

Figure 4

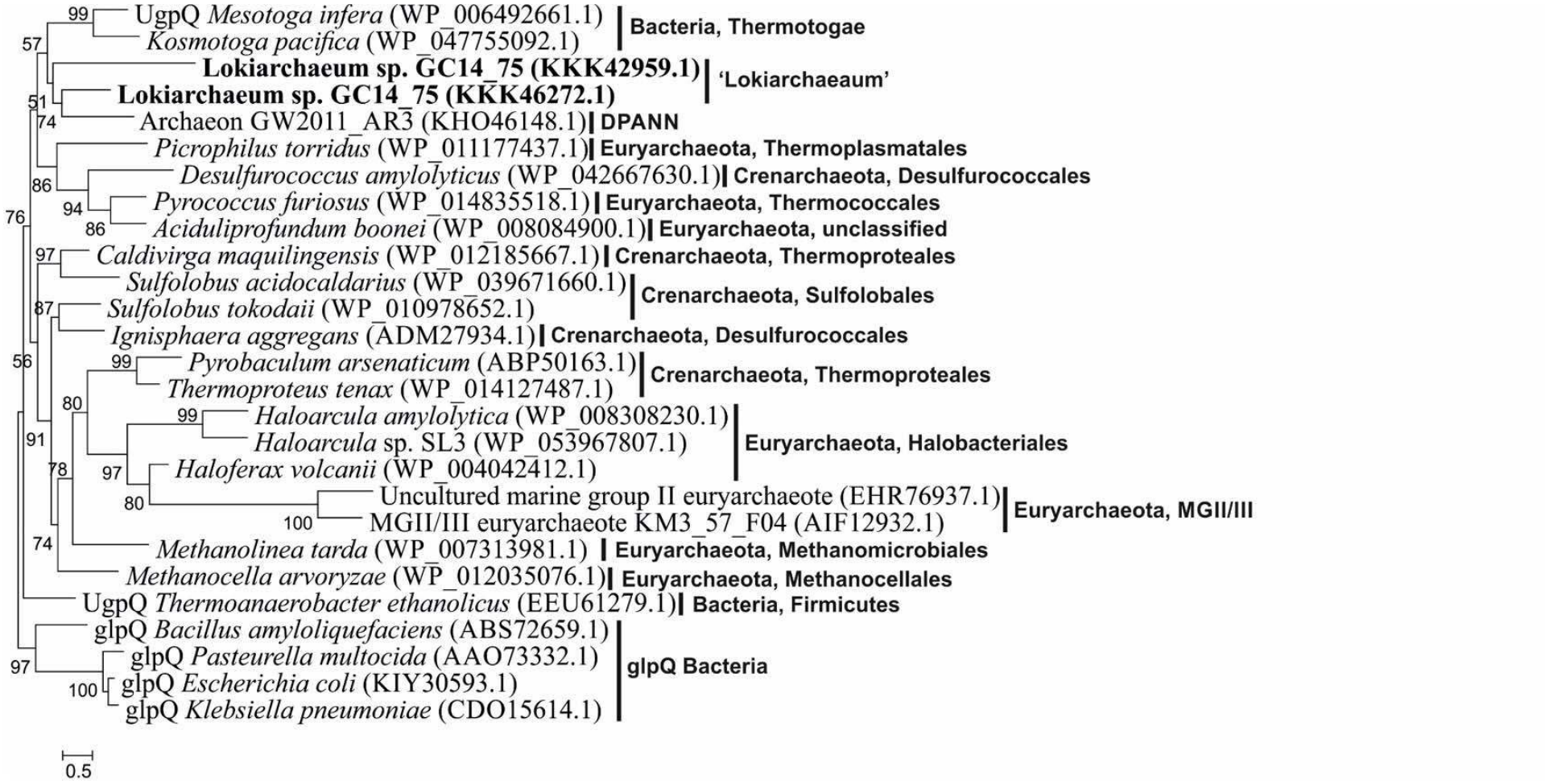

UgpQ Mesotoga infera (WP_006492661.1)
Kosmotoga pacifica (WP_047755092.1)
Bacteria, Thermotogae
Lokiarchaeum sp. GC14_75 (KKK42959.1)
Lokiarchaeum sp. GC14_75 (KKK46272.1)
'Lokiarchaeaum'
Archaeon GW2011_AR3 (KHO46148.1)
DPANN
Picrophilus torridus (WP_011177437.1)
Euryarchaeota, Thermoplasmatales
Desulfurococcus amylolyticus (WP_042667630.1)
Crenarchaeota, Desulfurococcales
Pyrococcus furiosus (WP_014835518.1)
Euryarchaeota, Thermococcales
Aciduliprofundum boonei (WP_008084900.1)
Euryarchaeota, unclassified
Caldivirga maquilingensis (WP_012185667.1)
Crenarchaeota, Thermoproteales
Sulfolobus acidocaldarius (WP_039671660.1)
Sulfolobus tokodaii (WP_010978652.1)
Crenarchaeota, Sulfolobales
Ignisphaera aggregans (ADM27934.1)
Crenarchaeota, Desulfurococcales
Pyrobaculum arsenaticum (ABP50163.1)
Thermoproteus tenax (WP_014127487.1)
Crenarchaeota, Thermoproteales
Haloarcula amylolytica (WP_008308230.1)
Haloarcula sp. SL3 (WP_053967807.1)
Haloferax volcanii (WP_004042412.1)
Euryarchaeota, Halobacteriales
Uncultured marine group II euryarchaeote (EHR76937.1)
MGII/III euryarchaeote KM3_57_F04 (AIF12932.1)
Euryarchaeota, MGII/III
Methanolinea tarda (WP_007313981.1)
Euryarchaeota, Methanomicrobiales
Methanocella arvoryzae (WP_012035076.1)
Euryarchaeota, Methanocellales
UgpQ Thermoanaerobacter ethanolicus (EEU61279.1)
Bacteria, Firmicutes
glpQ Bacillus amyloliquefaciens (ABS72659.1)
glpQ Pasteurella multocida (AAO73332.1)
glpQ Escherichia coli (KIY30593.1)
glpQ Klebsiella pneumoniae (CDO15614.1)
glpQ Bacteria
0.5

Figure 5

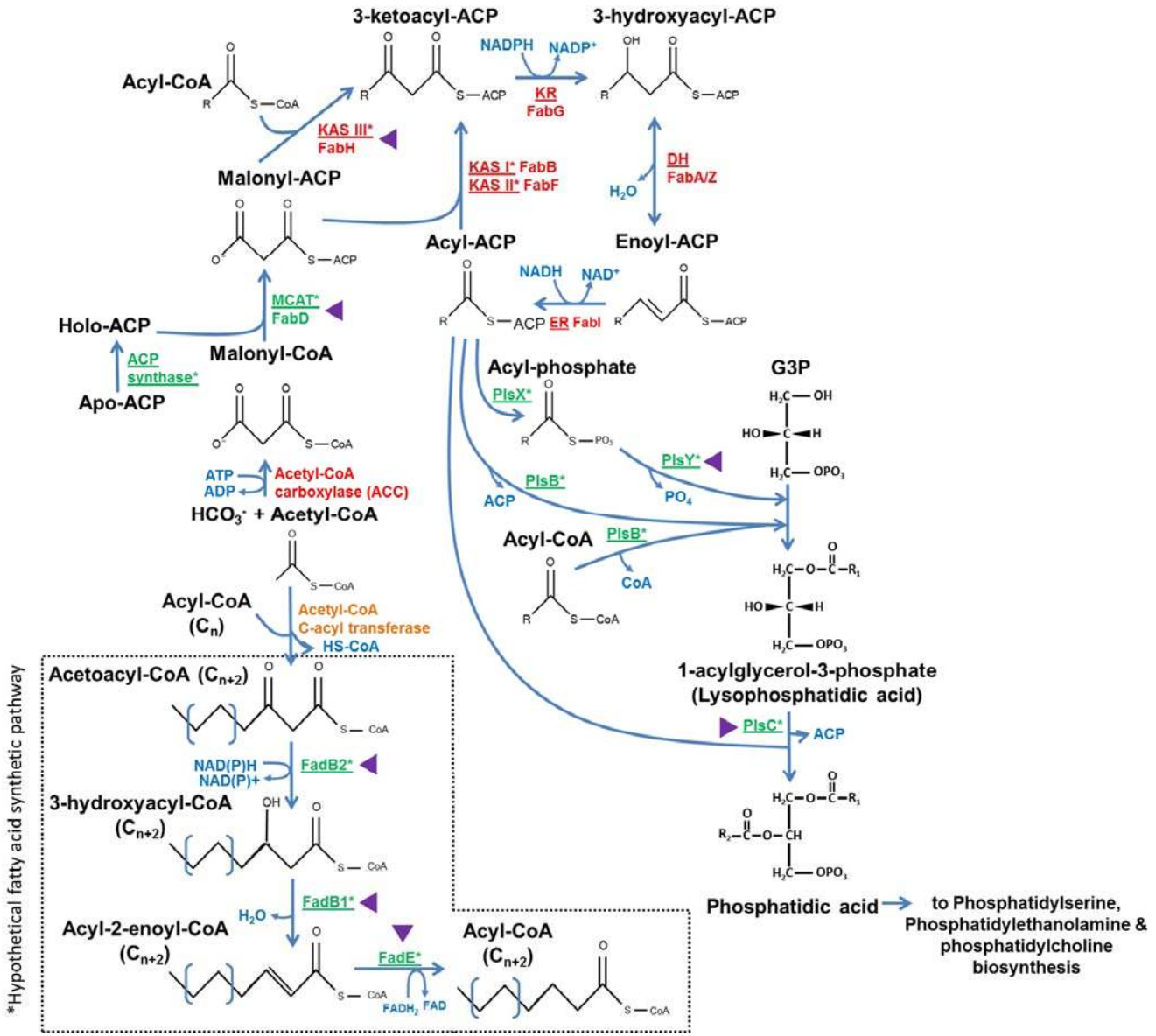
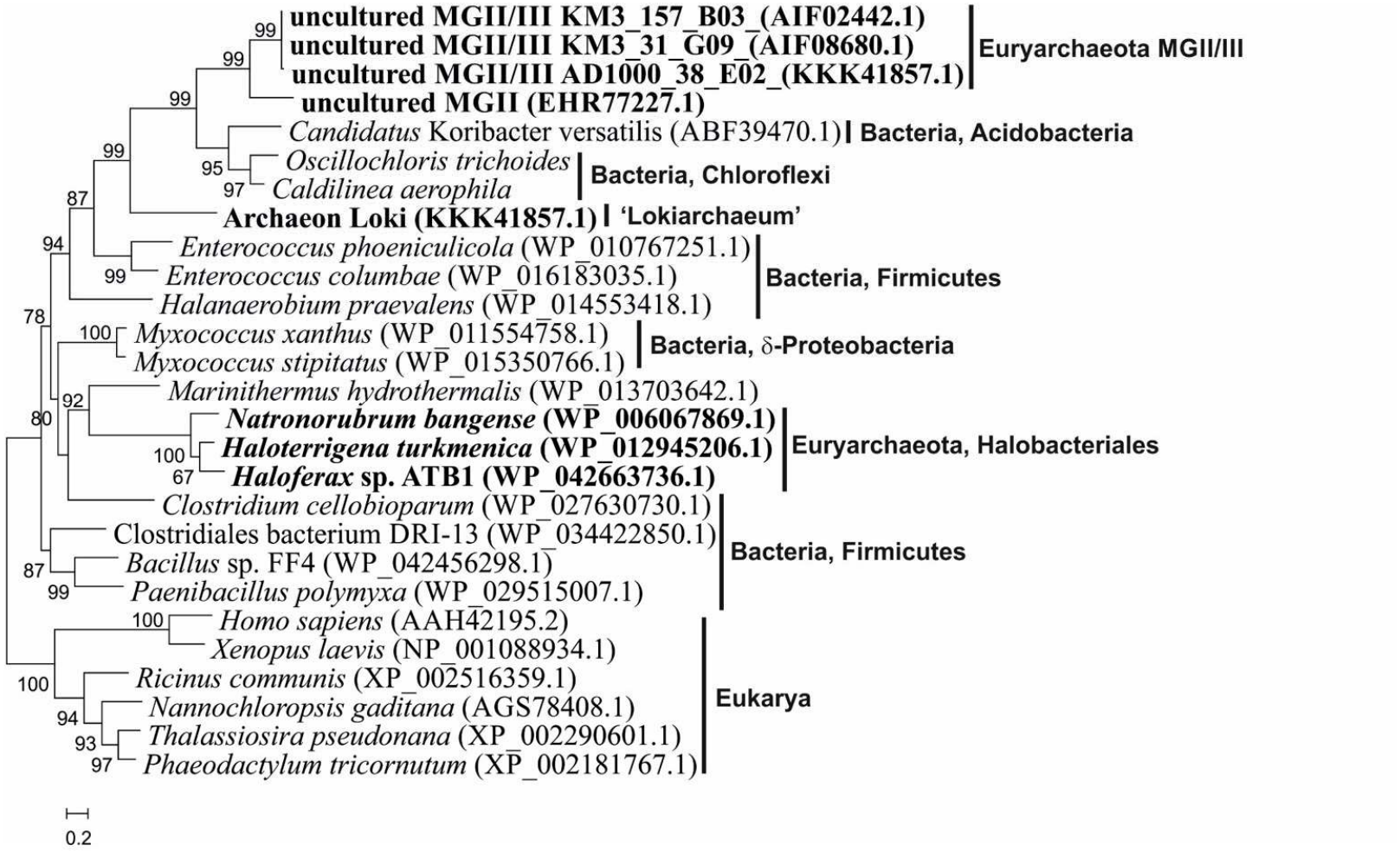

3-ketoacyl-ACP
3-hydroxyacyl-ACP
Acyl-CoA
NADPH
NADP+
KR
FabG
KAS III*
FabH
Malonyl-ACP
KAS I* FabB
KAS II* FabF
DH
FabA/Z
H2O
Acyl-ACP
Enoyl-ACP
NADH
NAD+
ER FabI
MCAT*
FabD
Holo-ACP
Malonyl-CoA
ACP
synthase*
Apo-ACP
Acyl-phosphate
G3P
PlsX*
PlsY*
PO4
ACP
PlsB*
Acyl-CoA
CoA
ATP
ADP
Acetyl-CoA
carboxylase (ACC)
HCO3- + Acetyl-CoA
Acyl-CoA
(Cn)
Acetyl-CoA
C-acyl transferase
HS-CoA
1-acylglycerol-3-phosphate
(Lysophosphatidic acid)
PlsC*
ACP
Acetoacyl-CoA (Cn+2)
NAD(P)H
NAD(P)+
FadB2*
3-hydroxyacyl-CoA
(Cn+2)
FadB1*
H2O
Acyl-2-enoyl-CoA
(Cn+2)
FadE*
FADH2
FAD
Acyl-CoA
(Cn+2)
Phosphatidic acid
to Phosphatidylserine,
Phosphatidylethanolamine &
phosphatidylcholine
biosynthesis
*Hypothetical fatty acid synthetic pathway

Figure 6


uncultured MGII/III KM3_157_B03_(AIF02442.1)
uncultured MGII/III KM3_31_G09_(AIF08680.1)
uncultured MGII/III AD1000_38_E02_(KKK41857.1)
Euryarchaeota MGII/III
uncultured MGII (EHR77227.1)
Candidatus Koribacter versatilis (ABF39470.1)
Bacteria, Acidobacteria
Oscillochloris trichoides
Caldilinea aerophila
Bacteria, Chloroflexi
Archaeon Loki (KKK41857.1)
'Lokiarchaeum'
Enterococcus phoeniculicola (WP_010767251.1)
Enterococcus columbae (WP_016183035.1)
Halanaerobium praevalens (WP_014553418.1)
Bacteria, Firmicutes
Myxococcus xanthus (WP_011554758.1)
Myxococcus stipitatus (WP_015350766.1)
Bacteria, δ-Proteobacteria
Marinithermus hydrothermalis (WP_013703642.1)
Natronorubrum bangense (WP_006067869.1)
Haloterrigena turkmenica (WP_012945206.1)
Haloferax sp. ATB1 (WP_042663736.1)
Euryarchaeota, Halobacteriales
Clostridium cellobioparum (WP_027630730.1)
Clostridiales bacterium DRI-13 (WP_034422850.1)
Bacillus sp. FF4 (WP_042456298.1)
Paenibacillus polymyxa (WP_029515007.1)
Bacteria, Firmicutes
Homo sapiens (AAH42195.2)
Xenopus laevis (NP_001088934.1)
Ricinus communis (XP_002516359.1)
Nannochloropsis gaditana (AGS78408.1)
Thalassiosira pseudonana (XP_002290601.1)
Phaeodactylum tricornutum (XP_002181767.1)
Eukarya
0.2

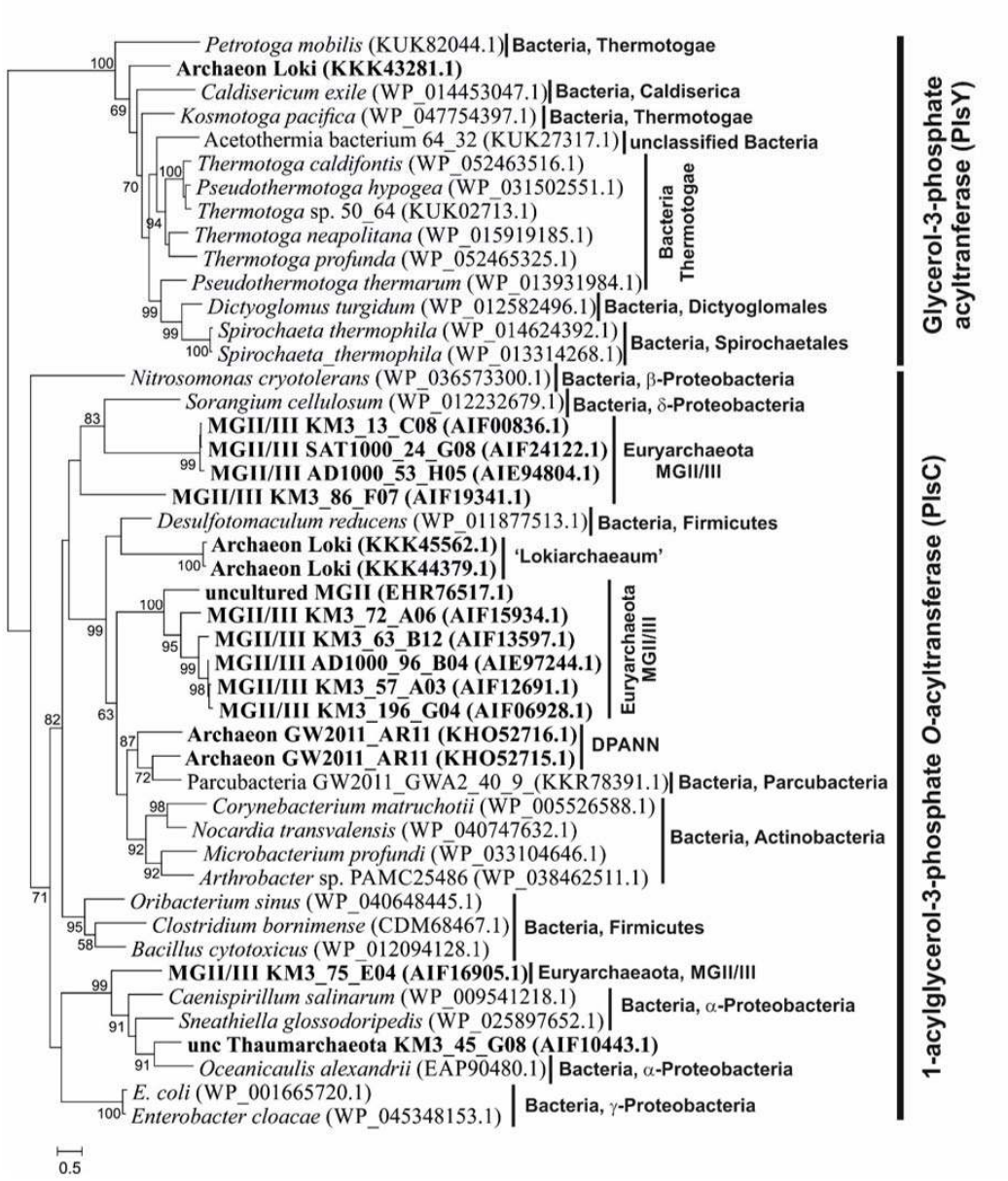
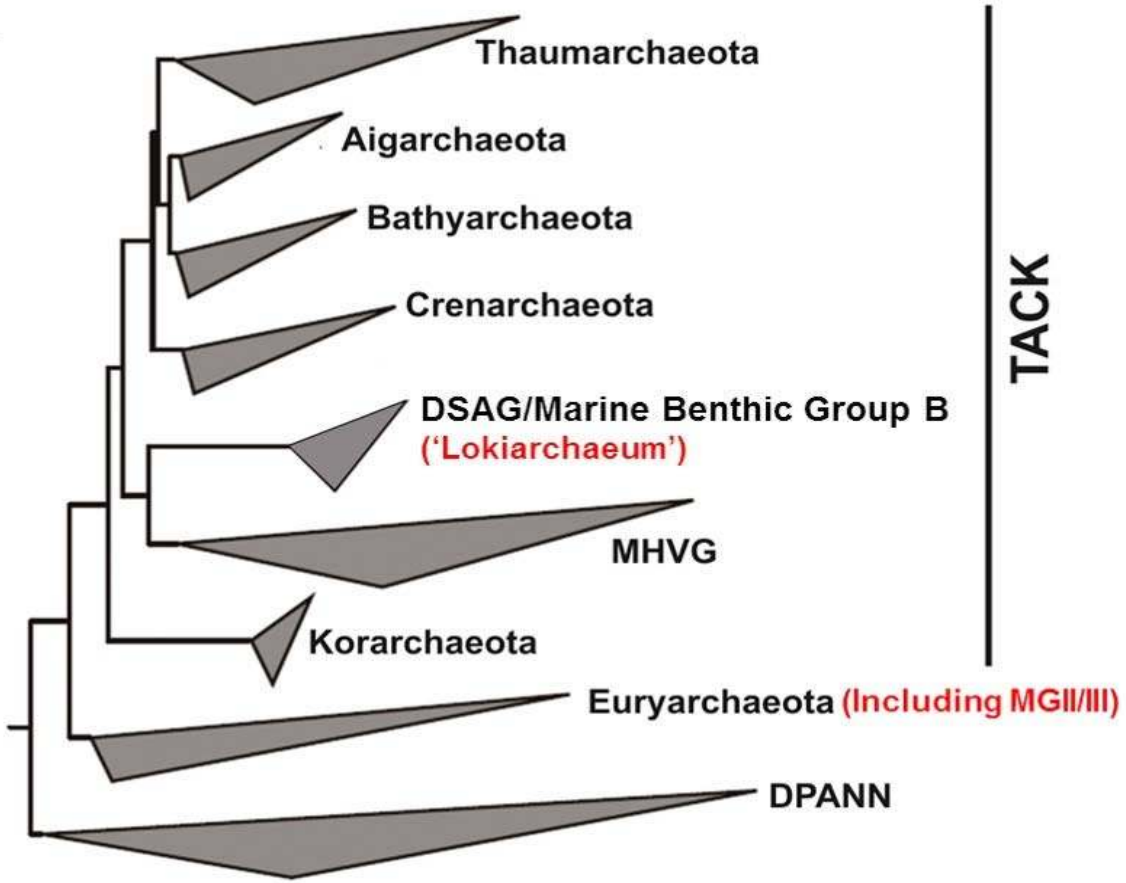

Petrotoga mobilis (KUK82044.1) | Bacteria, Thermotogae
Archaeon Loki (KKK43281.1)
Caldisericum exile (WP_014453047.1) | Bacteria, Caldiserica
Kosmotoga pacifica (WP_047754397.1) | Bacteria, Thermotogae
Acetothermia bacterium 64_32 (KUK27317.1) | unclassified Bacteria
Thermotoga caldifontis (WP_052463516.1)
Pseudothermotoga hypogea (WP_031502551.1)
Thermotoga sp. 50_64 (KUK02713.1)
Thermotoga neapolitana (WP_015919185.1)
Thermotoga profunda (WP_052465325.1)
Pseudothermotoga thermarum (WP_013931984.1)
Bacteria Thermotogae
Dictyoglomus turgidum (WP_012582496.1) | Bacteria, Dictyoglomales
Spirochaeta thermophila (WP_014624392.1)
Spirochaeta_thermophila (WP_013314268.1) | Bacteria, Spirochaetales
Glycerol-3-phosphate acyltranferase (PlsY)
Nitrosomonas cryotolerans (WP_036573300.1) | Bacteria, β-Proteobacteria
Sorangium cellulosum (WP_012232679.1) | Bacteria, δ-Proteobacteria
MGII/III KM3_13_C08 (AIF00836.1)
MGII/III SAT1000_24_G08 (AIF24122.1)
MGII/III AD1000_53_H05 (AIE94804.1)
Euryarchaeota MGII/III
MGII/III KM3_86_F07 (AIF19341.1)
Desulfotomaculum reducens (WP_011877513.1) | Bacteria, Firmicutes
Archaeon Loki (KKK45562.1)
Archaeon Loki (KKK44379.1) | 'Lokiarchaeaum'
uncultured MGII (EHR76517.1)
MGII/III KM3_72_A06 (AIF15934.1)
MGII/III KM3_63_B12 (AIF13597.1)
MGII/III AD1000_96_B04 (AIE97244.1)
MGII/III KM3_57_A03 (AIF12691.1)
MGII/III KM3_196_G04 (AIF06928.1)
Euryarchaeota MGII/III
Archaeon GW2011_AR11 (KHO52716.1)
Archaeon GW2011_AR11 (KHO52715.1)
DPANN
Parcubacteria GW2011_GWA2_40_9_(KKR78391.1) | Bacteria, Parcubacteria
Corynebacterium matruchotii (WP_005526588.1)
Nocardia transvalensis (WP_040747632.1)
Microbacterium profundi (WP_033104646.1)
Arthrobacter sp. PAMC25486 (WP_038462511.1)
Bacteria, Actinobacteria
Oribacterium sinus (WP_040648445.1)
Clostridium bornimense (CDM68467.1)
Bacillus cytotoxicus (WP_012094128.1)
Bacteria, Firmicutes
MGII/III KM3_75_E04 (AIF16905.1) | Euryarchaeaota, MGII/III
Caenispirillum salinarum (WP_009541218.1)
Sneathiella glossodoripedis (WP_025897652.1)
Bacteria, α-Proteobacteria
unc Thaumarchaeota KM3_45_G08 (AIF10443.1)
Oceanicaulis alexandrii (EAP90480.1) | Bacteria, α-Proteobacteria
E. coli (WP_001665720.1)
Enterobacter cloacae (WP_045348153.1) | Bacteria, γ-Proteobacteria
1-acylglycerol-3-phosphate O-acyltransferase (PlsC)
0.5

Figure 7

Figure S1


Thaumarchaeota
Aigarchaeota
Bathyarchaeota
Crenarchaeota
DSAG/Marine Benthic Group B
('Lokiarchaeum')
MHVG
Korarchaeota
Euryarchaeota (Including MGII/III)
DPANN
TACK

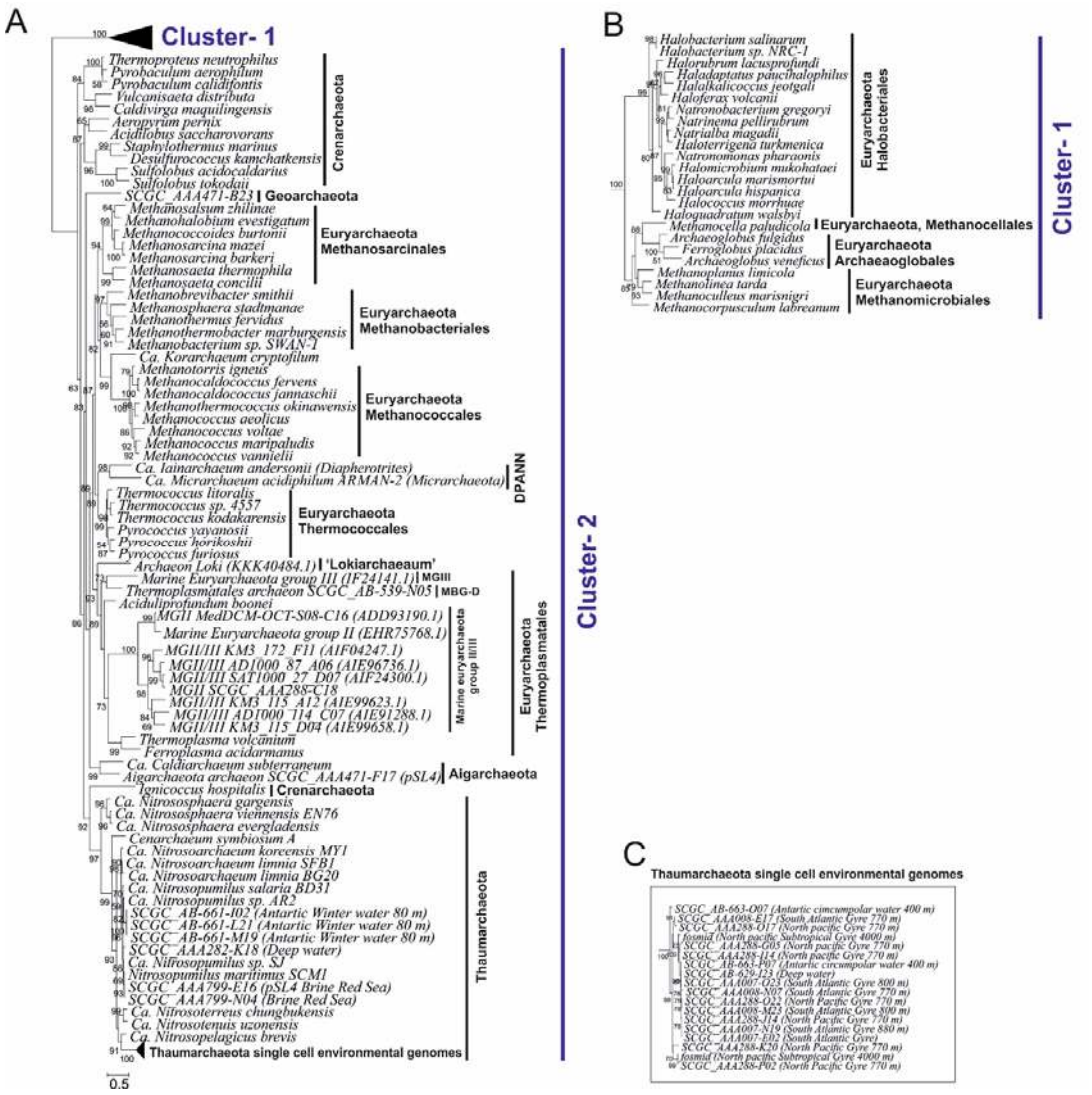
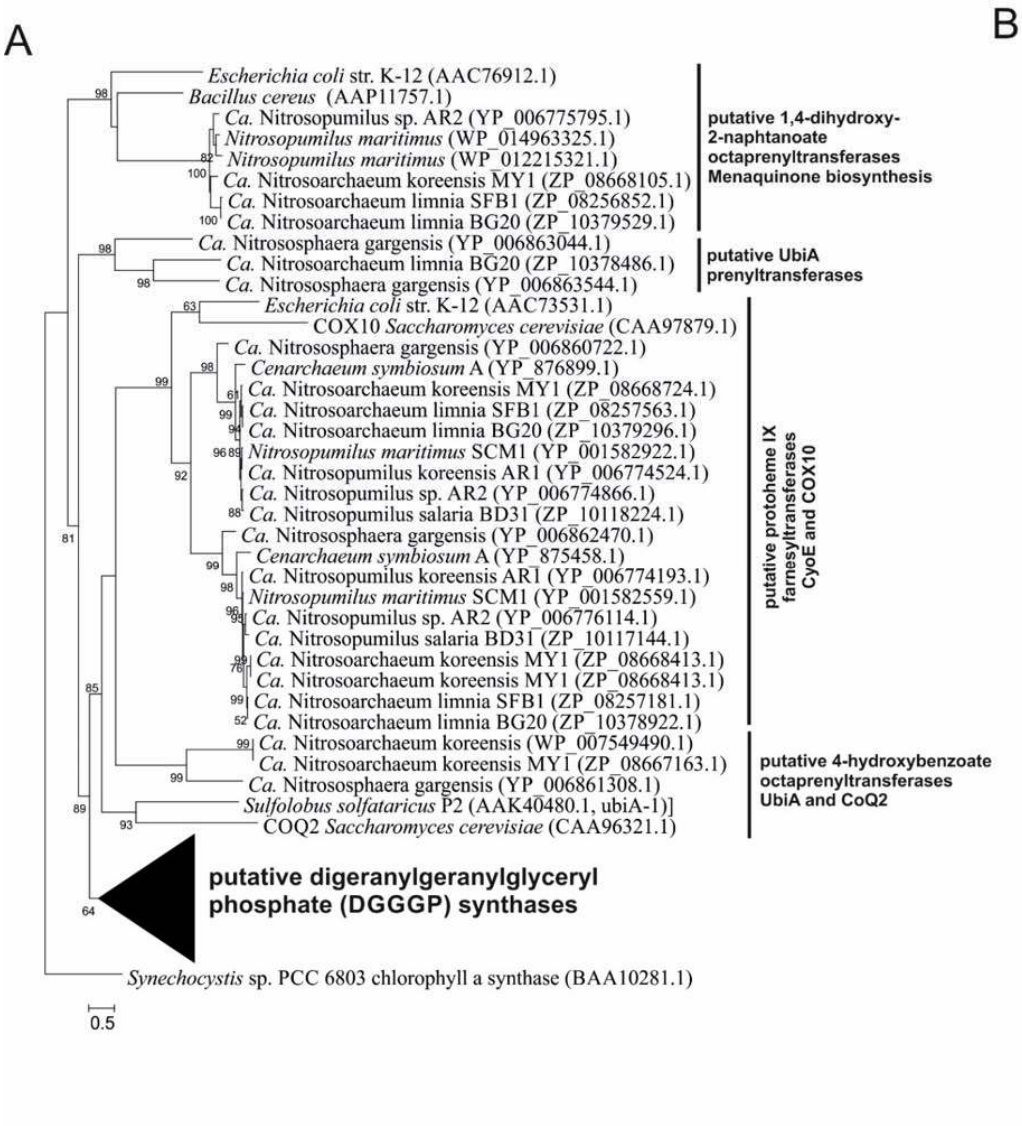
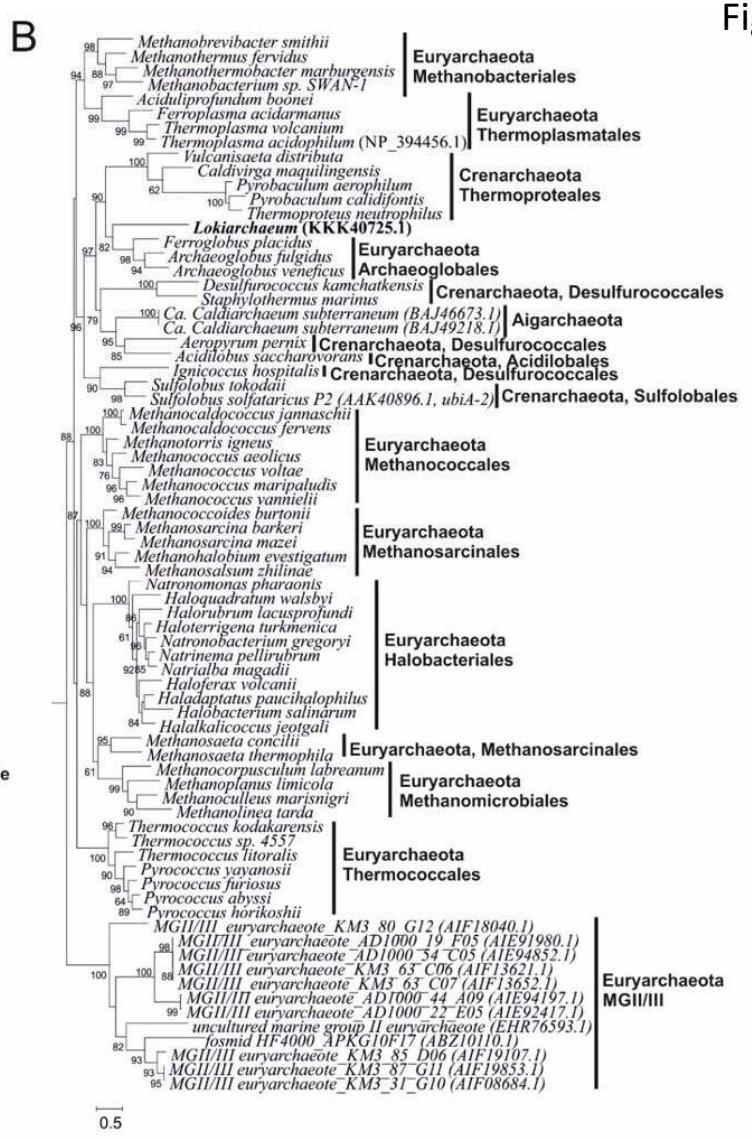

A
Cluster- 1
Thermoproteus neutrophilus
Pyrobaculum aerophilum
Pyrobaculum calidifontis
Vulcanisaeta distributa
Caldivirga maquilingensis
Aeropyrum pernix
Acidilobus saccharovorans
Staphylothermus marinus
Desulfurococcus kamchatkensis
Sulfolobus acidocaldarius
Sulfolobus tokodaii
Crenarchaeota
SCGC_AAA471-B23 | Geoarchaeota
Methanosalsum zhilinae
Methanohalobium evestigatum
Methanococcoides burtonii
Methanosarcina mazei
Methanosarcina barkeri
Methanosaeta thermophila
Methanosaeta concilii
Euryarchaeota
Methanosarcinales
Methanobrevibacter smithii
Methanosphaera stadtmanae
Methanothermus fervidus
Methanothermobacter marburgensis
Methanobacterium sp. SWAN-1
Euryarchaeota
Methanobacteriales
Ca. Korarchaeum cryptofilum
Methanotorris igneus
Methanocaldococcus fervens
Methanocaldococcus jannaschii
Methanothermococcus okinawensis
Methanococcus aeolicus
Methanococcus voltae
Methanococcus maripaludis
Methanococcus vannielii
Euryarchaeota
Methanococcales
Ca. Iainarchaeum andersonii (Diapherotrites)
Ca. Micrarchaeum acidiphilum ARMAN-2 (Micrarchaeota)
DPANN
Thermococcus litoralis
Thermococcus sp. 4557
Thermococcus kodakarensis
Pyrococcus yayanosii
Pyrococcus horikoshii
Pyrococcus furiosus
Euryarchaeota
Thermococcales
Archaeon Loki (KKK40484.1) | 'Lokiarchaeaum'
Marine Euryarchaeota group III (IF24141.1) | MGIII
Thermoplasmatales archaeon SCGC_AB-539-N05 | MBG-D
Aciduliprofundum boonei
MGII MedDCM-OCT-S08-C16 (ADD93190.1)
Marine Euryarchaeota group II (EHR75768.1)
MGII/III KM3_172_F11 (AIF04247.1)
MGII/III_AD1000_87_A06 (AIE96736.1)
MGII/III SAT1000_27_D07 (AIF24300.1)
MGII SCGC_AAA288-C18
MGII/III KM3_115_A12 (AIE99623.1)
MGII/III AD1000_114_C07 (AIE91288.1)
MGII/III KM3_115_D04 (AIE99658.1)
Marine euryarchaeota group II/III
Euryarchaeota
Thermoplasmatales
Thermoplasma volcanium
Ferroplasma acidarmanus
Ca. Caldiarchaeum subterraneum
Aigarchaeota archaeon SCGC_AAA471-F17 (pSL4)
Aigarchaeota
Ignicoccus hospitalis | Crenarchaeota
Ca. Nitrososphaera gargensis
Ca. Nitrososphaera viennensis EN76
Ca. Nitrososphaera evergladensis
Cenarchaeum symbiosum A
Ca. Nitrosoarchaeum koreensis MY1
Ca. Nitrosoarchaeum limnia SFB1
Ca. Nitrosoarchaeum limnia BG20
Ca. Nitrosopumilus salaria BD31
Ca. Nitrosopumilus sp. AR2
SCGC_AB-661-I02 (Antartic Winter water 80 m)
SCGC_AB-661-L21 (Antartic Winter water 80 m)
SCGC_AB-661-M19 (Antartic Winter water 80 m)
SCGC_AAA282-K18 (Deep water)
Ca. Nitrosopumilus sp. SJ
Nitrosopumilus maritimus SCM1
SCGC_AAA799-E16 (pSL4 Brine Red Sea)
SCGC_AAA799-N04 (Brine Red Sea)
Ca. Nitrosoterreus chungbukensis
Ca. Nitrosotenuis uzonensis
Ca. Nitrosopelagicus brevis
Thaumarchaeota single cell environmental genomes
Thaumarchaeota
0.5
Cluster- 2
B
Halobacterium salinarum
Halobacterium sp. NRC-1
Halorubrum lacusprofundi
Haladaptatus paucihalophilus
Halalkalicoccus jeotgali
Haloferax volcanii
Natronobacterium gregoryi
Natrinema pellirubrum
Natrialba magadii
Haloterrigena turkmenica
Natronomonas pharaonis
Halomicrobium mukohataei
Haloarcula marismortui
Haloarcula hispanica
Halococcus morrhuae
Haloquadratum walsbyi
Euryarchaeota
Halobacteriales
Methanocella paludicola | Euryarchaeota, Methanocellales
Archaeoglobus fulgidus
Ferroglobus placidus
Archaeoglobus veneficus
Euryarchaeota
Archaeaoglobales
Methanoplanus limicola
Methanolinea tarda
Methanoculleus marisnigri
Methanocorpusculum labreanum
Euryarchaeota
Methanomicrobiales
Cluster- 1
C
Thaumarchaeota single cell environmental genomes
SCGC_AB-663-O07 (Antartic circumpolar water 400 m)
SCGC_AAA008-E17 (South Atlantic Gyre 770 m)
SCGC_AAA288-O17 (North pacific Gyre 770 m)
fosmid (North pacific Subtropical Gyre 4000 m)
SCGC_AAA288-G05 (North pacific Gyre 770 m)
SCGC_AAA288-I14 (North pacific Gyre 770 m)
SCGC_AB-663-P07 (Antartic circumpolar water 400 m)
SCGC_AB-629-I23 (Deep water)
SCGC_AAA007-D23 (South Atlantic Gyre 800 m)
SCGC_AAA008-N07 (South Atlantic Gyre 770 m)
SCGC_AAA288-O22 (North Pacific Gyre 770 m)
SCGC_AAA008-M23 (South Atlantic Gyre 800 m)
SCGC_AAA288-J14 (North Pacific Gyre 770 m)
SCGC_AAA007-N19 (South Atlantic Gyre 880 m)
SCGC_AAA007-E02 (South Atlantic Gyre)
SCGC_AAA288-K20 (North Pacific Gyre 770 m)
fosmid (North pacific Subtropical Gyre 4000 m)
SCGC_AAA288-P02 (North Pacific Gyre 770 m)

Figure S2

Figure S3


A
Escherichia coli str. K-12 (AAC76912.1)
Bacillus cereus (AAP11757.1)
Ca. Nitrosopumilus sp. AR2 (YP_006775795.1)
Nitrosopumilus maritimus (WP_014963325.1)
Nitrosopumilus maritimus (WP_012215321.1)
Ca. Nitrosoarchaeum koreensis MY1 (ZP_08668105.1)
Ca. Nitrosoarchaeum limnia SFB1 (ZP_08256852.1)
Ca. Nitrosoarchaeum limnia BG20 (ZP_10379529.1)
putative 1,4-dihydroxy-2-naphtanoate octaprenyltransferases Menaquinone biosynthesis
Ca. Nitrososphaera gargensis (YP_006863044.1)
Ca. Nitrosoarchaeum limnia BG20 (ZP_10378486.1)
Ca. Nitrososphaera gargensis (YP_006863544.1)
putative UbiA prenyltransferases
Escherichia coli str. K-12 (AAC73531.1)
COX10 Saccharomyces cerevisiae (CAA97879.1)
Ca. Nitrososphaera gargensis (YP_006860722.1)
Cenarchaeum symbiosum A (YP_876899.1)
Ca. Nitrosoarchaeum koreensis MY1 (ZP_08668724.1)
Ca. Nitrosoarchaeum limnia SFB1 (ZP_08257563.1)
Ca. Nitrosoarchaeum limnia BG20 (ZP_10379296.1)
Nitrosopumilus maritimus SCM1 (YP_001582922.1)
Ca. Nitrosopumilus koreensis AR1 (YP_006774524.1)
Ca. Nitrosopumilus sp. AR2 (YP_006774866.1)
Ca. Nitrosopumilus salaria BD31 (ZP_10118224.1)
Ca. Nitrososphaera gargensis (YP_006862470.1)
Cenarchaeum symbiosum A (YP_875458.1)
Ca. Nitrosopumilus koreensis AR1 (YP_006774193.1)
Nitrosopumilus maritimus SCM1 (YP_001582559.1)
Ca. Nitrosopumilus sp. AR2 (YP_006776114.1)
Ca. Nitrosopumilus salaria BD31 (ZP_10117144.1)
Ca. Nitrosoarchaeum koreensis MY1 (ZP_08668413.1)
Ca. Nitrosoarchaeum koreensis MY1 (ZP_08668413.1)
Ca. Nitrosoarchaeum limnia SFB1 (ZP_08257181.1)
Ca. Nitrosoarchaeum limnia BG20 (ZP_10378922.1)
putative protoheme IX farnesyltransferases CyoE and COX10
Ca. Nitrosoarchaeum koreensis (WP_007549490.1)
Ca. Nitrosoarchaeum koreensis MY1 (ZP_08667163.1)
Ca. Nitrososphaera gargensis (YP_006861308.1)
Sulfolobus solfataricus P2 (AAK40480.1, ubiA-1)]
COQ2 Saccharomyces cerevisiae (CAA96321.1)
putative 4-hydroxybenzoate octaprenyltransferases UbiA and CoQ2
putative digeranylgeranylglyceryl phosphate (DGGGP) synthases
Synechocystis sp. PCC 6803 chlorophyll a synthase (BAA10281.1)
0.5


B
Methanobrevibacter smithii
Methanothermus fervidus
Methanothermobacter marburgensis
Methanobacterium sp. SWAN-1
Euryarchaeota Methanobacteriales
Aciduliprofundum boonei
Ferroplasma acidarmanus
Thermoplasma volcanium
Thermoplasma acidophilum (NP_394456.1)
Euryarchaeota Thermoplasmatales
Vulcanisaeta distributa
Caldivirga maquilingensis
Pyrobaculum aerophilum
Pyrobaculum calidifontis
Thermoproteus neutrophilus
Crenarchaeota Thermoproteales
Lokiarchaeum (KKK40725.1)
Ferroglobus placidus
Archaeoglobus fulgidus
Archaeoglobus veneficus
Euryarchaeota Archaeoglobales
Desulfurococcus kamchatkensis
Staphylothermus marinus
Crenarchaeota, Desulfurococcales
Ca. Caldiarchaeum subterraneum (BAJ46673.1)
Ca. Caldiarchaeum subterraneum (BAJ49218.1)
Aigarchaeota
Aeropyrum pernix
Crenarchaeota, Desulfurococcales
Acidilobus saccharovorans
Crenarchaeota, Acidilobales
Ignicoccus hospitalis
Crenarchaeota, Desulfurococcales
Sulfolobus tokodaii
Sulfolobus solfataricus P2 (AAK40896.1, ubiA-2)
Crenarchaeota, Sulfolobales
Methanocaldococcus jannaschii
Methanocaldococcus fervens
Methanotorris igneus
Methanococcus aeolicus
Methanococcus voltae
Methanococcus maripaludis
Methanococcus vannielii
Euryarchaeota Methanococcales
Methanococcoides burtonii
Methanosarcina barkeri
Methanosarcina mazei
Methanohalobium evestigatum
Methanosalsum zhilinae
Euryarchaeota Methanosarcinales
Natronomonas pharaonis
Haloquadratum walsbyi
Halorubrum lacusprofundi
Haloterrigena turkmenica
Natronobacterium gregoryi
Natrinema pellirubrum
Natrialba magadii
Haloferax volcanii
Haladaptatus paucihalophilus
Halobacterium salinarum
Halalkalicoccus jeotgali
Euryarchaeota Halobacteriales
Methanosaeta concilii
Methanosaeta thermophila
Euryarchaeota, Methanosarcinales
Methanocorpusculum labreanum
Methanoplanus limicola
Methanoculleus marisnigri
Methanolinea tarda
Euryarchaeota Methanomicrobiales
Thermococcus kodakarensis
Thermococcus sp. 4557
Thermococcus litoralis
Pyrococcus yayanosii
Pyrococcus furiosus
Pyrococcus abyssi
Pyrococcus horikoshii
Euryarchaeota Thermococcales
MGII/III_euryarchaeote_KM3_80_G12 (AIF18040.1)
MGII/III_euryarchaeote_AD1000_19_F05 (AIE91980.1)
MGII/III euryarchaeote_AD1000_54_C05 (AIE94852.1)
MGII/III euryarchaeote_KM3_63_C06 (AIF13621.1)
MGII/III_euryarchaeote_KM3_63_C07 (AIF13652.1)
MGII/III euryarchaeote_AD1000_44_A09 (AIE94197.1)
MGII/III euryarchaeote_AD1000_22_E05 (AIE92417.1)
uncultured marine group II euryarchaeote (EHR76593.1)
fosmid HF4000_APKG10F17 (ABZ10110.1)
MGII/III euryarchaeote_KM3_85_D06 (AIF19107.1)
MGII/III euryarchaeote_KM3_87_G11 (AIF19853.1)
MGII/III euryarchaeote_KM3_31_G10 (AIF08684.1)
Euryarchaeota MGII/III
0.5