This is a preprint of:

Villanueva, L., Schouten, S. & Sinninghe Damsté, J.S. (2016). Phylogenomic analysis of lipid biosynthetic genes of Archaea shed light on the 'lipid divide'. *Environmental Microbiology*, 19(1), 54-69.

Published version: dx.doi.org/10.1111/1462-2920.13361

Link NIOZ Repository: www.vliz.be/nl/imis?module=ref&refid=259343

[Article begins on next page]

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

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'.

17 **Introduction**

18   Membrane lipids are essential building blocks for the cell since membranes define the 'inside'

19 and the 'outside' of the cell. Membranes are involved in many biological processes such as

20 establishing and maintaining trans-membrane gradients, compartmentalizing biochemical reactions

21 into distinct functional domains, controlling transport into and out of cells, and inter- and intra-

22 cellular communication. The membrane lipid composition is also one of the most remarkable traits

23 distinguishing the three domains of life, Archaea, Bacteria, and Eukarya (Woese and Fox, 1977).

24 Membrane lipids of Bacteria and Eukarya share a large number of structural similarities as they are

25 typically composed of two fatty acid chains that are linked to a glycerol moiety via ester bonds and

26 are organized in a bilayer structure (Lombard *et al*., 2012a). On the other hand, membrane lipids of

27 Archaea are characterized by ether linkages between the glycerol moiety and isoprene-based alkyl

28 chains in either a bilayer or monolayer (Koga and Morii, 2007; Lombard *et al*., 2012a). These traits

29 are not fully exclusive to these groups since membrane-spanning lipids have also been reported in

30 some members of the Bacteria (Sinninghe Damsté *et al*., 2002, 2007; Weijers *et al*., 2006), and fatty

31 acids in some archaeal species (Gattinger *et al*., 2002). An exclusive distinction in the structures of

32 membrane lipids of archaea and bacteria/eukaryotes is the opposite stereochemistry of the glycerol

33 phosphate backbone, being *sn*-glycerol-1-phosphate (G1P) in archaea, and *sn*-glycerol-3-phosphate

34 (G3P) in bacteria and eukaryotes (Kates, 1993). The biosynthesis of G3P and G1P is catalyzed by

35 two entirely different enzymes (i.e. glycerol-1- and glycerol-3-phosphate dehydrogenase, G1PDH,

36 and G3PDH, respectively) that, based on differences in the catalytic reaction and protein sequence

37 (Koga *et al*., 2003; Han *et al*., 2005), are not evolutionary related (Koga *et al*., 1998). This

38 differentiation of lipid structures between Archaea, on the one hand, and Bacteria and Eukarya, on

39 the other, has been coined as the 'the lipid divide'.

40   This 'lipid divide' has posed some fundamental questions on microbial evolution. Since

41 Archaea and Bacteria are believed to stem from a common ancestor (the cenancestor or last

42 universal common ancestor, LUCA), their completely different membrane lipid structures represent

43 a conundrum. Koga *et al*. (Koga *et al*., 1998) proposed that the cenancestor lacked a membrane and

3

44   that the specific archaeal and bacterial membrane lipid biosynthetic pathways emerged later and

45   independently in the lineages leading to Archaea and Bacteria. Martin & Russell (2003)

46   hypothesized that the cenancestor had mineral monosulfide compartments instead of lipids.

47   Wächtershäuser (2003) suggested that the cenancestor had a lipid heterochiral membrane containing

48   both stereochemical forms of the glycerol phosphate backbone, which progressively diverged into a

49   more stable homochiral membrane leading to the differentiation between archaea and bacterial

50   membranes. However, experiments with liposomes containing both archaeal and bacterial

51   membrane lipids showed that heterochiral membranes are also stable (Fan *et al.*, 1995; Shimada *et*

52   *a*l., 2011), suggesting that there exists no evolutionary pressure to select for organisms with a

53   homochiral membrane. A recent study by Sojo *et al.* (2014) based on modeling of membrane

54   bioenergetics suggested that LUCA did not have membranes with glycerol phosphate headgroups,

55   which would have reduced proton permeability, but rather a lipid bilayer composed of both fatty

56   acids and isoprenes, and that modern membranes in Bacteria and Archaea arose later and

57   independently.

58   Another conundrum is the similarity of membrane lipids of the Eukarya with those of Bacteria

59   rather than with those of the Archaea, which are believed to be the predecessors of the Eukarya

60   (Pereto *et al.*, 2004). According to the classical Woesian three-domain phylogeny, the last common

61   ancestor of archaea and eukaryotes would have had an archaeal membrane that was later replaced

62   by a bacterial-like membrane in eukaryotes, or alternatively that an ancestral mixed membrane with

63   G1P- and G3P-based membrane lipids evolved to an archaeal membrane in archaea and to a

64   bacterial-like membrane in eukaryotes. However, both options are difficult to reconcile as they

65   would involve an intensive horizontal gene transfer of the genes required, while the mixed

66   membrane model would imply that bacterial-like membranes evolved twice from the cenancestor in

67   bacteria and in eukaryote. Currently, the most accepted early life evolutionary theory considers

68   Archaea and Bacteria as primary branches derived directly from the cenancestor, while Eukarya

69   would have evolved secondarily as a chimeric organism derived from the endosymbiosis of one

70   bacterium (the ancestor of mitochondria) within a host cell (Gray and Doolittle, 1982; Golding and

4

71 Gupta, 1995; among others). However, the origin of the host cell is still under debate. In any case,

72 most models of the origin of eukaryotic cells require a transition from an archaeal-like membrane to

73 a bacterial membrane, including a reversal of the glycerol stereochemistry of membrane lipids

74 Nonetheless, no evidence for this kind of transition has ever been found in bacteria or archaea,

75 casting some doubts on this mechanism.

76 The growing availability of genomes could shed light on the evolutionary processes leading to

77 the 'lipid divide'. Recent studies based on environmental metagenomics have defined several new

78 archaeal lineages (Castelle *et al.*, 2015). Currently, the domain Archaea is represented by two

79 superphyla (DPANN and TACK) and the phylum Euryarchaeota (Guy and Ettema, 2011; Fig. S1).

80 The TACK-superphylum is comprised of Thaumarchaeota, Aigarchaeota, Crenarchaeota,

81 Korarchaeota, and some other phyla (e.g. Guy and Ettema, 2011; Williams *et al.*, 2012; Martijn and

82 Ettema, 2013). Some phylogenomic studies have provided evidence that the archaeal 'ancestor' of

83 the eukaryotic cell emerged from the TACK superphylum (Guy and Ettema, 2011). Furthermore, a

84 recent study suggests that the novel candidate archaeal phylum 'Lokiarchaeota' (Deep-Sea Archaeal

85 Group/Marine Benthic Group B, DSAG/MBG-B; Spang *et al.*, 2015; Fig. S1), a deep branching

86 clade of the TACK superphylum, forms a monophyletic group with the eukaryotes. Indeed, the

87 'Lokiarchaeum' composite genome codes a remarkable number of eukaryotic signature proteins,

88 supporting the hypothesis that the eukaryotic cell evolved from an archaeal ancestor of this group

89 (Spang *et al*., 2015).

90 The recent discovery of Lokiarchaeum, which potentially shares a common ancestor with

91 eukaryotes, prompted us to re-examine the 'lipid divide' conundrum. We investigated two key

92 aspects of the lipid divide: the specific stereoconfiguration of archaeal lipids and the capacity for

93 fatty acid synthesis in archaea. We searched for homologs of genes encoding for enzymes involved

94 in membrane lipid and fatty acid biosynthetic pathways in archaeal genomes and performed

95 phylogenomic analyses with the annotated homologs. The results reveal differences in the lipid

96 biosynthetic pathway, especially concerning the stereochemistry of the glycerol phosphate

5

97 backbone, in certain uncultured archaeal groups at key evolutionary phylogenetic positions with

98 substantial implications for our understanding of the 'lipid divide'.

Results and Discussion

100 *Enzymes involved in the glycerol phosphate stereospecific biosynthesis in Archaea*

101 The stereoconfiguration of archaeal lipids is established by the enzyme G1PDH. This enzyme is

102 thought to be restricted to archaea (e.g. Pereto *et al*., 2004; Koga and Morii, 2007; Matsumi *et al*.,

103 2011) although a G1PDH homolog (AraM) has been found in *Bacillus* sp., and some related

104 bacterial species (Guldan *et al*., 2008). Our survey of archaeal genomes revealed that the gene

105 coding for G1PDH (*egs*A, Fig.1) is present in almost all examined archaea but, interestingly, is

106 absent in marine group II euryarchaeota (MGII; Iverson *et al*., 2012), in fosmid sequences of the

107 MGII and marine group III euryarchaeota (MGIII) (Deschamps *et al.,* 2014), 'Lokiarchaeum'

108 (Spang *et al*., 2015), and all examined species of the DPANN superphylum (Castelle *et al*., 2015)

109 (Table 1, Table S1). This suggests that these uncultivated archaea may not have the ability to

110 synthesize the G1P backbone of archaeal lipids. For the members of the DPANN this is not

111 surprising because they have simplified genomes of reduced size and are thought to rely on host

112 cells or cell debris for the synthesis of their lipids (Waters *et al*., 2003; Jahn *et al*., 2004). In this

113 respect, it is notable that in some of the DPANN genomes some homologs of enzymes involved in

114 the archaeal membrane lipid biosynthesis are present, although they lack the MVK gene encoding

115 for mevalonate kinase (Table 1; Table S1), an essential enzyme for isoprenoid biosynthesis. This

116 situation may represent an intermediate stage of progressively losing those genes. However, truly

117 exceptional is the lack of G1PDH in MGII and MGIII euryarchaeota and in 'Lokiarchaeum', as we

118 found that genomes of these groups of archaea still harbor all the other known genes coding for the

119 enzymes of the archaeal lipid biosynthetic pathway (i.e. geranylgeranylglyceryl phosphate synthase,

120 GGGP; digeranylgeranylglyceryl phosphate synthase, DGGGP; geranyl reductase, GR, among

121 others; Table 1; Table S1, S2). It should be noted that the current genome assembly of

122 'Lokiarchaeum' (*Lokiarchaeaum* sp. GC14_75), is 92% complete (Spang *et al*., 2015), and thus

6

123   there is a small chance that the gene coding for G1PDH would be located in the unsequenced part.

124   However, we also did not detect any homologs of the G1PDH coding gene in the larger

125   metagenome dataset (LCGC14AMP, 56.6 Gbp) reported in the same study (Spang *et al.*, 2015).

126   The lack of G1PDH homologs in the fosmid sequences of MGIII (Deschamps *et al.,* 2014) could

127   potentially be due to their lack of completeness, however this can be ruled out for the genome of

128   MGII reported by Iverson *et al.*, 2012, which is closed. Our analysis strongly suggests that MGII

129   and MGIII euryarchaeota and 'Lokiarchaeum' are capable of synthesizing isoprenoid-based ether

130   lipids for their membranes but not with G1P as the glycerol building block.

131   G3PDH (encoded by the *gps* gene) catalyzes the conversion of dihydroxyacetone phosphate

132   (DHAP) into G3P and this enzyme is responsible in Bacteria and Eukarya for the stereochemistry

133   of the glycerol units of their membrane lipids. *Gps*-coded G3PDH homologs were previously

134   detected in the euryarchaeota *Archaeoglobus fulgidus* and *Methanothermobacter*

135   *themoautotrophicum,* in addition to G1PDH, but their metabolic function is unknown (Pereto *et al.*,

136   2004). Some (heterotrophic) archaea have been reported to synthesize a 'G3PDH' enzyme encoded

137   by the *glp* gene (Pereto *et al.*, 2004; Koga and Morii, 2007) but this enzyme catalyzes the

138   conversion of G3P into DHAP, the reverse of the reaction catalyzed by *gps*-coded G3PDH (Fig. 1).

139   It is believed that this enables heterotrophic archaea to feed glycerol into the glycolysis pathway

140   (Nishihara *et al.*, 1999), as it is also observed for bacteria able to metabolize glycerol. These

141   heterotrophic archaea still biosynthesize membrane lipids with the archaeal stereochemistry as they

142   also harbor G1PDH (Fig. 1).

143   In our survey of archaeal genomes, *gps*-encoded G3PDH homologs were detected in MGII/III

144   euaryarchaeota, and in some archaea of the orders Archaeoglobales and Methanobacteriales

145   (*Methanobrevibacter* sp.), as well as in species of the DPANN superphylum (AR9 and AR11

146   genomes of the Woesearchaeota), but not in 'Lokiarchaeum' (Table 1; Table S1). Homologs of *glp*-

147   encoded G3PDH were found in several archaeal genomes including members of the

148   Thermococcales, Archaeoglobales, Halobacteriales, Thermoplasmatales, Korarchaeota, some

149   Crenarchaeota genomes, as well as in the 'Lokiarchaeum' genome, which contains three putative

7

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.

8

176   G3P can also be synthesized by degradation of glycerophosphodiesters by a

177   glycerophosphodiester phosphodiesterase (GDPD) producing the corresponding alcohols and G3P

178   (Larson *et al*., 1983; Fig. 1). Glycerophosphodiesters are enzymatically produced by

179   phospholipases A_1 and A_2 from membrane phospholipids (e.g. Istivan *et al*., 2006).

180   Glycerophosphodiester phosphodiesterase activities have been characterized in bacteria as well as

181   in eukaryotes (Tommassen *et al*., 1991; Fisher *et al*., 2005; van der Rest *et al*., 2002), and genomic

182   analyses have revealed a wide distribution of this protein family from bacteria and Archaea to

183   metazoans, plants, and fungi (Santelli *et al*., 2004). In the bacterium *Escherichia coli*, the

184   transformation of glycerophosphodiesters into G3P is thought to be catalyzed by two homologous

185   enzymes, a periplasmic GDPD GlpQ, and a cytosolic GDPD UgpQ with a broad substrate

186   specificity toward various glycerophosphodiesters (e.g. Tommassen *et al*., 1991). Here we detected

187   archaeal UgpQ homologs in genomes of the Crenarchaeota, and in euryarchaeotal genomes of the

188   Methanobacteriales, Thermoccoccales, Methanomicrobiales, Halobacteriales and

189   Thermoplasmatales (Table 1; Table S1). Two putative UgpQ GDPD homologs were also detected

190   in two genomes of the uncultured marine euryarchaeota group II/III, which were in turn closely

191   related to putative UgpQ GDPD of the euryarchaeota Halobacteriales (Fig. 4). In addition, two

192   putative homologs of UgpQ GDPD were detected in the 'Lokiarchaeum' genome (Table 1; Table

193   S1-S2), which were closely related to another putative UgpQ GDPD annotated in the DPANN

194   Woesearchaeota genome GW2011_AR3 as well as UgpQ detected in bacterial genomes of the

195   Thermotogae (Fig. 4). The detection of two putative GDPD UgpQ in the 'Lokiarchaeum' genome

196   opens the possibility of the formation of a G3P backbone by degradation of glycerophosphodiesters,

197   as indicated in Fig. 1.

198   Putting this genetic evidence together, the data indicate that the unique archaeal groups that lack

199   the gene encoding for G1PDH, i.e. the MGII/III euryarchaeota and 'Lokiarchaeum', harbor

200   homologs of genes involved in the synthesis of the G3P backbone in Bacteria and Eukarya, either

201   through the catalysis of DHAP to G3P in the case of MGII/III euryarchaeota (G3PDH encoded by

202   *gps*), or in 'Lokiarchaeum' by degradation of glycerophosphodiesters by glycerophosphodiester

9

203   phosphodiesterase (UgpQ GDPD) or by metabolism of glycerol into G3P via glycerol kinase

204   (encoded by the *glp*K gene). Consequently, it is tempting to speculate that both MGII and MGIII

205   euryarchaeota and 'Lokiarchaeum' synthesize archaeal ether-linked membrane lipids, since they

206   have the complete set of the genes coding for the enzymes of the archaeal lipid biosynthetic

207   pathway (Table 1; Table S1, S2), but with the G3P stereochemistry typical for Bacteria and

208   Eukarya, as they lack genes for G1PDH but possess genes encoding for enzymes for G3P

209   formation. However, the synthesis of these kinds of lipids would be only possible if enzymes

210   similar to GGGP and DGGGP synthases would catalyze the formation of ether bonds between

211   isoprenoid chains and the G3P backbone as specified in the hypothetical pathway in Fig. 1.

212   *Enzymes involved in the fatty acid biosynthetic pathway in Archaea*

213   In bacterial fatty acid biosynthesis, the acetyl-CoA carboxylase (ACC) converts acetyl-

214   coenzyme A (CoA) into malonyl-CoA (Fig. 5). Secondly, the peptide cofactor acyl carrier protein

215   (ACP) has to be activated by an ACP synthase. Finally, the malonyl-CoA:ACP transacylase

216   (MCAT) charges the malonyl-CoA to holo-ACP, resulting in malonyl-ACP building blocks needed

217   by fatty acid synthases (Fig. 5). Although the occurrence of diacyl glycerols is generally believed to

218   be restricted to Bacteria and Eukarya, the presence of minor amounts of free fatty acids have been

219   previously reported in some archaea (Kates *et al*., 1968; Langworthy *et al*., 1974; Gattinger *et al*.,

220   2002). In line with this, homologs of several of bacterial enzymes of the fatty acid biosynthetic

221   pathway (i.e. ACC; beta-ketoacyl synthase, KAS, FabH; beta-ketoacyl reductase, KR, FabG; DH,

222   beta-hydroxyacyl dehydratase; and enoyl reductase, ER) have been detected in some archaeal

223   genomes (Pereto *et al*., 2004; Fig. 5). In addition, a study by Iverson *et al*. (2012) annotated

224   homologs of the genes coding for ACC, KAS, KR, and ER in the genome of a MGII euryarchaeota.

225   Lombard *et al*. (2012b) observed that, except for a few unrelated species that probably acquired

226   ACP by independent HGT (i.e. some species of the Methanomicrobiales and Halobacteriales), no

227   homologs of the ACP-processing machinery (ACP synthase and MCAT) could be detected in

228   archaeal genomes. More recently, Dibrova *et al*. (2014) suggested a hypothetical archaeal fatty acid

229   pathway based on the presence of the gene coding for archaeal acetyl-CoA-acetyltransferase (also

230   known as acetyl-CoA C-acyl transferase) in all archaeal genomes up to date with the exception of

231   *Nanoarchaeum equitans* (Matsumi *et al.*, 2011). This enzyme catalyzes the first condensation

232   reaction in the mevalonate pathway producing acetoacyl-CoA, which then would be further reduced

233   and dehydrated by bacterial-type enzymes involved in the β-oxidation of fatty acids (acyl-CoA

234   dehydrogenase, FadE; enoyl-CoA hydratase, FadB1; 3-hydroxyacyl-CoA dehydrogenase, FadB2;

235   Fig. 5), operating in reverse direction.

236   To shed further light on the potential capacity of archaea to synthesize fatty acids we performed

237   an extensive search for archaeal homologs of the enzymes involved in the ACP-processing

238   machinery (i.e. ACP synthase and MCAT), which are generally lacking in archaeal genomes and

239   could potentially prohibit a complete the archaeal fatty acid biosynthesis. We also searched for

240   archaeal homologs of the beta-ketoacyl synthase (KAS), which have been previously annotated in

241   several archaeal genomes. Moreover, we also performed a more extensive search for the homologs

242   potentially involved in the hypothetical archaeal fatty acid synthesis pathway as proposed by

243   Dibrova *et al.*, (2014) in available archaeal genomes.

244   Putative homologs of ACP synthase were detected in some species of Methanomicrobiales,

245   Archaeoglobales, Halobacteriales, Thermoplasmatales, in some Crenarchaeota species, and in a

246   genome of a species in the DPANN superphylum (*i.e.* AR5 Aenigmarchaeota; Table 1; Table S1).

247   In our survey, we also detected MCAT (FabD) homologs in MGII and MGIII genomes (Table 1;

248   Table S1), as well as in some Halobacteriales (cf. Lombard *et al.*, 2012b). In the case of the MGII,

249   we detected the MCAT (FabD) gene as part of the previously annotated polyketide synthase, which

250   also contained a beta-ketoacyl synthase (KAS) and a polyketide synthase dehydratase (FabA/Z

251   dehydratase) domains. We have also observed this distribution of the MCAT (FabD) coding region

252   for several of the MGII/III genomes released by Deschamps *et al.* (2014) (Table 1; Table S1). In

253   addition, a putative MCAT homolog was also detected in the 'Lokiarchaeum' genome within the

254   protein previously annotated as phenol phthiocerol synthesis polyketide synthase type I (Table S1,

255   S2). This protein harbors a MCAT (FabD; comprised between residues 14604−15485; Table S2),

11

256   and KAS regions (FabBI and FabFII; residues 13674−14525). We performed phylogenetic analyses

257   of the MCAT (FabD) proteins detected in archaeal genomes to determine their evolutionary

258   relationships between each other and with bacterial homologs (Fig. 6). MCAT homologs of the

259   MGII and MGIII, and 'Lokiarchaeum' were related to bacterial MCAT homologs of the

260   Acidobacteria, Chloroflexi, and Firmicutes, while the MCAT homologs detected in Halobacteriales

261   genomes were quite different from the rest of the archaeal homologs as well as the bacterial ones,

262   suggesting a different evolutionary origin for these MCAT proteins.

263   Archaeal homologs of the genes coding for the acyl-CoA dehydrogenase FadE, enoyl-CoA

264   hydratase FadB1, and 3-hydroxyacyl-CoA dehydrogenase FadB2 were detected in some of the

265   genomes of the euryarchaeota Archaeoglobales, Halobacteriales, Thermoplasmatales, uncultured

266   marine euryarchaeota group II/III, MBG-D, and most of the Crenarchaeotal and Thaumarchaeotal

267   genomes (Table 1; Table S1). One putative homolog of FadE was detected in the 'Lokiarchaeum'

268   genome, as well as multiple copies of putative homologs of FadB1 and FadB2 (Table 1; Table S1,

269   S2), which would also suggest that 'Lokiarchaeum' harbors the potential for fatty acid synthesis

270   with the hypothetical pathway proposed by Dibrova *et al*. (2014).

271   Our survey of the occurrence of genes coding for key fatty acid biosynthetic enzymes (i.e. ACP

272   synthase, MCAT and KAS) in archaeal genomes (Table 1; Table S1) suggests that only species of

273   the Halobacteria, MGII/III euryarchaeota and 'Lokiarchaeum' have all the key genes required to

274   potentially synthesize fatty acids. Species of the Halobacteria possess all three key genes (Table 1;

275   Table S1). Although MGII/III and 'Lokiarchaeum' lack annotated homologs of ACP synthase

276   (Table 1; Table S1), these archaeal groups could potentially still synthesize bacterial-like fatty acids

277   by an ACP-independent pathway as proposed by Lombard *et al*. (2012b). Furthermore, the

278   extensive search we performed for the genes coding for the FadE, FadB1 and FadB2, also showed

279   that most of the archaeal genomes, including MGII/III euryarchaeota and 'Lokiarchaeum', harbor

280   the potential for hypothetical fatty acid biosynthesis pathway proposed by Dibrova *et al*. (2014)

281   (Fig. 5).

282   Based on the genetic potential of some archaea to produce fatty acids, we further investigated

283   archaeal genomes for genes coding for enzymes catalyzing the esterification of fatty acids and G3P

284   required for the formation of glycerol ester lipids (Fig. 5). There are two families of enzymes

285   responsible for the acylation of the 1-position of the G3P. The PlsB acyltransferase, found in the

286   bacteria *Escherichia coli* and in many eukaryotes, primarily uses acyl-acyl carrier protein (ACP)

287   end products of fatty acid biosynthesis (acyl-ACP) as acyl donors but may also use acyl-CoA

288   derived from exogenous fatty acids (Fig. 5). The other family concerns the PlsY acyltransferase

289   which is more widely distributed in Bacteria and uses as donor acyl-phosphate produced from acyl-

290   ACP by the PlsX, an acyl-ACP:PO_4 transacylase enzyme (Fig. 5). The acylation in the 2-position

291   of the G3P is carried out by the 1-acylglycerol-3-phosphate *O*-acyltransferase (PlsC) (Fig. 5).

292   We performed genomic searches of putative archaeal homologs of the bacterial acyl-ACP

293   transferases involved in the formation of ester bonds between fatty acids and the G3P backbone in

294   the phospholipid synthesis (Fig. 5). No homologs of the PlsB acyltransferase were detected in any

295   of the archaeal genomes with the exception of one species of the DPANN superphylum (AR1

296   Pacearchaeota; Table 1; Table S1). Archaeal homologs to the PlsY acyltransferase were only found

297   in the 'Lokiarchaeum' genome (i.e. one putative PlsY homolog; Table 1; Table S1, S2). Molecular

298   phylogeny (Fig. 7) indicated that the 'Lokiarchaeum' PlsY homolog was closely related to PlsY

299   enzymes of bacterial groups such as Thermotogae, Spirochaetales and Dictioglomales which

300   suggest that this enzyme was acquired by lateral gene transfer from Bacteria. A wide variety of

301   putative homologs of the PlsC 1-acylglycerol-3-phosphate *O*-acyltransferase were found in MGII

302   and MGIII genomes, a DPANN genome (AR11 Woesearchaeota), and two putative PlsC were

303   found in the 'Lokiarchaeum' genome (Table 1; Table S1, S2). Moreover, PlsC protein homologs of

304   the MGII/III euryarchaeota, AR11 DPANN, and 'Lokiarchaeum' were closely related to those of

305   the α-, β-Proteobacteria or Actinobacteria, the newly proposed Parcubacteria group (Brown *et al.*,

306   2015), and a species of the Firmicutes, respectively (Fig. 7).

307   Our data suggest that, in addition to the apparent ability to synthesize 'bacterial' G3P, MGII/III

308   euryarchaeota and 'Lokiarchaeum' also possess a putative fatty acid synthetic pathway.

13

309 Furthermore, the detection of homologs of the acyl transferase PlsY in the 'Lokiarchaeum' genome

310 and PlsC homologs in both MGII/III euryarchaeota and 'Lokiarchaeum' genomes suggests that

311 their biosynthetic machinery would be able to form ester-bonded fatty acid membrane lipids with

312 G3P stereochemistry. Based on the enzyme inventory (Table 1; Table S1), these archaea are

313 predicted to produce chimeric membrane lipids, such as di- or tetraether-linked isoprenoidal

314 membrane lipids with a bacterial/eukaryote G3P stereochemistry, or lipids with one ether-linked

315 isoprenoidal chain at position *sn*-1 of a G3P backbone and one ester-bound fatty acid at position *sn*-

316 2 (see "hypothetical part" of Fig. 1). Mixed ether/ester membrane lipids have been previously

317 detected in aerobic and anaerobic bacteria such as anammox bacteria, sulfate-reducing bacteria,

318 members of the bacterial order Thermotogales and Acidobacteria (Rütters *et al*., 2001; Sinninghe

319 Damsté *et al*., 2002, 2007, 2011, 2014). In fact, the presence of these types of lipids in the order

320 Thermotogales, an early-branching clade of the Bacteria, was interpreted as an indication that the

321 ability to produce both ether and ester-linked membrane lipids developed relatively early during

322 microbial evolution (Sinninghe Damsté *et al*., 2007). However, the early branching in the tree of

323 life of Thermotogales and Aquificales has been questioned and it has even been proposed that the

324 majority of the genes of these groups shows affinities to Archaea and Firmicutes (Zhaxybayeva *et*

325 *al*., 2009, among others).

326 In addition, 'chimeric' tetraether lipids containing both n-alkyl and isoprenoidal chains have

327 been previously detected in the environment (Schouten *et al*., 2000; Liu *et al*., 2012). In the case of

328 'Lokiarchaeum', it is also possible that they synthesize bacterial-like fatty acids ester-bound at the

329 *sn*-1 and *sn*-2 positions of the G3P, as we have detected both putative acyltranferases (PlsY and

330 PlsC) in its genome. However, we did not detect a putative homolog of the PlsX protein involved in

331 the transformation of acyl-ACP to acylphosphate needed for the catalysis mediated by PlsY.

332 Therefore it remains unknown if 'Lokiarchaeum' is able to mediate the formation of bacterial-like

333 ester-bond fatty acid in the *sn*-1 position. The formation of the hypothetical chimeric tetraether

334 lipids (see "hypothetical part" of Fig. 1) would follow a biosynthetic pathway in which enzymes

335 similar to GGGP and DGGGP synthases would catalyze the formation of ether bonds between

14

336   isoprenoid chains and G3P backbone in *sn*-2 position instead of with the expected G1P. Since

337   GGGP synthase has proven to be selective for G1P (Peterhoff *et al.*, 2014), it would suggest that

338   this step is mediated by a totally different enzyme

339   *Implications for the 'lipid divide'*

340   The potential capacity of synthesis of isoprenoidal ether and fatty acid ester lipids with a G3P

341   backbone within a single organism, i.e. MGII/III euryarchaeota species and 'Lokiarchaeum', sheds

342   new light on the current 'lipid divide'. This is especially relevant for 'Lokiarchaeum' as its genome

343   codes a remarkable number of eukaryotic signature proteins, which has been used as an argument to

344   support the hypothesis that the eukaryotic cell evolved from an archaeal ancestor of this group

345   (Spang *et al.*, 2015). Our genome mining study suggests that 'Lokiarchaeum' has the biosynthetic

346   capacity to synthesize archaeal ether-linked and fatty acid ester linked membrane lipids with the

347   bacterial/eukaryotic G3P stereochemistry further supports the hypothesis of Spang *et al.* (2015). If

348   'Lokiarchaeum' was indeed a descendant of the archaeal ancestor leading to the eukaryotic cell,

349   then this ancestor may have possessed the capacity for both isoprenoidal ether and fatty acid ester

350   lipids with a G3P backbone. After the endosymbiosis of the archaeal ancestor with a bacterium, the

351   capacity for isoprenoid ether lipid synthesis may have been lost, leaving the fatty acids ester lipids

352   with a G3P backbone as the main membrane lipid.

353   It is not clear why phylogenetically distant archaeal groups such as MGII/III euryarchaeota and

354   'Lokiarchaeum' both harbor these particular lipid biosynthetic capacities. Extensive bacteria-to-

355   archaea gene transfer has occurred in MGII/III euryarchaeota, Thaumarchaeota, Halobacteria and

356   mesophilic methanogens (López-García *et al.*, 2004; Brochier-Armanet *et al.*, 2011; Nelson-Sathi *et*

357   *al.*, 2012; Deschamps *et al.*, 2014). It has been proposed that this has promoted their adaptation to a

358   mesophilic lifestyle (López-García *et al.*, 2015). The Lokiarchaeum genome also contains a

359   relatively high fraction of genes that display a high similarity to genes of bacterial origin (i.e. 29%

360   of all genes; Spang *et al.,* 2015), which is comparable to that in MGII/III euryarchaeota

361   (Deschamps *et al.*, 2014). This high level inter-domain gene exchange between Bacteria and

15

362 Archaea may have substantially impacted the membrane lipid biosynthetic pathway in both

363 MGII/III euryarchaeota and 'Lokiarchaeum' to such an extent that they produce 'chimeric'

364 membrane lipids. Acquisition of only one more bacterial gene (PlsB) by an ancestor of

365 'Lokiarchaeum' would result in a 'full' bacterial/eukaryotic lipid membrane pathway, paving the

366 road leading to the development of a 'truly' eukaryotic cell membrane. Our study suggests that the

367 'lipid divide' between the domain Archaea, on the one hand, and those of the Bacteria and Eukarya,

368 on the other, is less clear cut as previously thought.

369   The required next step following our phylogenomic study is to provide confirmation of our

370 hypothesis by identification of the 'chimeric' lipids predicted here. However, such an endeavor is

371 strongly hindered by the lack of cultivated representatives of MGII/III euryarchaeota and

372 'Lokiarchaeota'. Future studies should focus on determining 'unusual' membrane lipids in natural

373 environments with high abundances of these uncultured archaeal groups, in particular by

374 determining the stereochemistry of their glycerol membrane lipids (cf. Weijers *et al.,* 2006). This

375 may also provide further insight in the evolution of Eukarya from the prokaryotes.

376 **Experimental procedures**

377 *Computational analysis*

378 Putative homologs of the enzymes mentioned in the text (Table 1; Table S1, S2) were detected by

379 tblastn (search translated nucleotide databases using a protein query) and blastp (protein query

380 against protein databases) searches using annotated enzymes as query sequences and with a

381 minimum e-value of $1e^{-25}$. The identity of the putative homologs was further investigated by visual

382 inspection of the alignment.

383 *Phylogenetic analyses*

384 Putative and annotated partial homologs aligned by Muscle (Edgar, 2004) in Mega6 software

385 (Tamura *et al*., 2013) and edited manually. Phylogenetic reconstruction was performed by

386 maximum likelihood in PhyML v3.0 (Guindon *et al*., 2010) using the best model according to AIC

387 indicated by ProtTest 2.4 (Abascal *et al*., 2005) as indicated in the figure legends.

16

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.

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

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

435 Iverson, V., Morris, R.M., Frazar, C.D., Berthiaume, C.T., Morales, R.L., and Armbrust, E.V.
436 (2012) Untangling genomes from metagenomes: revealing an uncultured class of marine
437 Euryarchaeota. Science 335: 587-590.

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

441 Jung, M.Y., Park, S.J., Kim, S.J., Kim, J.G., Sinninghe Damsté, J.S., Jeon, C.O., and Rhee, S.K.
442 (2014) A mesophilic, autotrophic, ammonia-oxidizing archaeon of thaumarchaeal group I.1a
443 cultivated from a deep oligotrophic soil horizon. Appl Environ Microbiol 80: 3645-3655.

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

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

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

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

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

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

460 Larson, T.J., Ehrmann, M., and Boos, W. (1983) Periplasmic glycerophosphodiester
461 phosphodiesterase of Escherichia coli, a new enzyme of the glp regulon. J Biol Chem 258: 5428-
462 5432.

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

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

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

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

18

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

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

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

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

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

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

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

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

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

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

502 Rütters, H., Sass, H., Cypionka, H., and Rullkotter, J. (2001) Monoalkylether phospholipids in the
503 sulfate-reducing bacteria Desulfosarcina variabilis and Desulforhabdus amnigenus. Arch Microbiol
504 176: 435-442.

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

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

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

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

517 Sinninghe Damsté, J. S., Rijpstra, W.I., Hopmans, E.C., Schouten, S., Balk, M., and Stams, A.J.
518 (2007) Structural characterization of diabolic acid-based tetraester, tetraether and mixed ether/ester,
519 membrane-spanning lipids of bacteria from the order Thermotogales. Arch Microbiol. 188: 629-
520 641.

19

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

525 Sinninghe Damsté, J.S., Rijpstra. W.I., Hopmans, E.C., Weijers, J.W., Foesel, B.U., Overmann, J..
526 and Dedysh, S.N. (2011) 13,16-Dimethyl octacosanedioic acid (iso-diabolic acid), a common
527 membrane-spanning lipid of Acidobacteria subdivisions 1 and 3. Appl Environ Microbiol 77: 4147-
528 4154.

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

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

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

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

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

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

547 Wächtershäuser, G. (2003) From pre-cells to Eukarya–a tale of two lipids. Mol Microbiol 47: 13-
548 22.

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

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

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

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

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

565 **Acknowledgments**

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.

Phylogenetic classification	%**	Isoprenoid biosynthesis MVK	Glycerol backbone biosynthesis G1PDH	G3PDH *gps*	Glycerol & glycerol phosphodiester catabolism G3PDH *glp*	*glp*K	GDPD	Ether lipid production GGGP	DGGGP	GR	Fatty acid biosynthesis ACP synthase	FabD	KAS	FadE	FadB1	FadB2	Ester–bond formation glycerol and fatty acids 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	√	√	×	√	√	×	√	√	√	×	×	×	√	×	√	×	×	×

22

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).

24

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.
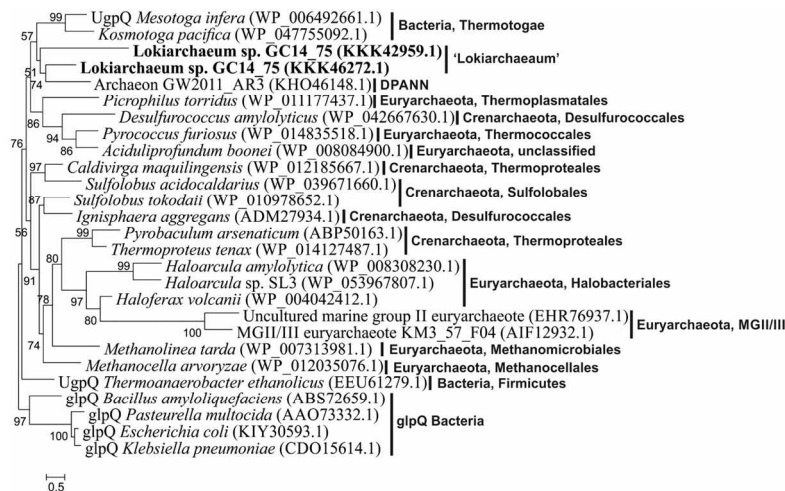
170x186mm (150 x 150 DPI)

Figure 1

Figure 2

Figure 3

Figure 4

99 — UgpQ *Mesotoga infera* (WP_006492661.1) | Bacteria, Thermotogae
57 — *Kosmotoga pacifica* (WP_047755092.1)
— **Lokiarchaeum sp. GC14_75 (KKK42959.1)** | 'Lokiarchaeaum'
51 — **Lokiarchaeum sp. GC14_75 (KKK46272.1)**
74 — Archaeon GW2011_AR3 (KHO46148.1) | DPANN
— *Picrophilus torridus* (WP_011177437.1) | Euryarchaeota, Thermoplasmatales
86 — *Desulfurococcus amylolyticus* (WP_042667630.1) | Crenarchaeota, Desulfurococcales
94 — *Pyrococcus furiosus* (WP_014835518.1) | Euryarchaeota, Thermococcales
86 — *Aciduliprofundum boonei* (WP_008084900.1) | Euryarchaeota, unclassified
76 — 97 — *Caldivirga maquilingensis* (WP_012185667.1) | Crenarchaeota, Thermoproteales
— *Sulfolobus acidocaldarius* (WP_039671660.1) | Crenarchaeota, Sulfolobales
87 — *Sulfolobus tokodaii* (WP_010978652.1)
— *Ignisphaera aggregans* (ADM27934.1) | Crenarchaeota, Desulfurococcales
56 — 99 — *Pyrobaculum arsenaticum* (ABP50163.1) | Crenarchaeota, Thermoproteales
80 — *Thermoproteus tenax* (WP_014127487.1)
99 — *Haloarcula amylolytica* (WP_008308230.1)
91 — *Haloarcula* sp. SL3 (WP_053967807.1) | Euryarchaeota, Halobacteriales
78 — 97 — *Haloferax volcanii* (WP_004042412.1)
80 — Uncultured marine group II euryarchaeote (EHR76937.1) | Euryarchaeota, MGII/III
100 — MGII/III euryarchaeote KM3_57_F04 (AIF12932.1)
74 — *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)
97 — glpQ *Pasteurella multocida* (AAO73332.1) | glpQ Bacteria
100 — glpQ *Escherichia coli* (KIY30593.1)
— glpQ *Klebsiella pneumoniae* (CDO15614.1)

0.5

Figure 5

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)
uncultured MGII (EHR77227.1)
Euryarchaeota MGII/III

Candidatus Koribacter versatilis (ABF39470.1) **Bacteria, Acidobacteria**

Oscillochloris trichoides **Bacteria, Chloroflexi**
Caldilinea aerophila

Archaeon Loki (KKK41857.1) **'Lokiarchaeum'**

Enterococcus phoeniculicola (WP_010767251.1)
Enterococcus columbae (WP_016183035.1) **Bacteria, Firmicutes**
Halanaerobium praevalens (WP_014553418.1)

Myxococcus xanthus (WP_011554758.1) **Bacteria, δ-Proteobacteria**
Myxococcus stipitatus (WP_015350766.1)

Marinithermus hydrothermalis (WP_013703642.1)

***Natronorubrum bangense* (WP_006067869.1)**
***Haloterrigena turkmenica* (WP_012945206.1)** Euryarchaeota, Halobacteriales
***Haloferax* sp. ATB1 (WP_042663736.1)**

Clostridium cellobioparum (WP_027630730.1)
Clostridiales bacterium DRI-13 (WP_034422850.1) **Bacteria, Firmicutes**
Bacillus sp. FF4 (WP_042456298.1)
Paenibacillus polymyxa (WP_029515007.1)

Homo sapiens (AAH42195.2)
Xenopus laevis (NP_001088934.1)
Ricinus communis (XP_002516359.1)
Nannochloropsis gaditana (AGS78408.1) **Eukarya**
Thalassiosira pseudonana (XP_002290601.1)
Phaeodactylum tricornutum (XP_002181767.1)

0.2

Figure 7

Figure S1

Figure S2

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. Nitrosoarchaeum 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 (YP_007549490.1)
Ca. Nitrosoarchaeum koreensis MY1 (ZP_08667163.1)
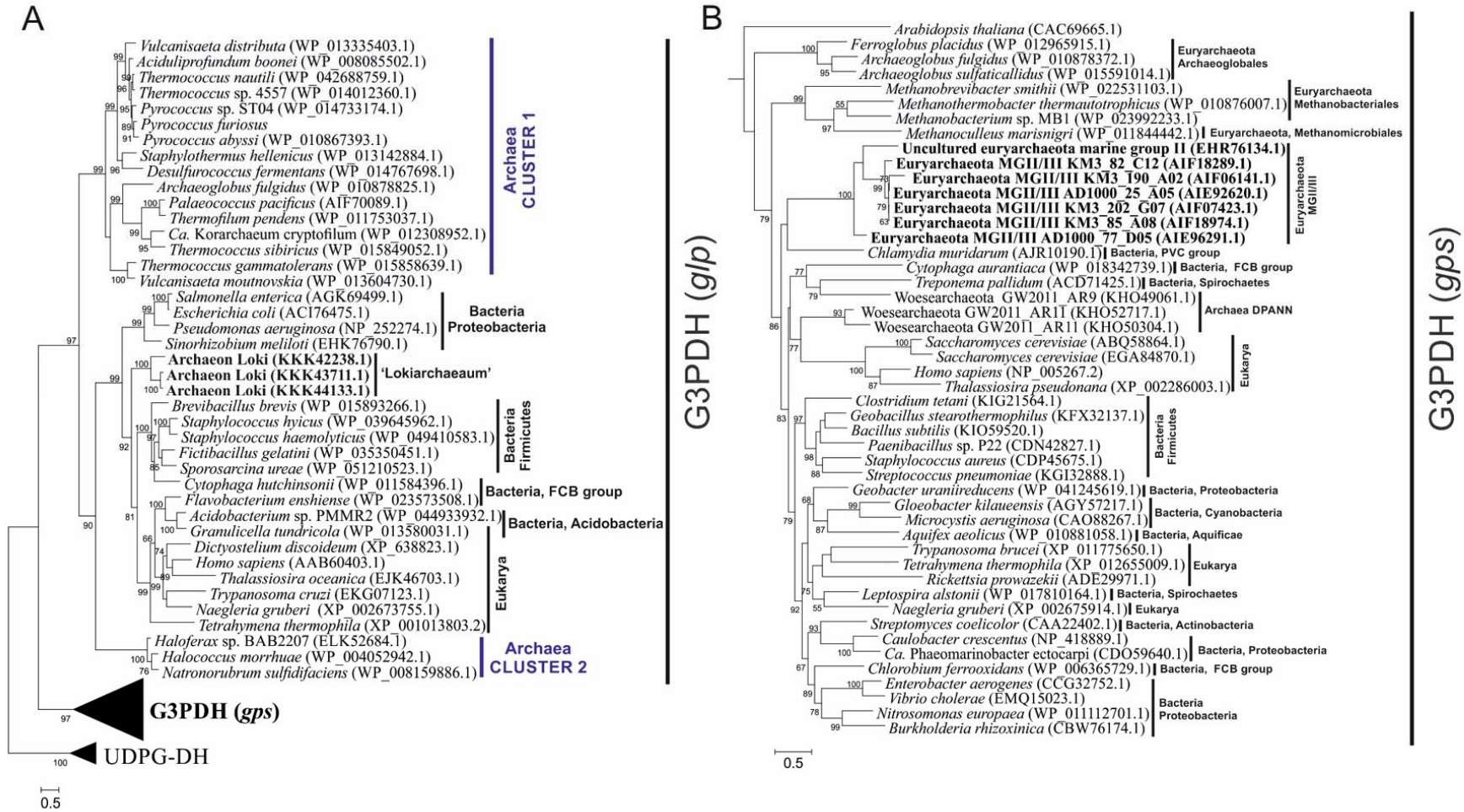*Ca.* Nitrososphaera gargensis (YP_006861308.1)
Sulfolobus solfataricus P2 (AAK40480.1, ubiA-1)]
COQ2 *Saccharomyces cerevisiae* (CAA96321.1)

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
Ignicoccus hospitalis

Crenarchaeota, Acidilobales
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