A Re-evaluation of the Archaeal Membrane Lipid Biosynthetic Pathway

Laura Villanueva, Jaap S. Sinninghe Damsté,

and Stefan Schouten

NIOZ, Royal Netherlands Institute for Sea Research. Department of Marine Organic Biogeochemistry. P.O. Box 59, NL-1790 AB Den Burg, The Netherlands. Correspondence to L.V. e-mail: laura.villanueva@nioz.nl

Note: Terms highlighted in grey are included in the glossary.

Abstract

Archaea produce unique membrane lipids in which isoprenoid alkyl chains are bound through ether linkages to glycerol moieties. With the increasing availability of cultured representatives of the archaea over the past decade, archaeal genomic and membrane lipid composition data have become available. Here, we compare the amino acid sequences of the key enzymes of the archaeal ether lipid biosynthesis pathway and criticality evaluate past studies on the biochemical functioning of these enzymes. Considering these evidences we propose an alternative archaeal lipid biosynthetic pathway based on a multiple-key, multiple-lock mechanism.

In 1977 Woese & Fox[1] proposed a new domain of life, the Archaea (at that time called archaeabacteria), in addition to the Eubacteria (Bacteria) and Eukarya. Although initially it was thought that archaea were confined to extreme environments (e.g. high temperature or salinity, extreme pH), subsequent studies have shown that they occur ubiquitously in non-extreme settings, such as the ocean[2], where they can be substantial contributors to total microbial biomass[3]. Archaea have now been shown to play important roles in global biogeochemical cycles such as the methane[4] and nitrogen[5] cycles.

In addition to their genomic make-up, the domain Archaea has also other traits that distinguish them from Bacteria and Eukarya. One of the most intriguing is the unique structure of their membrane lipids[6]. Bacterial and eukaryotic membranes are composed of fatty acid chains that are linked to the glycerol moiety through ester bonds. These bacterial and eukaryotic lipids are organized in a bilayer structure. In contrast, archaeal membrane lipids are characterized by (i) ether instead of ester linkages between the glycerol moiety and the alkyl chains, (ii) isoprene-based alkyl chains instead of acetate-

based straight alkyl chains as building blocks of the apolar side-chains, and (iii) an opposite stereochemistry of the glycerol phosphate backbone, i.e. *sn*-glycerol-1-phosphate (G1P)[7]. Soon after the discovery of archaeal membrane ether lipids, it was suggested that they could provide an advantage in surviving in extreme environments (e.g. high temperature, high salinity or extreme pH)[7], based on the fact that the ether-linked lipids present in Archaea are chemically more stable than the ester-linked lipids present in Bacteria and Eukarya[8]. This is most likely due to restrictions in the hydrocarbon chain mobility in ether-linked membranes which also may result in reduced permeability of this membranes. However, the discovery of ether lipids in ubiquitous mesophilic/neutrophilic archaea found in the ocean[9] suggested that this hypothesis needed to be re-evaluated.

In addition to an exact answer as to why archaea are producing ether membrane lipids, we also lack an answer to the important question of how they biochemically produce them. Many steps in the archaeal membrane lipid biosynthetic pathway are still unknown and most studies have focused mainly on evolutionary processes involved in the differentiation of bacterial and archaeal membranes[6]. Phylogenetic analyses of the enzymes involved in the archaeal membrane lipid biosynthetic pathway have been performed[10, 11] but were limited to a small number of archaeal genomes available. In light of the recent availability of many more archaeal genome sequences, in particular of members of mesophilic and environmentally important archaea, and the much more detailed information available on membrane lipid composition of archaea[12], it is timely to analyze the relationship between archaeal membrane ether lipid composition and the enzymes involved in their biosynthesis. The analysis of amino acid sequences of key biosynthetic enzymes presented here, as well as a critical evaluation of the current conception of the archaeal membrane ether lipids biosynthetic pathway based on

enzymatic studies in specific archaeal isolates, indicates that the concept of the archaeal membrane lipid biosynthesis pathway has to be reconsidered.

Archaeal phylogeny and membrane lipids

Initial studies based on 16S rRNA gene sequences originally supported a deep split within the Archaea forming two major phyla: Crenarchaeota and Euryarchaeota[13]. Based on culture studies and the analysis of environmental gene sequences Crenarchaeota are thought to consist mostly of hyperthermophiles and thermoacidophiles[14]. Most hyperthermophilic Crenarchaeota have been isolated from geothermally heated soils or waters, sulfur-rich springs, or hydrothermal vents, where they obtain their energy mainly from sulfur-containing compounds[15]. Euryarchaeota are abundant in a wide range of environments and have widely diverse physiological strategies (e.g. halophilic, thermophilic, methanogenic[16]). Horizontal gene transfer (HGT) is thought to have been especially important in the evolution of certain members of the Euryarchaeota. For example, Halobacteriales have acquired several genes from Bacteria and it has been proposed that HGT transformed a methanogen into the common ancestor of the Halobacteria[17]. The evolution of another order of the Euryarchaeota, the Thermoplasmatales, is believed to have involved extensive HGT from Sulfolobales (hyperthermophilic Crenarchaeota) and Bacteria[18–19].

In the last decade several other archaeal phyla have been discovered, e.g. Korarchaeota and Nanoarchaeota[20–21], Thaumarchaeota[22], and the recently proposed 'Aigarchaeota' phylum[23]. Species of the Korarchaeota, Nanoarchaeota and 'Aigarchaeota' have a limited environmental distribution, being mainly found in hot springs, and their physiology is not clear (e.g. REF 20). Thaumarchaeota, by contrast, are widespread in marine, lacustrine and terrestrial environments, as revealed by environmental genomics[24].

Although there is a wide variety of archaeal membrane lipids[25], they typically feature a variation of two main core structures, i.e. *sn*-2,3-diphytanylglycerol diether (archaeol) with phytanyl (C_{20}) chains in a bilayer structure, and *sn*-2,3-dibiphytanyl diglycerol tetraether (also known as glycerol dibiphytanyl glycerol tetraether, GDGT), in which the two glycerol moieties are connected by two C_{40} isoprenoid chains, allowing the formation of monolayer membranes[26–27]. GDGTs can contain 0–8 cyclopentane moieties (i.e. GDGT-x, x equals the number of cyclopentane moieties; REF 12; Table 1). The presence of these cyclopentane moieties is thought to be essential in maintaining functional membranes and cellular homeostasis in situations of extreme pH or thermal stress; the number of cyclopentane moieties increases as growth temperature increases[28] and pH decreases[29–30].

Comparison of an archaeal reference phylogeny with the membrane lipid composition distribution shows that most lipids are not specific for a certain phylogenetic group (Table 1; REF 12). Only the GDGT crenarchaeol[9], containing four cyclopentane moieties and a cyclohexane moiety, is considered to be characteristic of the Thaumarchaeota[31], suggesting that the biosynthesis of the cyclohexane moiety is unique for this phylum. GDGTs are the dominant lipid species in Crenarchaeota and Thaumarchaeota, while euryarchaeotal orders synthesize archaeol (Methanococcales, Halobacteriales, Methanosarcinales), GDGTs (Methanopyrales, Thermoplasmatales, Archaeoglobales, Methanomicrobiales), or both (Thermococcales, Methanobacteriales) (Table 1). GDGT-0 is found in all (hyper)thermophilic Crenarchaeota and several thermophilic euryarchaeotal orders, in some mesophilic methanogenic Euryarchaeota, and in Thaumarchaeota. GDGTs with 1–4 cyclopentane moieties are synthesized by hyperthermophilic Crenarchaeota, Thaumarchaeota, in the thermophilic euryarchaeotal order Thermoplasmatales, and the euryarchaeote '*Ca.* Aciduliprofundum boonei'

100 (member of the DHVE-2 cluster, closely related to the Thermoplasmatales order)[32].
101 However, they are apparently not synthesized by methanogenic Euryarchaeota (Table
102 1). GDGTs with more than four cyclopentane moieties (GDGTs 5–8, Table 1) are rare
103 and only found in hyperthermophilic Crenarchaeota and some hyperthermophilic
104 Euryarchaeota of the Thermoplasmatales order. GDGTs are absent in Halobacteriales
105 (Euryarchaeota) that mainly contain archaeol or extended archaeol with one C_{25}
106 isoprenoid chain[33].

107 ## Archaeal lipid synthesis

108 Previous studies have characterized some of the enzymes involved in the biosynthesis
109 of archaeal membrane ether lipids (FIG. 1). Isopentenyl diphosphate and dimethylallyl
110 diphosphate (DMAPP) serve as basic building blocks of the isoprenoid chains and are
111 synthesized by the mevalonate pathway[6, 34]. DMAPP is thought to be consecutively
112 condensed with several isopentenyl diphosphate units to form geranylgeranyl
113 diphosphate (GGPP, C_{20}) by a short -chain (C_{20}) isoprenyl diphosphate synthase, GGPP
114 synthase (FIG. 1). The subsequent ether bond formation is catalyzed by two
115 prenyltransferases: GGPP is attached to the glycerol-1-phosphate (G1P) to form
116 geranylgeranylglyceryl phosphate (GGGP) catalyzed by the GGGP synthase. The
117 attachment of the second side chain to GGGP generates digeranylgeranylglyceryl
118 phosphate (DGGGP) and is catalyzed by the DGGGP synthase (FIG. 1). This is thought
119 to be followed, after addition of a polar headgroup to the glycerol moiety, by a
120 reduction of the unsaturated isoprenoid chains mediated by geranylgeranyl reductases[6],
121 forming archaeol. The formation of GDGTs is thought to initially involve a coupling of
122 two archaeol molecules through head-to-head condensation of the phytanyl chains (FIG.
123 1). Cyclopentane moieties are subsequently thought to be formed by internal

124 cyclization. These latter two steps are highly unusual since they involve non-activated
125 methyl groups and the enzymes involved are unknown[6, 35].
126 Evidence for the head-to-head coupling of archaeol comes from pulse-chase
127 experiments with cell extracts of the euryarchaeon *Thermoplasma acidophilum*
128 (Thermoplasmatales order) incubated with ^{14}C-mevalonate, which showed incorporation
129 of radioactivity first into archaeol and then into GDGT-0[36]. Furthermore, pulse-chase
130 experiments performed with cell extracts of *T. acidophilum* labeled with ^{14}C-
131 mevalonate and using a squalene epoxidase inhibitor (terbinafine) led to accumulation
132 of archaeol, with a modified headgroup, rather than GDGTs[37]. These experiments
133 suggest archaeol as the precursor of GDGTs. However, Poulter *et al*.[38] studied the *in*
134 *vivo* incorporation of radiolabeled archaeol into cells of the euryarchaeon
135 *Methanospirillum hungatei* (order Methanomicrobiales) and found no incorporation of
136 radioactivity in GDGT-0. Furthermore, radiolabeled phytol, in which there is one
137 double bond, was not incorporated into archaeol and GDGT-0, while geranylgeraniol
138 was efficiently incorporated into both. Similar results were obtained by incorporation of
139 deuterium-labeled DGGGP analogs in *Methanothermobacter thermoautotrophicus*
140 (order Methanobacteriales)[39–40]. The deuterium-labeled DGGGP analogs with a terminal
141 double bond or with a saturated terminal isoprene unit were not incorporated into
142 GDGT-0, and only the DGGGP analog with a terminal isopropylidene group was
143 incorporated into the GDGT. These studies thus suggest that the presence of double
144 bonds in the DGGGP molecule is a prerequisite for the formation of GDGTs, which
145 contradicts the idea that fully saturated phytanyl chains are coupled.
146 Below we focus on three known key enzymes in the formation of glycerol ether
147 lipids formation, i.e. GGPP, GGGP, and DGGGP synthases. We searched for
148 homologues of those enzymes in all archaeal genomes available up to date, compared

149 the amino acid moieties involved in the selection of the substrate, used maximum
150 likelihood analyses to reveal their phylogeny, and compared this with the distribution of
151 ether membrane lipids (Table 1).

152 Isoprenyl diphosphate synthase

153 Isoprenyl diphosphate (IPP) synthases catalyze consecutive condensations of
154 isopentenyl diphosphates with allylic primer substrates to form isoprenoid compounds,
155 including steroids, triterpenoids, carotenoids, prenylated proteins and quinones[41]. IPP
156 synthases synthesize short (i.e. C_{10}–C_{20}) or longer (> C_{20}) prenyl groups. IPP synthases
157 harbour two conserved aspartate-rich motifs typical of prenyltransferases, which form a
158 deep hydrophobic cleft or substrate-binding pocket[42]. Short-chain (up to C_{20}) IPP
159 synthases are characterized by the presence of 'bulky' amino acids, i.e. phenylalanine
160 (F) or tyrosine (Y), as the 5th amino acid residue before the first aspartate-rich motif,
161 which limits the degree of isoprenoid chain elongation to the 20 carbon atoms of the
162 GGPP[42]. Some IPP synthases are flexible in the chain length they synthesize, e.g. the
163 single bifunctional short-chain IPP synthase of *M. thermoautotrophicus* synthesizes
164 both the C_{15} precursor for the synthesis of squalene and GGPP (C_{20}) for the synthesis of
165 archaeal membrane lipids[43].

166 We searched for homologues of IPP synthases in 43 archaeal genomes (Table S1).
167 Some of the identified sequences harbor a small amino acid residue (alanine, A; valine,
168 V; serine, S) in the 5th amino acid residue before the first aspartate-rich motif,
169 classifying them as putative long-chain IPP synthase (Table S1). Long-chain IPP
170 synthases were only detected in species of the Thaumarchaeota phylum, in most orders
171 of the Crenarchaeota, and in the orders Halobacteriales, Methanosarcinales,
172 Archaeoglobales and Thermoplasmatales of the Euryarchaeota (Table S1). The role of
173 the long-chain IPP synthase in these groups is unknown but it has been hypothesized

174 that is related to the synthesis of isoprenoid chains other than for ether lipids[34, 44], such
175 as respiratory quinones[45].

176 Putative short-chain IPP synthases harboring a 'bulky' amino acid residue (Y or F)
177 at position 5 (FIG. 2) were detected in all the archaeal orders (Table S1), suggesting that
178 the archaeal lipid biosynthetic pathway starts with the formation of isoprenoid chains
179 with 20 carbon atoms (GGPP). According to the current picture of the archaeal lipid
180 biosynthetic pathway (FIG. 1), short-chain IPP synthases should always encounter the
181 same substrate (isopentenyl diphosphate units) and yield the same product, i.e. GGPP.
182 However, the substantial differences between IPP synthases at the amino acid level (e.g.
183 FIG. 2) seems at odds with this idea. Rather, the observed large amino acid sequence
184 variability of the IPP synthases and, thus the expected plasticity in the structure of this
185 enzyme suggests structural diversity in the intermediates synthesized from isopentenyl
186 diphosphate units.

187 ## Geranylgeranylglyceryl phosphate synthase

188 The next step in the proposed biosynthetic pathway consists of the formation of an ether
189 linkage between the C-3 of the G1P and GGPP to form GGGP (FIG. 1). This step is
190 mediated by the GGGP synthase, which is selective for the G1P acceptor but also for
191 the isoprenoid chain added, strongly favoring GGPP over shorter or longer chains[47].
192 GGGP synthase represents the first identified triose phosphate isomerase (TIM) barrel
193 structure with a prenyltransferase function, which is thought to be unique to the
194 archaea[48]. GGGP synthase is a homologue of PcrB protein that catalyzes the
195 condensation of G1P with C_{35} heptaprenyl pyrophosphate (HepPP) to
196 heptaprenylglyceryl phosphate (HepGP) in Gram-positive bacteria (e.g. *Bacillus*
197 *subtilis*)[49].

198 The only GGGP synthase characterized in detail so far is that of the euryarchaeon
199 *Archaeoglobus fulgidus*[48], which produces archaeol and GDGT-0 as membrane lipids
200 (Table 1). The crystal structure of this enzyme displays a unique fold acting as a 'greasy
201 slide' and a 'swinging door' due to the replacement of a helix α-3 by a strand that
202 creates a large gap for the product of IPP synthase to fit in[48]. It is thought that a 'bulky'
203 hydrophobic amino acid residue, i.e. tryptophan (W), at position 99 (α4a helix of *A.*
204 *fulgidus*; here referred to as the 'chain-length determination area'; FIG. 3), usually
205 marks the end of the gaps in the barrel and would presumably select for the chain length
206 of the substrate (in this case presumably GGPP, C_{20}). The GGGP synthase of *A. fulgidus*
207 and PcrB from *B. subtilis* share 35% sequence identity and the binding sites for G1P are
208 conserved (FIG. 3; REF 49). Interestingly, the residue corresponding to alanine (A) at
209 position 100 (A_{100}) in PcrB from *Bacillus*, as well as the tyrosine (Y) 104, allow the
210 binding of substrates longer than GGPP, i.e. $>C_{20}$ (REF 50). This A_{100} residue
211 corresponds to W_{99} in the *A. fulgidus* IPGP synthase (FIG. 3). The conversion of A_{100} to
212 W_{100} in PcrB from *Bacillus* has been proven to prevent the formation of C_{35} products,
213 and the one from Y_{104} to A_{104} to allow the formation of longer products up to C_{40} (REF
214 50). Guldan *et al*.[49] also showed that the conversion of W_{99} to A_{99} in the *A. fulgidus*
215 GGGP synthase allowed the protein to use substrates longer than GGPP.
216 We searched for GGGP synthase homologues in 72 archaeal genomes and aligned
217 them with the GGGP synthase sequences from *A. fulgidus* (simplified alignment in FIG.
218 3). Interestingly, the 'bulky' W_{99} amino acid residue found in *A. fulgidus* GGGP
219 synthase, which is believed to restrict the length to C_{20} substrates, was only detected in
220 sequences of the euryarchaeotal orders Archaeoglobales, Halobacteriales and
221 Methanomicrobiales, while in the remaining sequences either a glycine (G) or alanine
222 (A), both small amino acid residues, were found in the corresponding position. This

223 amino acid position also coincides with the presence of the A_{100} residue found in PcrB
224 of *Bacillus*, which allows it to use longer (>C_{20}) isoprenyl chains as substrates. Indeed,
225 the secondary structure of the partial amino acid sequences (FIG. 3) showed that the
226 G/A_{99} residue observed in most archaeal sequences (other than the euryarchaeotal
227 orders Archaeoglobales, Halobacteriales and Methanomicrobiales) was included in an
228 α-helix structure as in the case of W_{99} of *A. fulgidus* (α4a helix according to[48]).
229 Moreover, the amino acid sequence alignment of GGGP synthases (FIG. 3), also reveals
230 the presence of a 'bulky' tryptophan (W) residue in the α5' helix (as defined for *A.*
231 *fulgidus*) in all the thaumarchaeotal sequences (data not shown), while in the
232 corresponding position in the rest of the sequences there is a small amino acid residue
233 (glycine, G or alanine, A). In fact, the protein secondary structure analysis does not
234 predict the existence of an α-helix in this position in the archaeal GGGP synthases other
235 than *A. fulgidus* and the PcrB of *Bacillus subtilis* (FIG. 3). This amino acid change in
236 the thaumarchaeotal sequences would certainly affect the positioning of the isoprenyl
237 substrate in the GGGP synthase TIM-barrel structure.
238 These key differences in the amino acid composition of GGGP synthases suggest
239 that their structure, as well as the amino acid interactions between the isoprenyl
240 substrate and the TIM-barrel structure of the GGGP synthase, are likely to be quite
241 different from the enzyme characterized in the euryarchaeon *A. fulgidus*. Our analysis of
242 the amino acid sequence diversity of archaeal GGGP synthases strongly suggests that
243 they harbor functional plasticity and enable the selection of substrates that are longer
244 than GGPP.
245 The phylogeny of GGGP synthase reveals two main clusters (FIG. 4). Cluster 1
246 includes the euryarchaeotal orders Archaeoglobales, Methanomicrobiales and
247 Halobacteriales while cluster 2 can be further subdivided into cluster 2A, which

includes the Thaumarchaeota and the Crenarchaeota, and cluster 2B, which includes the remaining euryarchaeotal groups. The large difference between the three euryarchaeotal orders in cluster 1 and the other Archaea (FIG. 4) has been previously related to the presence of an ancestral divergent type of GGGP synthase in Halobacteria[10]. Interestingly, GGGP synthase sequences seem to roughly cluster according to the presence/absence of ring moieties in the membrane lipids with the notable exception of GGGP synthases from the Thermoplasmatales order. However, the phylogenetic positioning of GGGP synthases of the Thermoplasmatales order has probably been strongly affected by events of vertical inheritance from an euryarchaeotal ancestor[18].

Digeranylgeranylglyceryl phosphate synthase

The next step in the proposed pathway consists of the catalysis of GGGP by the DGGGP synthase to form DGGGP (FIG. 1). DGGGP synthase is a member of the UbiA prenyltransferase family, which, apart from being involved in the archaeal ether lipid formation, also transfers prenyl groups to hydrophobic ring structures such as quinones, hemes, chlorophylls, vitamin E, or shikonin[56].

We searched for putative DGGGP synthases in archaeal genomes based on protein homology with the DGGGP synthase of the crenarchaeota *Sulfolobus solfataricus* which function has been previously tested experimentally[56]. DGGGP synthases were highly divergent between archaeal orders and no clustering was observed (FIG. 5). The most striking observation, however, is the lack of homologues of DGGGP synthases in the Thaumarchaeota, as observed previously for a more limited database[34]. However, several putative protoheme IX farnesyltransferases and other prenyltransferases were identified in thaumarchaeotal genomes (FIG. 5). The inability to clearly identify DGGGP synthases in thaumarchaeotal genomes suggests the existence of very divergent DGGGP synthases in this phylum compared to others. Interestingly,

Thaumarchaeota are the only archaea capable of biosynthesizing GDGTs containing a cyclohexane moiety (crenarchaeol). Sinninghe Damsté *et al*.[9] showed that this additional cyclohexane ring led to a 'bulge' in one of the biphytanyl chains that prevents the dense packing of the biphytanyl chains in the thaumarchaeotal GDGT membranes. Possibly, this 'bulky' biphytanyl chain can only be accommodated by a DGGGP synthase that is rather different from those using regular biphytanyl chains as substrates.

An alternative pathway for ether lipid biosynthesis

Our results, together with the, sometimes contradicting, circumstantial evidences on e.g. the substrates utilized for formation of GDGTs (REF 37 *vs* REFs 38, 40), are difficult to reconcile with the current ether membrane lipid biosynthetic pathway (FIG. 1). We, therefore, propose an alternative pathway that better explains our, and earlier[6], observations, while at the same time circumventing some unresolved issues (i.e. head-to-head condensation of saturated phytanyl chains; ring formation) in the currently proposed pathway. The new hypothetical pathway is based on a multiple-key, multiple-lock mechanism for which multiple keys with different structures, due to the presence/absence of rings, must accommodate and specifically interact at the molecular level with different locks (i.e. GGGP and DGGGP synthases) (FIG. 6). The large difference in amino acid sequences of IPP, GGGP and DGGGP synthases indicate a larger functional plasticity than previously anticipated. One explanation could be that the rings are already present in the prenyl chains before they are coupled to the glycerol unit (for GGGP and DGGGP synthases). Formation of ring structures at this early stage would avoid the need to form them by internal cyclization of saturated chains.

296 Potentially this cyclization may happen simultaneously with the chain elongation using
297 isopentenyl diphosphate (FIG. 6).

298 The presence of small amino acid residues in the chain-length determination area of
299 archaeal GGGP synthases (cluster 2; FIG. 4) indicates that these synthases could
300 accommodate prenyl substrates longer than C_{20}. This implies that the substrates of
301 GGGP synthases could be C_{40} prenyl substrates containing ring moieties. Thus, head-to-
302 head condensation of two C_{20} isoprenyl molecules may take place prior to attachment to
303 the glycerol unit. These C_{20} isoprenyl molecules contain an isopropylidene double bond
304 required for such condensation[39–40], except for the unusual C_{20} isoprenyl unit with a
305 cyclohexane moiety hypothesized for Thaumarchaeota (FIG. 6). This eliminates the
306 need for an unusual (and experimentally poorly supported) condensation of the two
307 saturated phytanyl chains of archaeol (FIG. 1). This proposed head-to-head
308 condensation of two C_{20} isoprenyl molecules could be potentially catalyzed by phytoene
309 synthase that converts two GGPP C_{20} into phytoene (C_{40}) by tail-to-tail coupling in the
310 second step in the biosynthesis of carotenoids[57]. Interestingly, homologues of phytoene
311 synthase have been annotated in archaeal genomes (Table S2) of the orders Sulfolobales
312 and Themoproteales of the Crenarchaeota phylum, and in the orders
313 Thermoplasmatales, Methanomicrobiales, Methanobacteriales, Methanosarcinales and
314 Halobacteriales of the Euryarchaeota phylum, but not in any of the available genomes of
315 the Thaumarchaeota phylum. The latter might not be surprising as the intermediate C_{20}
316 GGGP containing the cyclohexane moiety, as hypothesized in our pathway, does not
317 possess a terminal isopropylidene moiety (FIG. 6).

318 After formation of the GGGP, the second IPP unit is attached to the glycerol moiety.
319 The potential presence of ring moieties before the catalysis mediated by GGGP and
320 DGGGP synthases would again explain the diversification of DGGGP synthases

observed in our study. It would also explain the apparent lack of the DGGGP synthase-coding gene in genomes of the Thaumarchaeota phylum by the presence of a more divergent DGGGP synthase that can accommodate the bulky presence of the unique cyclohexane moiety of the biphytanyl chain. Finally, a second glycerol moiety is attached followed by saturation of the isoprenyl chains and attachment of the headgroup. Considering the alternative pathway presented here, we propose to rename the GGGP and DGGGP synthases as isoprenylglyceryl phosphate (IPGP) synthase and di-isoprenylglyceryl phosphate (DIPGP) synthase, respectively, in order to reflect the the more general nature of these enzymes and their independence with respect to the chain length of their substrate (FIG. 6).

The proposed pathway is consistent with the analysis of the sequences of key enzymes of the pathway observed in our study, as well as most of the experimental evidence for the different GDGT biosynthetic steps. Furthermore, the isoprenoid glycerol dialkanol diethers (compounds with C_{40} isoprenoid chains and ring moieties but only attached to one glycerol group), recently detected in archaeal cultures[58–59], as well as the biphytane diols detected in the environment[60], are all products of potential intermediates that fit well with our proposed biosynthetic pathway. Clearly, the steps proposed in our hypothetical biosynthetic scheme require experimental verification using archaeal cultures, specifically of the Thaumarchaeota phylum. Such results, together with further genomic data mining, will shed further light on the unique membrane lipid pathway of the Archaea.

References

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

2. DeLong, E. F. & Pace, N. R. Environmental diversity of Bacteria and Archaea. *Syst. Biol.* **50**, 470–478 (2001).

3. Karner, M. B., DeLong, E. F. & Karl, D. M. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**, 507–510 (2001).

4. Jarrell, K. F., Walters, A. D., Bochiwal, C., Borgia, J. M., Dickinson, T. & Chong, J. P. Major players on the microbial stage: why archaea are important. *Microbiology* **157**, 919–936 (2011).

5. Wuchter, C., Abbas, B., Coolen, M. J. L., Herfort, L., Timmers, P., Strous, M., van Bleijswijk, J., Teira, E., Herndl, G. J., Middelburg, J. J., Schouten, S., & Sinninghe Damsté, J. S. Archaeal nitrification in the ocean. *Proc. Nat. Acad. Sci. USA* **103**, 12317–12322 (2006).

6. Koga, Y. & Morii, H. Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations. *Microbiol. Mol. Biol. Rev.* 71, 97–120 (2007).
An extensive review of the different steps in the archaeal ether lipid biosynthetic pathway and the experimental data supporting this.

7. Kates, M. Biology of halophilic bacteria, Part II. Membrane lipids of extreme halophiles: biosynthesis, function and evolutionary significance. *Experientia* **49**, 1027–1036 (1993).

8. Thompson, D. H., Wong, K. F., Humphry-Baker, R.,Wheeler , J. J., Kim, J.-M. & Rannanavare, S. B. Tetraether bolaform amphiphiles as models of archae-bacterial membrane lipids: Raman spectroscopy, ^{31}P NMR, X-rays scattering, and electron microscopy. *J. Am. Chem. Soc.* **114**, 9035–9042 (1992).

9. Sinninghe Damsté, J. S., Hopmans, E. C., Schouten, S., van Duin, A. C. T. & Geenevasen, J. A. J. Crenarchaeol: the characteristic core glycerol dibiphytanyl glycerol tetraether membrane lipid of cosmopolitan pelagic crenarchaeota. *J. Lipid Res.* **43**, 1641–1651 (2002).
First description of crenarchaeol structure as a characteristic lipid of marine Thaumarchaeota.

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

11. Lombard, J., Lopez-Garcia, P., & Moreira D. Phylogenomic investigation of phospholipid synthesis on Archaea. *Archaea* **2012**, 630910 (2012).

12. Schouten, S, Hopmans, E. C. & Sinninghe Damsté, J. S. The organic geochemistry of glycerol dialkyl glycerol tetraether lipids: A review. *Org. Geochem*. **54**, 19–61 (2013).
A recent review on the sources, analysis methods, and applications of archaeal ether lipids in organic geochemistry.

13. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA*. **87**, 4576–4579 (1990).

14. Gribaldo, S. & Brochier-Armanet, C. The origin and evolution of Archaea: a state of the art. *Philos Trans. R. Soc. Lond. B Biol. Sci.* **361**, 1007–1022 (2006).

15. Stetter, K.O., Fiala, G., Huber, G., Huber, R. & Segerer, A. Hyperthemophilic microorganisms. *FEMS Microbiol. Rev.* **75**, 117–124 (1990).

16. Madsen, E. L. Environmental Microbiology: From Genomes to Biogeochemistry. Wiley-Blackwell eds . (2008).

17. Nelson-Sathi, S., Dagan, T., Janssen, A., Steel, M, McInerney, J. O., Deppenmeier, U. & Martin, W. F. Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. *Proc. Natl. Acad. Sci. USA.* doi:10.1073/pnas.1209119109 (2012).

18. Ruepp, A., Graml, W., Santos-Martinez, M. L., Koretke, K. K., Volker, C., Mewes, H. W, et al. The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature* **407**, 508–5139 (2000).

19. Brochier-Armanet, C., Forterre, P. & Gribaldo, S. Phylogeny and evolution of the Archaea: one hundred genomes later. *Curr. Opin. Microbiol*. **14**, 274–281 (2011).

20. Elkins, J. G., Podar, M., Graham, D. E., Makarova, K. S., Wolf, Y., Randau, L., Hedlund, B. P., Brochier-Armanet, C., Kunin, V., Anderson, I et al. A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl. Acad. Sci. USA*. **105**, 8102–8107 (2008).

21. Huber, H., Hohn, M. J., Stetter, K. O., & Rachel, R. The phylum Nanoarchaeota: present knowledge and future perspectives of a unique form of life. *Res. Microbiol*. **154**, 165–171 (2003).

22. Brochier-Armanet, C., Boussau, B., Gribaldo, S. & Forterre, P. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat. Rev. Microbiol.* **6**, 245–252 (2008).
Analysis report suggesting the classification of Thaumarchaeota as a separate phylum of the Archaea.

23. Nunoura, T.,Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., Chee, G-J., Hattori, M., Kanai, A., Atomi, H., Takai, K. & Takami, H. Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res.* 39, 3204–3223 (2011).

24. Pester, M., Schleper, C., & Wagner, M. The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. *Curr. Opin. Microbiol.* **14**, 300–306 (2011).

25. Koga, Y., & Morii, H. Recent advances in structural research on ether lipids from Archaea including comparative and physiological aspects. *Biosci. Biotechnol. Biochem.* **69**, 2019–2034 (2005).

26. Koga, Y., Akagawa-Matsushita, M., Ohga, M. & Nishihara, M. Taxonomic significance of the distribution of component parts of polar ether lipids in methanogens. *Syst. Appl. Microbiol*. **16**, 342–351 (1993).

425 27. Langworthy, T.A. Lipids of archaebacteria. In The Bacteria (Woese, C. R. & Wolfe,
426 R. S., ed.) 459–497 (New York, Academic Press, 1985).

427 28. Uda, I, Sugai, A., Itoh, Y. H., Itoh, T. Variation in molecular species of polar lipids
428 from *Thermoplasma acidophilum* depends on growth temperature. *Lipids* **36**, 103–
429 105 (2001).

430 29. Chong, P. L. G. Archaebacterial bipolar tetraether lipids: physico-chemical and
431 membrane properties. *Chem. Phys. Lip.* **163**, 253–65 (2010).

432 30. Macalady, J. L., Vestling, M. M., Baumler, D., Boekelheide, N., Kaspar, C. W. &
433 Banfield, J. F. Tetraether-linked membrane monolayers in *Ferroplasma* spp: a key
434 to survival in acid. *Extremophiles* **8**, 411–419 (2004).

435 31. Pitcher, A., Hopmans, E. C., Mosier, A. C., Park, S., Rhee, S., Francis, C. A.,
436 Schouten, S. & Sinninghe Damsté, J. S. Core and Intact Polar Glycerol Dibiphytanyl
437 Glycerol Tetraether Lipids of Ammonia-Oxidizing Archaea Enriched from Marine
438 and Estuarine Sediments. *Appl. Environ. Microbiol.* **77**, 3468–3477 (2011).

439 32. Reysenbach, A. L., Liu, Y., Banta, A. B., Beveridge, T. J., Kirshtein, J. D.,
440 Schouten, S., Tivey, M. K., Von Damm, K. & Voytek, M. A. Isolation of a
441 ubiquitous obligate thermoacidophilic archaeon from deep-sea hydrothermal vents.
442 *Nature* **442**, 444–447 (2006).

443 33. De Rosa, M., Gambacorta, A., Nicolaus, B., Ross, H. H. M., Grant, W. D. &
444 Bu'Lock, J. D. An asymmetric archaebacterial diether lipid from alkaliphilic
445 halophiles. *J. Gen. Microbiol.* **128**, 343–348 (1982).

446 34. Matsumi, R., Atomi, H., Driessen, A. J. & van der Oost, J. Isoprenoid biosynthesis
447 in Archaea--biochemical and evolutionary implications. *Res. Microbiol*. **162**, 39–52
448 (2011).

449 35. Lai, D. Isoprenoid ether lipid biosynthesis in the extremophile, *Archaeoglobus*
450 *fulgidus*. University of California, Los Angeles (2009) PhD dissertation.

451 36. Langworthy, T. A. Turnover of di-O-phytanylglycerol in Thermoplasma. *Rev.*
452 *Infect. Dis.* **4**, S266 (1982).

453 37. Nemoto, N., Shida, Y., Shimada, H., Oshima, T., & Yamagishi, A. Characterization
454 of the precursor of tetraether lipid biosynthesis in the thermoacidophilic archaeon
455 *Thermoplasma acidophilum*. *Extremophiles* **7**, 235–243 (2003).

456 38. Poulter, C. D., Aoki, T., & Daniels, L. Biosynthesis of isoprenoid membranes in the
457 methanogenic archaebacterium *Methanospirillum hungatei*. *J. Am. Chem. Soc.* **110**,
458 2620–2624 (1988).

459 39. Eguchi, T., Takyo, H., Morita, M., Kakinuma, K., & Koga, Y. Unusual double-bond
460 migration as plausible key reaction on the synthesis of the isoprenoid membrane
461 lipids of methanogenic archaea. *J. Chem. Soc. Chem. Commun.* **2000**, 1545–1546
462 (2000).
463

40. Eguchi, T., Nishimura, Y., & Kakinuma. Importance of the isopropylene terminal of geranylgeranyl group for the formation of tetraether lipid in methanogenic archaea. *Tetrahedron Lett.* **44**, 3275–3279 (2003).

41. Wang, K. C., & Ohnuma, S-I. Isoprenyl diphosphate synthases. *Biochim. Biophys. Acta*. **1529**, 33–48 (2000).
A review of the classification, catalytic mechanism, and chain length determination of isoprenyl diphosphate synthases.

42. Wang, K. & Ohnuma, S. Chain-length determination mechanism of isoprenyl diphosphate synthases and implications for molecular evolution. *Trends. Biochem. Sci.* **24**, 445–451 (1999).

43. Chen, A., & Poulter, C. D. Purification and characterization of farnesyl diphosphate/geranylgeranyl diphosphate synthase. A thermostable bifunctional enzyme from *Methanobacterium thermoautotrophicum*. *J. Biol. Chem.* **268**, 11002–11007 (1993).

44. Ogawa, T., Yoshimura, T., & Hemmi, H. Geranylfarnesyl diphosphate synthase from *Methanosarcina mazei*: different role, different evolution. *Biochem. Biophys. Res. Commun.* **393**, 16e20 (2010).

45. Peters, K. E., Wlaters C. C., and Moldowan, J. M. The biomarker guide. Cambridge University Press (2007).

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

47. Zhang, D. & Poulter, C. D. Biosynthesis of archaebacterial ether lipids. Formation of ether linkages by prenyltransferases. *J. Am. Chem. Soc.* **115**, 1270–1277 (1993).

48. Payadeh, J., Fujihashi, M., Gillon, W. & Pai, E. F. The crystal structure of (S)-3-O-geranylgeranylglyceryl phosphate synthase reveals an ancient fold for an ancient enzyme. *J. Biol. Chem* **281**, 6070–6078 (2006).
First report of the crystal structure of an archaeal geranylgeranylglyceryl phosphate synthase.

49. Guldan, H., Matysik, F-M., Bocola, M., Sterner, R., & Babinger, P. Functional assignment of an enzyme that catalyzes the synthesis of an Archaea-type ether lipid in Bacteria. *Angew. Chem. Int. Ed.* **50**, 8188–8191 (2011).

50. Ren, F., Feng, X., Ko, T-P., Huang, C-H., Hu, Y., Chan, H-C., Liu, Y-L., Wang, K., Chen, C-C., Pang, X., He, M., Li, Y., Oldfield, E., & Guo, R-T. Insights into TIM-barrel prenyl transferase mechanisms: Crystal structures of PcrB from *Bacillus subtilis* and *Staphylococcus aureus*. *ChemBioChem.* **14**, 195–199 (2013).

51. Cole, C., Barber, J. D., & Barton, G. J. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* **35**, W197–W201 (2008).

52. Talavera, G., & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).

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

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

55. Letunic, I., & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).

56. Hemmi, H., Shibuya, K., Takahashi, Y., Nakayama, T. & Nishino, T. J. (S)-2,3-Di-O-geranylgeranylglyceryl phosphate synthase from the thermoacidophilic archaeon *Sulfolobus solfataricus*. Molecular cloning and characterization of a membrane-intrinsic prenyltransferase involved in the biosynthesis of archaeal ether-linked membrane lipids. *Biol. Chem.* **279**, 50197–50203 (2004).

57. Klassen, J. L. Phylogenetic and evolutionary patterns in microbial carotenoid biosynthesis are revealed by comparative genomics. *Plos One* **5**, e11257 (2010).

58. Knappy, C.S. & Keely, B. J. Novel glycerol dialkanol triols in sediments: transformation products of glycerol diphytanyl glycerol tetraether lipids or biosynthetic intermediates? *Chem. Commun.* **48**, 841–843 (2012).

59. Liu, X. -L., Lipp, J. S., Schröder, J. M., Summons, R. E. & Hinrichs, K. -U. Isoprenoid glycerol dialkanol diethers: A series of novel archaeal lipids in marine sediments. *Org. Geochem.* **43**, 50–55 (2012).

60. Schouten, S., Hoefs, M. J. L., Koopmans, M. P., Bosch, H-J., & Sinninghe Damsté, J. S. Structural characterization, ocurrence and fate of archaeal ether-bound acyclic and cyclic biphytanes and corresponding diols in sediments. *Org. Geochem.* **29**, 1305–1319 (1998).

Competing interest statement

The authors declare no competing financial interest.

Figure legends

Figure 1. **Current conception of the archaeal lipid biosynthetic pathway** (after[6]).

The two basic building blocks are the five-carbon compound isopentenyl phosphate and its isomer dimethylallyl diphosphate (DMAPP) are synthesized by the mevalonate pathay in Archaea[6]. DMAPP consecutively condenses with several isopentenyl diphosphate units to form geranylgeranyl diphosphate (GGPP, C_{20}) by an isoprenyl diphosphate (IPP) synthase, GGPP synthase. Dihydroxyacetone phosphate (DHAP) is catalyzed to form glycerol-1-phosphate (G1P). The formation of the two ether bonds between G1P and the GGPP units is catalyzed by the geranylgeranylglyceryl phosphate (GGGP) synthase and the digeranylgeranylglyceryl phosphate (DGGGP) synthase. Then, CDP (cytidine-diphosphate)-diglyceride synthase replaces the phosphate group of the unsaturated DGGGP by CDP generating CDP-DGGGP (unsaturated), which is then replaced by a polar headgroup by a CDP-alcohol phosphatidyl transferase[34]. Saturation of the side chains is supposed to be mediated by geranylgeranyl reductases. The formation of GDGTs is thought to involve a head-to-head coupling between the two archaeol lipids followed by internal cyclization to form cyclopentane moieties. The latter reactions are highly unusual and the enzymes involved are unknown.

Figure 2. **Partial isoprenyl diphosphate (IPP) synthases protein alignment.**

Alignment of amino acid sequences of putative IPP synthases identified in genomes of different archaeal orders showing a 'bulky' amino acid residue (tyrosine, Y; phenylalanine, F) at the 5th position before the first aspartate (D)-rich motif, indicating that they are short-chain IPP synthases elongating the isoprenoid chain up to 20 carbon atoms.

Sequences were aligned by MUSCLE (multiple sequence comparison by log-expectation; REF 46). Species detailed in the alignment: S.acidocaldarius (*Sulfolobus acidocaldarius*), D.kamchatkensis (*Desulfurococcus kamchatkensis*), A.fulgidus (*Archaeoglobus fulgidus*), M.smithii (*Methanobrevibacter smithii*), M.thermophila (*Methanosaeta thermophila*), M.maripaludis (*Methanococcus maripaludis*), M.hungatei (*Methanospirillum hungatei*), T.acidophilum (*Thermoplasma acidophilum*), N.maritimus (*Nitrosopumilus maritimus*). Accession numbers are listed in Table S1.

Figure 3. **Partial geranylgeranylglyceryl phosphate (GGGP) synthase protein alignment**.

Annotated putative GGGP synthases of representatives of different archaeal orders are aligned and compared to the GGGP synthase protein of *A. fulgidus* which crystalline structure has been previously determined[48]. The black star indicates the position of the amino acid residue corresponding to the W_{99} position of *A. fulgidus*. The red star indicates the position containing a tryptophan (W) in the thaumarchaeotal sequences as discussed in the text. The chain-length determination area is arbitrary and indicated for clarification purposes. Amino acids: W (tryptophan), A (alanine), Y (tyrosine), E (glutamine). Location of α-helixes according to the *A. fulgidus* crystal structure[48] in the partial sequence is indicated above the alignment. Black boxes surrounding amino acid

sequences in the alignment correspond to α-helix prediction by the Jpred 3 server[51]. Cluster 1, 2A and 2B correspond to the clusters also indicated in Figure 4.

Sequences were aligned by MUSCLE[46]. Species detailed in the alignment: A.fulgidus (*Archaeoglobus fulgidus*), H.salinarium (*Halobacterium salinarium*), M.limicola (*Methanoplanus limnicola*), N.maritimus (*Nitrosopumilus maritimus*), T.neutrophilus (*Thermoproteus neutrophilus*), A.pernix (*Aeropyrum pernix*), S.acidocaldarius (*Sulfolobus acidocaldarius*), D.kamchatkensis (*Desulfurococcus kamchatkensis*), M.maripaludis (*Methanococcus maripaludis*), T.litoralis (*Thermococcus litoralis*), M.marburgensis (*Methanothermobacter marburgensis*), T.volcanicum (*Thermoplasma volcanicum*), unc. Euryarchaeota A.boonei (unclassified euryarchaeota *Aciduliprofundum boonei*), M.thermophila (*Methanosaeta thermophila*), PcrB Bs (PcrB protein of *Bacillus subtilis*; accession number YP_007532597.1).

Figure 4. **Maximum likelihood tree based on the protein sequences of archaeal putative geranylgeranylglyceryl phosphate (GGGP) synthases.**

Cluster 1 consists on divergent putative GGGP synthases of the euryarchaeotal orders Archaeoglobales, Methanomicrobiales and Halobacteriales. The second cluster is subdivided into cluster 2A, which includes GGGP synthases of the Thaumarchaeota and the Crenarchaeota, and cluster 2B, which includes the remaining GGGP synthases of other euryarchaeotal groups.

The scale bar represents number of substitutions per site. Abbreviations: THAUM: Thaumarchaeota; CREN: Crenarchaeota; EURY: Euryarchaeota. The colored symbols indicate the presence of the various membrane lipids (Table 1); Archaeol (dark blue circle); extended archaeol (light blue pentagon); GDGT-0 (red rectangle; GDGT-1–4 (yellow triangle); GDGT-5–8 (purple hexagon); Crenarchaeol (green cross). Sequences were aligned using MUSCLE[46]. Alignment was trimmed in Gblocks 0.91b with relaxed parameters[52] and manually curated. Phylogenetic tree was computed by PHYML v3.0[53] using the LG model plus gamma distribution and invariant site (LG+G+I) indicated by ProtTest 2.4[54]. Branch support was calculated with the approximate likelihood ratio test (aLRT) and indicated on the branches (color code in the nodes: red (≥90%), blue (≥70%, <90%) and green (≥50%, <70%), less than 50% is not shown). Trees were edited in iTOL[55].

Figure 5. **Maximum likelihood tree based on the protein sequences of archaeal putative digeranylgeranylglyceryl phosphate (DGGGP) synthases and thaumarchaeotal prenyltransferases.**

Star symbols indicate *Sulfolobus solfataricus* DGGGP synthase (AAK40896), and *S. solfataricus* UbiA-1 (AAK4048.1) previously tested[56]. †Ca. Caldiarchaeum subterraneum Aigarchaeota phylum[23]. For explanation of symbols and legends see FIG. 4. The tree was computed as described in the legend of FIG.4.

Figure 6. **Alternative archaeal lipid biosynthesis scheme based on a multiple-key, multiple-lock mechanism.**

Isoprenyl diphosphate synthases generate C_{20} isoprenoid units with or without ring moieties during their catalysis. Triangles indicate the introduction of the cyclohexane moiety in the precursor of crenarchaeol in Thaumarchaeota. Condensation of two C_{20}

614 isoprenoid units produce a variety of C_{40} substrates (multiple-keys) that are then used as
615 substrate by isoprenylglyceryl phosphate (IPGP) (*) and di-isoprenylglyceryl phosphate
616 (DIPGP) (**) synthases (multiple-locks), followed by the attachment of the 2nd glycerol
617 moiety, saturation of the isoprenoid chains, and final attachment of the headgroups.
618 Note that the hydrogenation step is indicated here after assembly of the GDGT but
619 potentially could also occur prior to attachment of IPGP to the glycerol moiety. The
620 formation of the cyclohexane moiety in Thaumarchaeota is indicated here during the
621 formation of the C_{20} isoprenoid but this leads to an intermediate without a terminal
622 double bond potentially inhibiting head-to-head-coupling of C_{20} isoprenyl units.

623

624 **Table 1.** Distribution of archaeal membrane lipids in different orders of the
625 Euryarchaeota, Crenarchaeota and Thaumarchaeota.
626
627 **Table S1.** Isoprenyl diphosphate synthases in archaeal genomes.

628 **Table S2.** Squalene/phytoene synthase homologues annotated in archaeal genomes.

629

Glossary

Isoprenoid: Group of natural products with diverse structures composed of various numbers of isopentenyl (C_5) pyrophosphate (IPP) units

Phytanyl: Saturated chain composed of 4 head-to-tail linked isoprene units (C_{20} isoprenoid).

Biphytanyl: Molecule composed of two head-to-head condensed phytanyl units (C_{40} isoprenoid).

Hyperthermophile: Organism that has an optimal growth temperature of at least 80°C.

Thermoacidophile: Combination of thermophile and acidophile (thrive under highly acidic conditions, around pH 2.0 or below), microorganisms that thrive in acid, sulfur rich, and high temperature environments.

Halophile: Extremophilic organism that thrives at high concentrations of salt.

Methanogen: Archaeon that produces methane under anoxic conditions.

Horizontal gene transfer: Transfer of genetic material between different species of microorganisms in which the acquired genes are transmitted to the next generation as the cell divides.

Mesophile: Organism that grows at a moderate temperature, typically between 20 and 45°C.

Diphosphate: Also known as pyrophosphate, ester containing two phosphate groups.

Allylic: Double bond at the terminal position of a carbon chain.

Prenyltransferases: Enzymes that transfer (iso)prenyl moieties to acceptor molecules.

Head-to-head condensation: Coupling of two isoprenyl units at position C1 of both units.

Isopropylidene: An isopropyl moiety with a terminal double bond.

Squalene: Biochemical precursor of the steroid and triterpenoid families. Synthesized by tail-to-tail condensation of farnesyl pyrophosphate (C_{15}) by squalene synthase.

Phytol: acyclic diterpene (terpene consist of two or more isoprene C_5H_8 units) alcohol.

Geranylgeraniol: diterpenoid alcohol (3,7,11,15-tetramethyl-2,6,10,14-hexadecatraen-1-ol).

Paralogues: Genes that derive from the same ancestral gene.

Online Summary

- Archaea were initially thought to be confined in extreme environments but now they are known to occur ubiquitously in nature and be important players in global biogeochemical cycles. Archaea are characterized by their unique membrane lipids containing isoprene units linked to the glycerol backbone by ether bonds (archaeol, C_{20}, in a bilayer and glycerol diakyl glycerol tetraether, GDGT, C_{40} in a monolayer).

- Comparison of the phylogenetic composition of Archaea with the distribution of membrane ether lipid shows that most lipids are not specific for a certain phylogenetic group. Only the GDGT crenarchaeol, containing four cyclopentane moieties and a cyclohexane moiety, is considered to be characteristic of the Thaumarchaeota, suggesting that the biosynthesis of the cyclohexane moiety is unique within this phylum.

- The current conception of the archaeal membrane ether lipid biosynthetic pathway involves the condensation of units of isopentenyl diphosphate to form geranylgeranyl (GGPP, C_{20}) by a GGPP synthase. The formation of the two ether bonds is catalyzed by the geranylgeranylglyceryl phosphate (GGGP) synthase and the digeranylgeranylglyceryl phosphate (DGGGP) synthase. The formation of GDGTs is thought to involve a head-to-head coupling between the two archaeol lipids followed by internal cyclization to form cyclopentane moieties. The latter reactions are highly unusual and the enzymes involved are unknown.

- The analysis of the amino acid sequence of most of the archaeal GGGP synthases suggest that they could accommodate substrates $>C_{20}$ and with rings already present.

- The synthesis of the unique cyclohexane moiety-containing GDGT crenarchaeol by Thaumarchaeota might explain the inability to annotate DGGGP synthases in thaumarchaeotal genomes, as a yet-unknown or divergent DGGGP synthase would be required to accommodate the isoprenyl chain containing the ‘bulky’ cyclohexane moiety.

- An alternative archaeal lipid biosynthetic pathway pathway is presented based on a multiple-key, multiple-lock mechanism for which multiple keys with different configurations due to the presence of rings, would need to accommodate and specifically interact at the molecular level with different locks (isoprenylglyceryl phosphate, IPGP and di-isoprenylglyceryl phosphate, DIPGP synthases). This pathway is consistent with most of the phylogenetic relationships observed in our study as well as with most of the experimental evidence for the different GDGT biosynthetic steps, and it is supported by possible intermediates previously described.

Author biographies

Dr. Laura Villanueva

Research Scientist in the Department of Marine Organic Biogeochemistry at the Royal Netherlands Institute for Sea Research (NIOZ). Her research interests are in the field of molecular geomicrobiology: Molecular evolution of lipid synthetic pathways, development of genetic tools to detect lipid biomarker-producing microorganisms, and microbial ecology of marine microorganisms especially those involved in the nitrogen cycle.

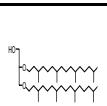
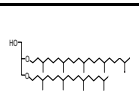
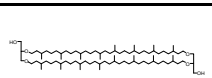
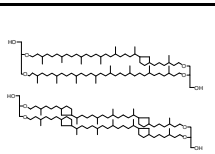
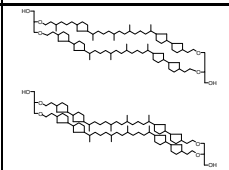
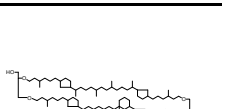
Prof. Dr. Ir. Jaap S. Sinninghe Damsté

Professor of Organic Geochemistry at Utrecht University (Faculty of Geosciences) and head of the Department of Marine Organic Biogeochemistry at NIOZ. He has been working for >25 years on the occurrence and characterization of lipids in microbes (bacteria, archaea, algae) and their application as biomarkers in oceanography, geochemistry, and palaeoclimatology.

Prof. Dr. Ir. Stefan Schouten

Professor in Molecular Palaeontology in Utrecht University (Faculty of Geosciences) and Senior Research Scientist at NIOZ. His research interest is the organic biogeochemistry of marine sediments, i.e. the reconstruction of present and past microbial communities, biosynthetic pathways, biogeochemical cycles, environments and climates by structural and stable isotopic analysis of organic compounds in microorganisms, marine waters and sediments.

Table 1. Distribution of archaeal membrane lipids in orders of the Euryarchaeota, Crenarchaeota and Thaumarchaeota phyla.

ARCHAEAL LIPIDS PHYLOGENY	Temperature	pH	Metabolism	Archaeol 	Ext archaeol 	GDGT-0 	GDGT-1 to 4 	GDGT-5 to 8 	Crenarchaeol 
Euryarchaeota									
Halobacteriales	M	N/Al	H	√	√				
Methanosarcinales	M	N	Met	√					
Methanopyrales	H	N	Met	√					
Methanococcales	M/T	N/Al	Met	√					
Thermococcales	T/H	N	S	√		√			
Methanobacteriales	M/T	N	Met	√		√			
Archaeoglobales	M/T	Al	S			√			
Methanomicrobiales	M	N	Met			√			
*Thermoplasmatales**	M/T	Ac	S			√	√	√	
Crenarchaeota									
Thermoproteales	T/H	N/Ac	S			√	√	√	
Sulfolobales	T/H	Ac	S			√	√	√	
Acidilobales	H	Ac	Org			√	√	√	
Desulfurococcales	H	N	S			√	√		
Thaumarchaeota									
Cenarchaeales						√	√		√
Nitrosopumilales						√	√		√
Nitrososphaerales						√	√		√

* DHVE-2 cluster (*Aciduliprofundum boonei*), closely related to the Thermoplasmatales order synthesize GDGT-0, GDGT-1/4 (REF 30). Temperature: M (Mesophile, 20–45°C); T (Thermophile, 45–80°C); H (Hyperthermophile, > 80°C). pH: N (Neutrophile, 5–8); Al (Alkalophile, >8); Ac (Acidophile, <5). Metabolism: H (Heterotrophy); Met (Methanogenesis); S (sulfur dependent); Nit (Nitrifier); Org (Organotroph). Archaeal membrane lipids distribution information from Schouten *et al.*[12].

Table S1. Isoprenyl diphosphate (IPP) synthases in archaeal genomes.

Phylum	Order	Genus, species	Short-chain IPP synthase‡	Amino acid†	Long-chain IPP synthase‡	Amino acid†
Crenarchaeota	Sulfolobales	*Sulfolobus acidocaldarius*	YP_254812.1	F	YP_255648.1	S
Crenarchaeota	Sulfolobales	*Sulfolobus solfataricus*	NP_341633.1	F	NP_343706.1	A
Crenarchaeota	Sulfolobales	*Sulfolobus tokodaii*	NP_378047.1	F	NP_376371.1	A
Crenarchaeota	Thermoproteales	*Pyrobaculum aerophilum*	NP_559016.1	Y	NP_560635.1	V
Crenarchaeota	Thermoproteales	*Thermoproteus neutrophilus*	YP_001793568.1	Y	YP_001794908	V
Crenarchaeota	Thermoproteales	*Pyrobaculum islandicum*	YP_930716	Y	YP_930070.1	V
Crenarchaeota	Thermoproteales	*Caldivirga maquilensis*	YP_001540467.1	Y	YP_001540335.1	A
Crenarchaeota	Desulfurococcales	*Aeropyrum pernix**	BAA88983.1	F		
Crenarchaeota	Desulfurococcales	*Desulfurococcus kamchatkensis*	YP_002429148.1	Y		
Crenarchaeota	Desulfurococcales	*Ignicoccus hospitalis*	YP_001435752.1	F	YP_001434928.1	S
Crenarchaeota	Desulfurococcales	*Staphylothermus marinus*	YP_001040510.1	Y		
Crenarchaeota	Acidilobales	*Acidilobus saccharovorans*	YP_003815770.1	F	YP_003815825	A
Euryarchaeota	Thermoplasmatales	*Thermoplasma volcanicum*	NP_110781.1	Y	NP_111576.1	A
Euryarchaeota	Thermoplasmatales	*Thermoplasma acidophilum*	NP_394768.1	Y	NP_393914	A
Euryarchaeota	Thermoplasmatales	*Ferroplasma acidarmanus*	ZP_05571075.1	F	ZP_05570405.1	A
Euryarchaeota	Thermoplasmatales	*Acididuliprofundum boonei*	ZP_04875656.1	Y	ZP_04875510.1	Y
Euryarchaeota	Thermococcales	*Thermococcus sp. AM4*	YP_002582296.1	Y	YP_002581574.1	A
Euryarchaeota	Thermococcales	*Pyrococcus horikoshii*	NP_142981.1	Y		
Euryarchaeota	Methanobacteriales	*Methanobrevibacter smithii*	ZP_05975848.2	F		
Euryarchaeota	Methanobacteriales	*Methanobacterium*	YP_004520238.1	F		
Euryarchaeota	Methanobacteriales	*Methanothermobacter marburgensis*	YP_003849447.1	F		
Euryarchaeota	Methanopyrales	*Methanopyrus kandleri*	NP_614058	F		
Euryarchaeota	Methanococcales	*Methanococcus maripadulis*	NP_987165.1	Y		

Euryarchaeota	Methanococcales	*Methanocaldococcus sp. FS406*	YP_003458715.1	Y		
Euryarchaeota	Methanosarcinales	*Methanosaeta thermophila*	YP_842903.1	F	YP_842784.1	A
Euryarchaeota	Methanosarcinales	*Methanosarcina barkeri*	YP_304956.1	F	YP_303957	A
Euryarchaeota	Methanosarcinales	*Methanosarcina mazei*	NP_633791.1	F	NP_632813	A
Euryarchaeota	Methanomicrobiales	*Methanospirillum hungatei*	YP_504297.1	F		
Euryarchaeota	Methanomicrobiales	*Methanoplanus limicola*	ZP_09700978.1	F		
Euryarchaeota	Archaeoglobales	*Archaeoglobus fulgidus*	AAD26851.1	F	NP_070380.1	A
Euryarchaeota	Archaeoglobales	*Archaeoglobus veneficus*	YP_004341338.1	F	YP_004340873	A
Euryarchaeota	Archaeoglobales	*Ferroglobus placidus*	YP_003435928.1	F	YP_003435725.1	A
Euryarchaeota	Halobacteriales	*Haladaptatus paucihalophilus*	ZP_08044024.1	F	ZP_08042560	A
Euryarchaeota	Halobacteriales	*Haloarcula hispanica*	YP_004795446.1	F	YP_004795026	A
Euryarchaeota	Halobacteriales	*Halomicrobium mukohataei*	YP_003178421.1	F	YP_003177096	A
Euryarchaeota	Halobacteriales	*Natronomonas pharaonis***	YP_327492.1	F	YP_325962	A
Euryarchaeota	Halobacteriales	*Halobacterium sp. NRC-1*	AAG19532.1	F	NP_280810	A
Thaumarchaeota	Cenarchaeales	*Cenarchaeum symbiosum*	YP_876540	F	YP_876597.1	E
Thaumarchaeota	Nitrosopumilales	*Ca. Nitrosoarchaeum limnia*	ZP_08257374	F	ZP_08257400.1	E
Thaumarchaeota	Nitrosopumilales	*Ca. Nitrosoarchaeum koreensis MY1*	ZP_08667311	F	ZP_08667284.1	E
Thaumarchaeota	Nitrosopumilales	*Ca. Nitrosopumilus salaria BD31*	ZP_10118543	F	ZP_10118596.1	E
Thaumarchaeota	Nitrosopumilales	*Nitrosopumilus maritimus*	YP_001581646	F	YP_001581621.1	E
Thaumarchaeota	Nitrososphaerales	*Ca. Nitrososphaera gargensis*	YP_006862746	F	YP_006861760	E

Short and Long-chain IPP synthases are defined in the text. ‡NCBI accession number. †Amino acid residue in the 5th position before the first aspartate-rich motif. *farnesylgeranyl diphosphate (FGPP) synthase of *Aeropyrum pernix* is involved in a pathway that only produces C_{25}–C_{25} diether lipids. It has been previously suggested that this FGPP synthase has evolved from an ancestral IPP synthase of Desulfurococcales (Tabichana *et al*., 2000). **bifunctional geranyl/farnesylgeranyl diphosphate synthase (C_{20} and C_{25}, respectively) described in the Halobacteriales *Natronomonas pharaonis* (Falb *et al*., 2005).

Table S2. Squalene/phytoene synthase homologues annotated in archaeal genomes.

Phylum	Order	Genus species	Accession number
Crenarchaeota	Sulfolobales	*Sulfolobus acidocaldarius*	YP_256333.1
Crenarchaeota	Sulfolobales	*Sulfolobus islandicus*	YP_002839942.1
Crenarchaeota	Sulfolobales	*Sulfolobus solfataricus*	NP_344224.1
Crenarchaeota	Sulfolobales	*Metallosphaera yellowstonensis*	WP_009069731.1
Crenarchaeota	Sulfolobales	*Metallophaera cuprina*	YP_004409628.1
Crenarchaeota	Sulfolobales	*Metallophaera sedula*	YP_001191163.1
Crenarchaeota	Thermoproteales	*Pyrobaculum oguniense*	YP_005260591.1
Crenarchaeota	Thermoproteales	*Pyrobaculum arsenaticum*	YP_001152662.1
Euryarchaeota	Thermoplasmatales	*Picrophilus torridus*	AAT_44120.1
Euryarchaeota	Methanobacteriales	*Methanobacterium sp.*	WP_008515272.1
Euryarchaeota	Methanobacteriales	*Methanothermobacter thermoautotrophicus*	NP_276914.1
Euryarchaeota	Methanomicrobiales	*Methanoculleus marisnigri*	YP_001046034.1
Euryarchaeota	Methanosarcinales	*Methanosalsum zhilinae*	YP_004616025.1
Euryarchaeota	Halobacteriales†	*Natrialba madadii*	YP_003482007.1
Euryarchaeota	Halobacteriales	*Natronobacterium gregoryi*	YP_007177834.1
Euryarchaeota	Halobacteriales	*Halobacterium sp.*	NP_280284.1
Euryarchaeota	Halobacteriales	*Haloferax prahovense*	WP_0080095376.1
Euryarchaeota	Halobacteriales	*Haloquadratum walsbyi*	YP_658569.1
Euryarchaeota	Halobacteriales	*Haloarcula marismortui*	YP_136629.1

Phytoene/squalene synthase are defined as tail-to-tail isoprenyl diphosphate synthases. Squalene and phytoene synthases catalyze the condensation of two C_{15} (farnesyl) and C_{20} (geranylgeranyl) isoprenyl diphosphates, respectively. †Halobacteriales: phytoene/squalene synthases are commonly found in members of the Halobacteriales order and here we just list some of them. They are believed to be involved in the formation of rhodopsins formed by halophilic Archaea (Peck *et al*., 2002).

References Supplementary Tables:

Falb, M., Pfeiffer, F., Palm, P., Rodewald, K., Hickmann, V., Tittor, J. & Oesterhelt, D. Living with two extremes: conclusions from the genome sequence of *Natronomonas pharaonis*. *Genome Res*. **15**, 1336–1343 (2005).

Peck, R.F., Johnson, E.A., & Krebs, M.P. Identification of a lycopene beta-cyclase required for bacteriorhodopsin biogenesis in the archaeon Halobacterium salinarum. *J. Bacteriol.* **184**, 2889–2897 (2002).

Tachibana, A., Yano, Y., Otani, S., Nomura, N., Sako, Y. & Taniguchi, M. Novel prenyltransferase gene encoding farnesylgeranyl diphosphate synthase from a hyperthermophilic archaeon, *Aeropyrum pernix*. Molecular evolution with alteration in product specificity. *Eur. J. Biochem.* **267**, 321–328 (2000).

Figure 1.

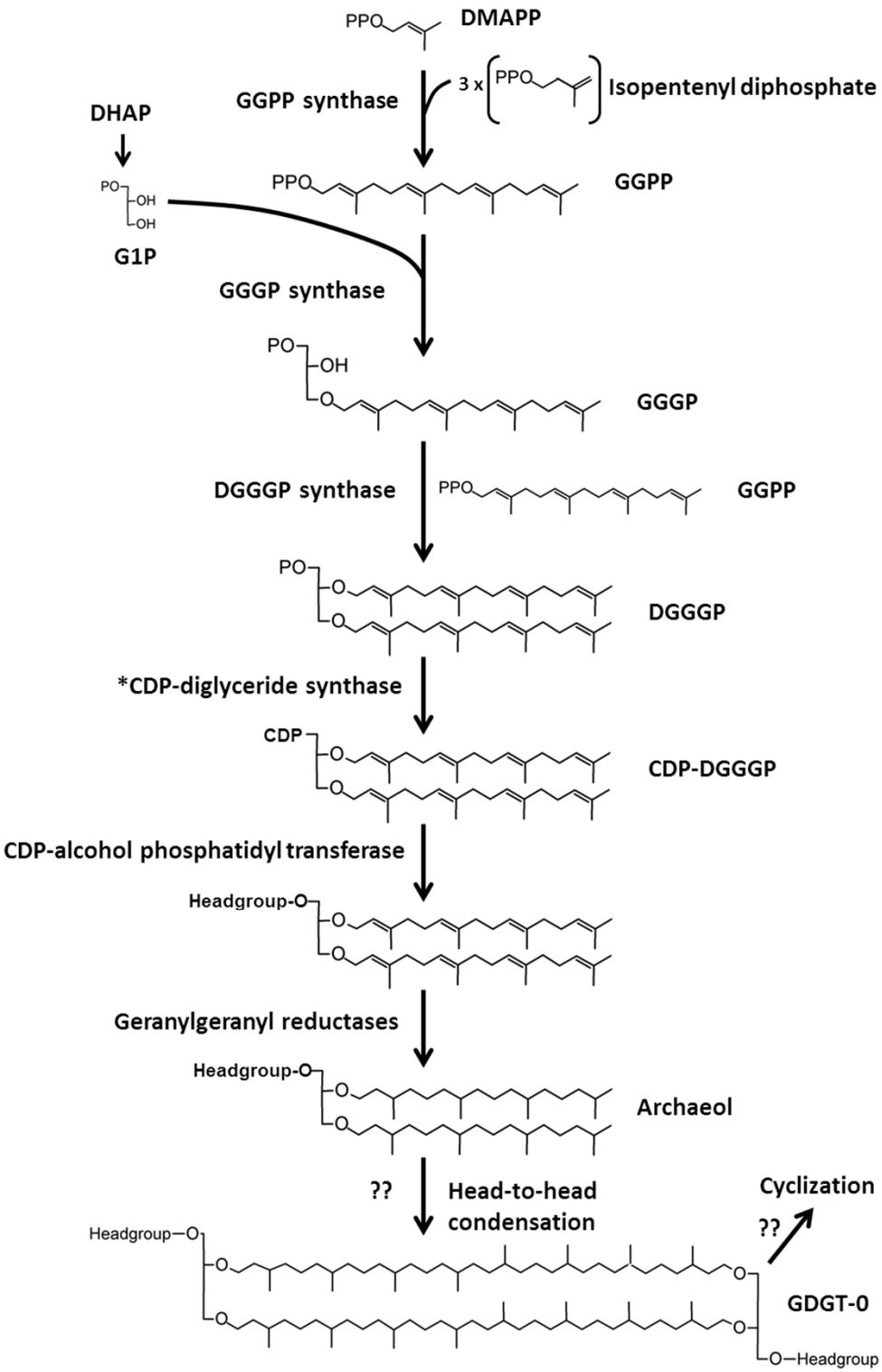

DMAPP
GGPP synthase
3 x
Isopentenyl diphosphate
DHAP
GGPP
G1P
GGGP synthase
GGGP
DGGGP synthase
GGPP
DGGGP
*CDP-diglyceride synthase
CDP-DGGGP
CDP-alcohol phosphatidyl transferase
Geranylgeranyl reductases
Archaeol
??
Head-to-head condensation
Cyclization
??
GDGT-0

Figure 2

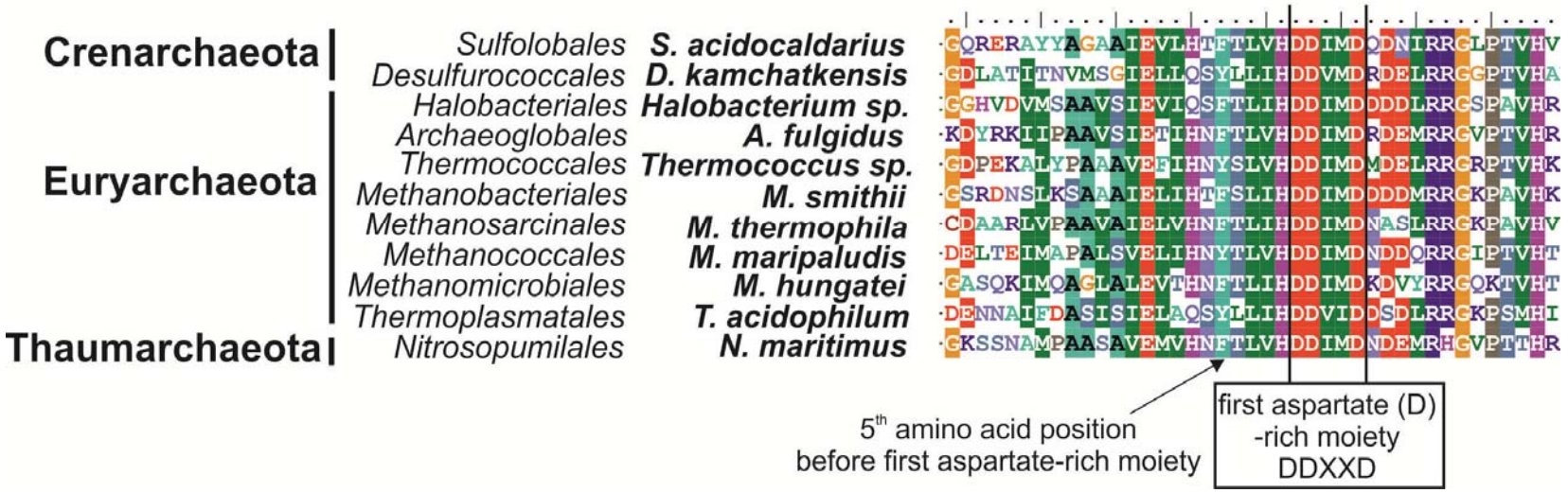

Crenarchaeota
Euryarchaeota
Thaumarchaeota
Sulfolobales S. acidocaldarius
Desulfurococcales D. kamchatkensis
Halobacteriales Halobacterium sp.
Archaeoglobales A. fulgidus
Thermococcales Thermococcus sp.
Methanobacteriales M. smithii
Methanosarcinales M. thermophila
Methanococcales M. maripaludis
Methanomicrobiales M. hungatei
Thermoplasmatales T. acidophilum
Nitrosopumilales N. maritimus
5th amino acid position
before first aspartate-rich moiety
first aspartate (D)
-rich moiety
DDXXD

Figure 3

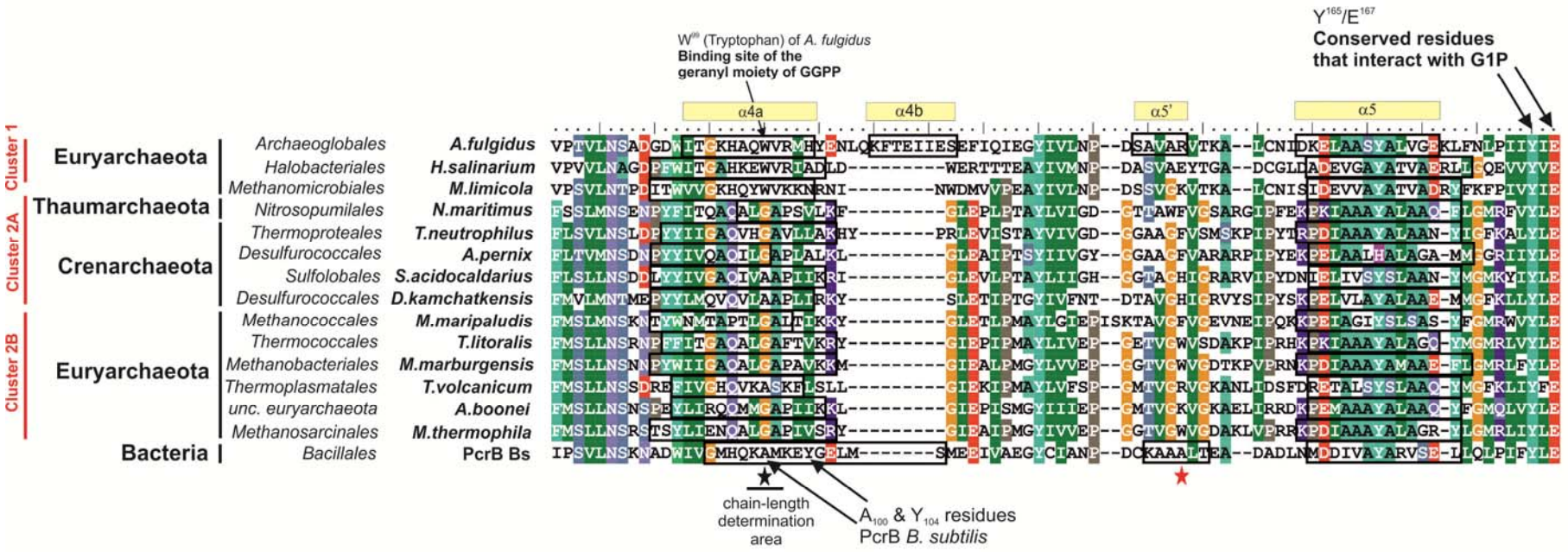

W[89] (Tryptophan) of A. fulgidus
Binding site of the geranyl moiety of GGPP
Y[165]/E[167]
Conserved residues that interact with G1P
α4a
α4b
α5'
α5
Cluster 1
Cluster 2A
Cluster 2B
Euryarchaeota
Thaumarchaeota
Crenarchaeota
Euryarchaeota
Bacteria
Archaeoglobales
Halobacteriales
Methanomicrobiales
Nitrosopumilales
Thermoproteales
Desulfurococcales
Sulfolobales
Desulfurococcales
Methanococcales
Thermococcales
Methanobacteriales
Thermoplasmatales
unc. euryarchaeota
Methanosarcinales
Bacillales
A.fulgidus
H.salinarium
M.limicola
N.maritimus
T.neutrophilus
A.pernix
S.acidocaldarius
D.kamchatkensis
M.maripaludis
T.litoralis
M.marburgensis
T.volcanicum
A.boonei
M.thermophila
PcrB Bs
chain-length determination area
A100 & Y104 residues PcrB B. subtilis

Figure 4

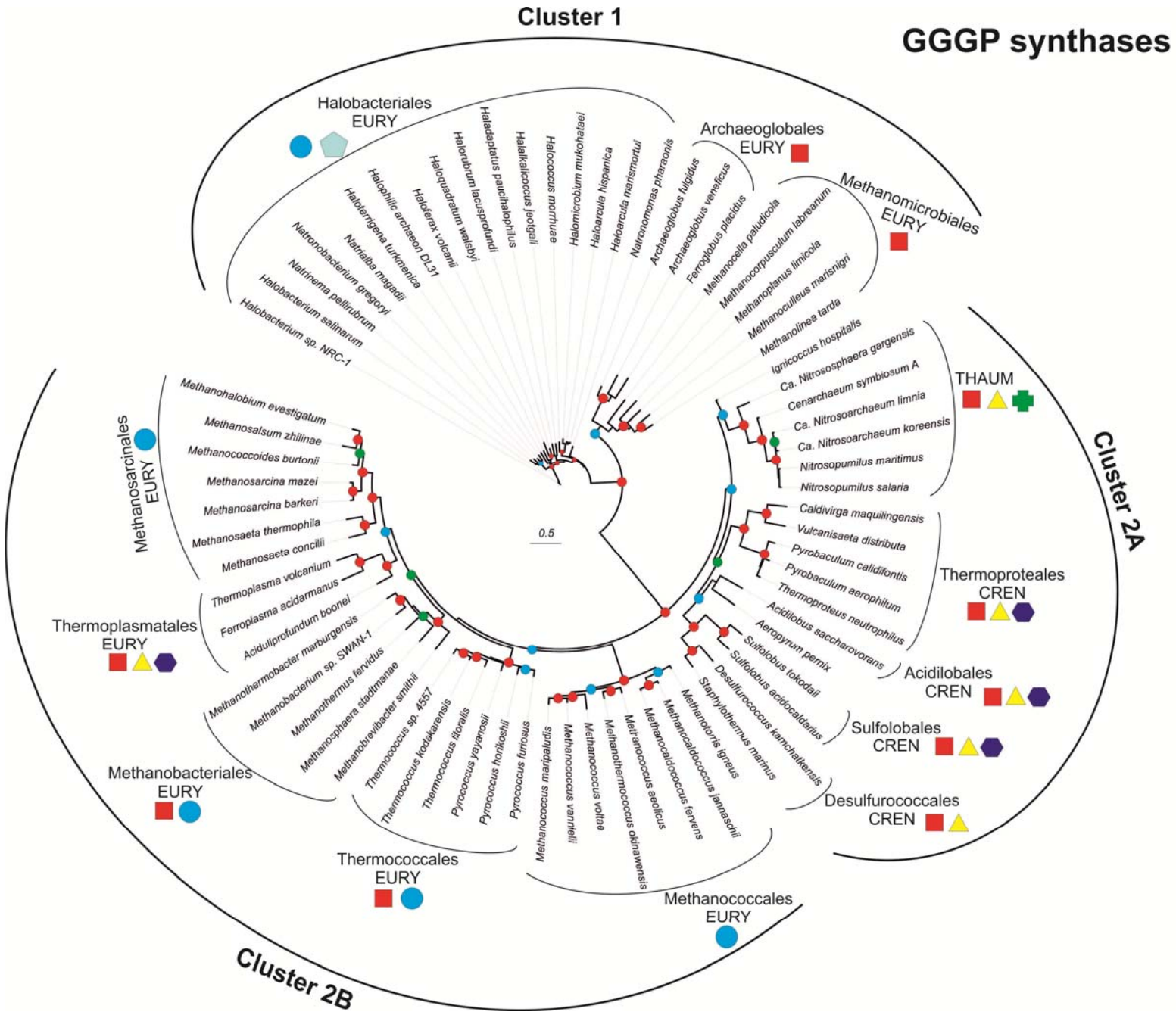

GGGP synthases
Cluster 1
Cluster 2A
Cluster 2B
Halobacteriales
EURY
Archaeoglobales
EURY
Methanomicrobiales
EURY
Halobacterium sp. NRC-1
Halobacterium salinarum
Natrinema pellirubrum
Natronobacterium gregoryi
Natrialba magadii
Haloterrigena turkmenica
Halophilic archaeon DL31
Haloferax volcanii
Haloquadratum walsbyi
Halorubrum lacusprofundi
Haladaptatus paucihalophilus
Halalkalicoccus jeotgali
Halococcus morrhuae
Halomicrobium mukohataei
Haloarcula hispanica
Haloarcula marismortui
Natronomonas pharaonis
Archaeoglobus fulgidus
Archaeoglobus veneficus
Ferroglobus placidus
Methanocella paludicola
Methanocorpusculum labreanum
Methanoplanus limicola
Methanoculleus marisnigri
Methanolinea tarda
Ignicoccus hospitalis
Ca. Nitrososphaera gargensis
Cenarchaeum symbiosum A
Ca. Nitrosoarchaeum limnia
Ca. Nitrosoarchaeum koreensis
Nitrosopumilus maritimus
Nitrosopumilus salaria
THAUM
Caldivirga maquilingensis
Vulcanisaeta distributa
Pyrobaculum calidifontis
Pyrobaculum aerophilum
Thermoproteus neutrophilus
Acidilobus saccharovorans
Thermoproteales
CREN
Aeropyrum pernix
Sulfolobus tokodaii
Sulfolobus acidocaldarius
Desulfurococcus kamchatkensis
Staphylothermus marinus
Methanotorris igneus
Methanocaldococcus jannaschii
Methanocaldococcus fervens
Methanococcus aeolicus
Methanothermococcus okinawensis
Methanococcus voltae
Methanococcus vannielii
Methanococcus maripaludis
Pyrococcus furiosus
Pyrococcus horikoshii
Pyrococcus yayanosii
Thermococcus litoralis
Thermococcus kodakarensis
Thermococcus sp. 4557
Methanobrevibacter smithii
Methanosphaera stadtmanae
Methanothermus fervidus
Methanothermobacter marburgensis
Methanobacterium sp. SWAN-1
Acidiuliprofundum boonei
Ferroplasma acidarmanus
Thermoplasma volcanium
Methanosaeta concilii
Methanosaeta thermophila
Methanosarcina barkeri
Methanosarcina mazei
Methanococcoides burtonii
Methanosalsum zhilinae
Methanohalobium evestigatum
Methanosarcinales
EURY
Thermoplasmatales
EURY
Methanobacteriales
EURY
Thermococcales
EURY
Methanococcales
EURY
Acidilobales
CREN
Sulfolobales
CREN
Desulfurococcales
CREN
0.5

Figure 5

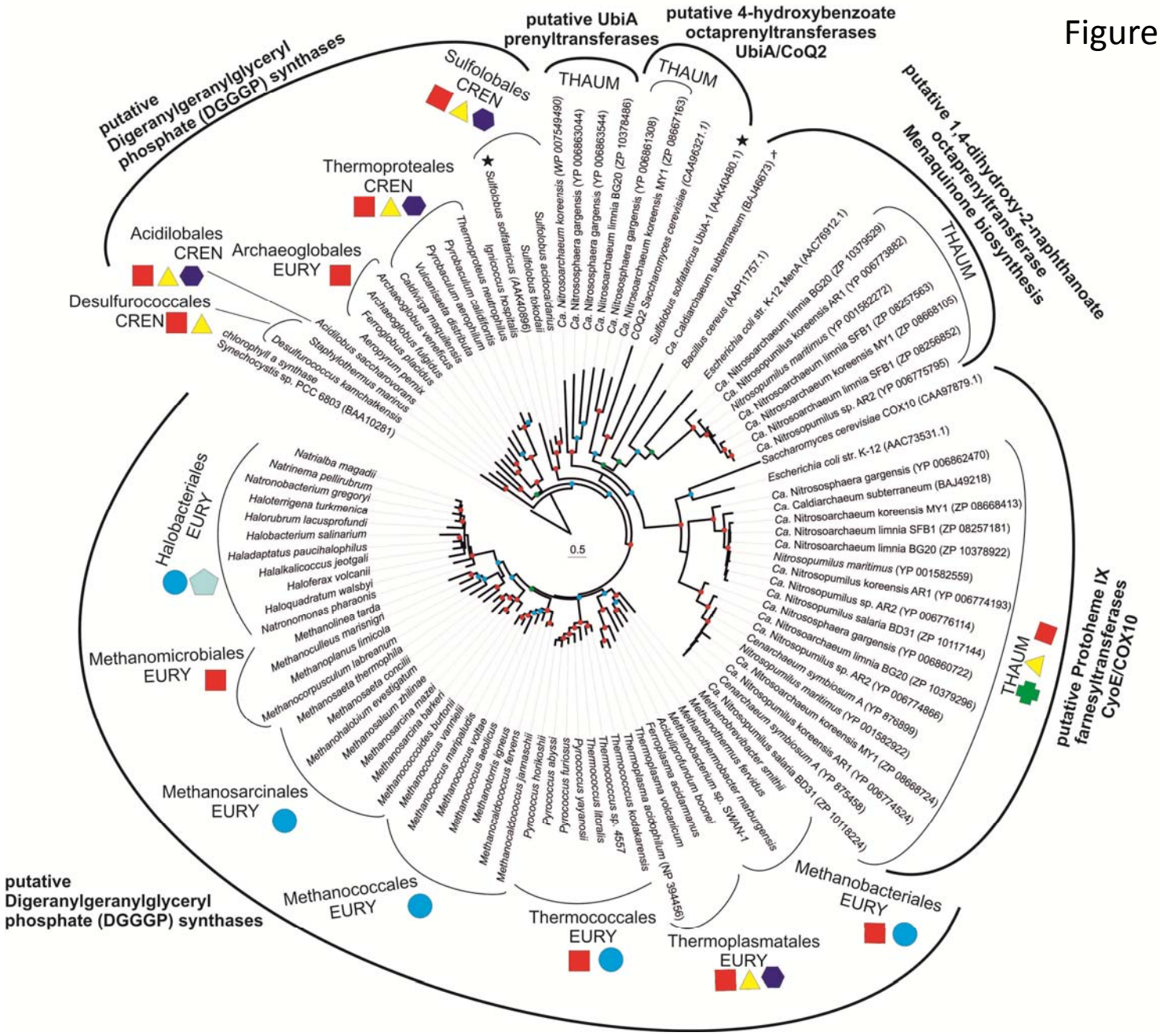
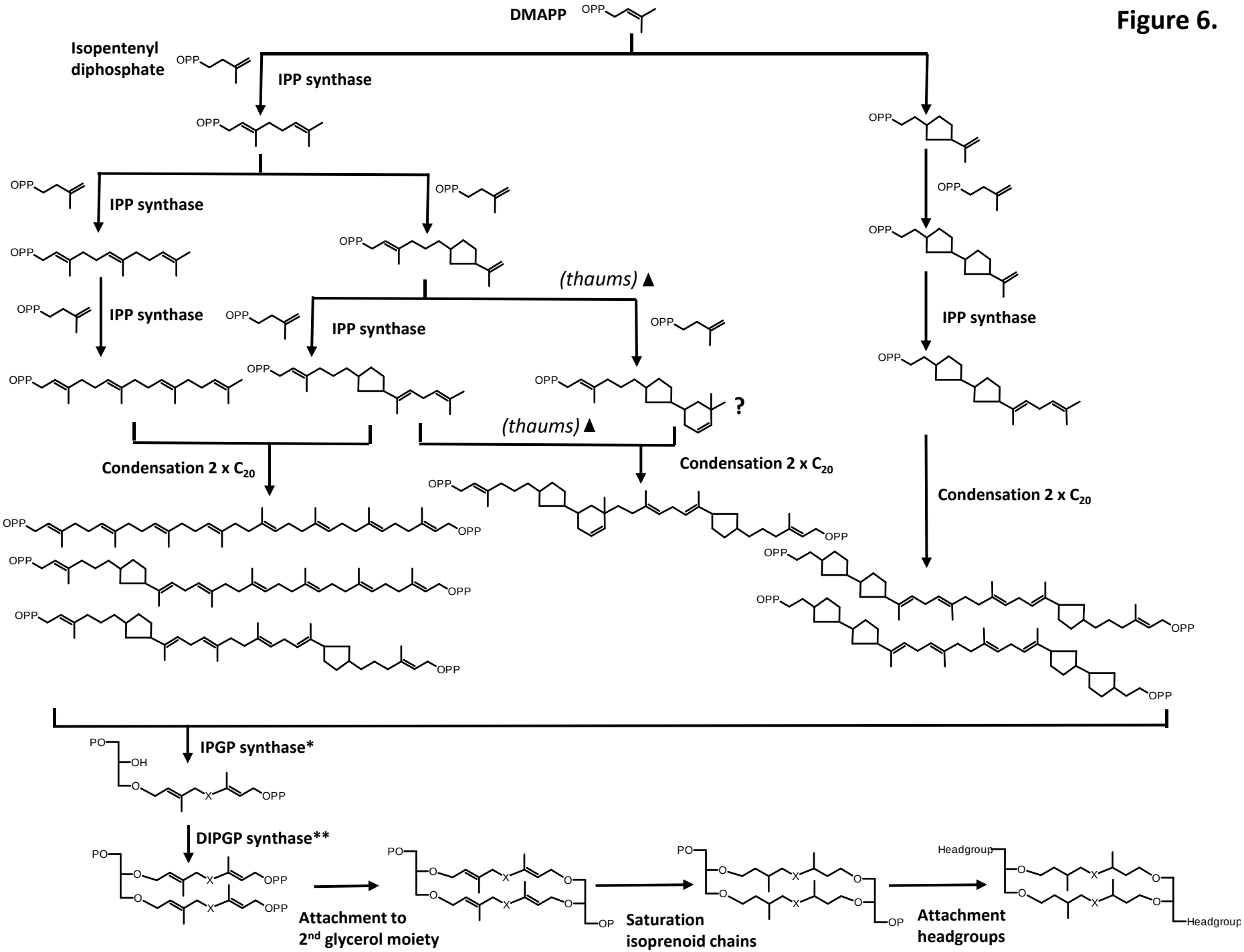

putative Digeranylgeranylglyceryl phosphate (DGGGP) synthases
putative UbiA prenyltransferases
putative 4-hydroxybenzoate octaprenyltransferases UbiA/CoQ2
putative 1,4-dihydroxy-2-naphthanoate octaprenyltransferase Menaquinone biosynthesis
putative Protoheme IX farnesyltransferases CyoE/COX10
Sulfolobales CREN
Thermoproteales CREN
Acidilobales CREN
Desulfurococcales CREN
Archaeoglobales EURY
Halobacteriales EURY
Methanomicrobiales EURY
Methanosarcinales EURY
Methanococcales EURY
Thermococcales EURY
Thermoplasmatales EURY
Methanobacteriales EURY
THAUM
0.5

Figure 6.


DMAPP
Isopentenyl diphosphate
IPP synthase
IPP synthase
IPP synthase
IPP synthase
IPP synthase
(thaums) ▲
(thaums) ▲
?
Condensation 2 x C20
Condensation 2 x C20
Condensation 2 x C20
IPGP synthase*
DIPGP synthase**
Attachment to 2nd glycerol moiety
Saturation isoprenoid chains
Attachment headgroups
Headgroup
Headgroup