On the Origin of Isoprenoid Biosynthesis

Yosuke Hoshino*,[1] and Eric A. Gaucher*,[1,2,3]

[1]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA
[2]School of Chemistry and Biochemistry, and Parker H. Petit Institute for Bioengineering and Biosciences, Georgia Institute of Technology, Atlanta, GA
[3]Department of Biology, Georgia State University, Atlanta, GA

*Corresponding authors: E-mails: yosuke.hoshino@biosci.gatech.edu; egaucher@gsu.edu.
Associate editor: James McInerney

Abstract

Isoprenoids and their derivatives represent the largest group of organic compounds in nature and are distributed universally in the three domains of life. Isoprenoids are biosynthesized from isoprenyl diphosphate units, generated by two distinctive biosynthetic pathways: mevalonate pathway and methylerthritol 4-phosphate pathway. Archaea and eukaryotes exclusively have the former pathway, while most bacteria have the latter. Some bacteria, however, are known to possess the mevalonate pathway genes. Understanding the evolutionary history of these two isoprenoid biosynthesis pathways in each domain of life is critical since isoprenoids are so interweaved in the architecture of life that they would have had indispensable roles in the early evolution of life. Our study provides a detailed phylogenetic analysis of enzymes involved in the mevalonate pathway and sheds new light on its evolutionary history. The results suggest that a potential mevalonate pathway is present in the recently discovered superphylum Candidate Phyla Radiation (CPR), and further suggest a strong evolutionary relationship exists between archaea and CPR. Interestingly, CPR harbors the characteristics of both the bacterial-type and archaeal-type mevalonate pathways and may retain signatures regarding the ancestral isoprenoid biosynthesis pathway in the last universal common ancestor. Our study supports the ancient origin of the mevalonate pathway in the three domains of life as previously inferred, but concludes that the evolution of the mevalonate pathway was more complex.

Key words: isoprenoid, mevalonate pathway, candidate phyla radiation, GHMP superfamily.

Introduction

Isoprenoids are the largest family of organic compounds in nature, encompassing over 65,000 compounds (Buckingham 2007) and have diverse roles in all organisms from single-celled bacteria and archaea to multicellular eukaryotes. Examples include quinones in electron transport chains, lipids in cell membranes, cell-signaling agents, hormones, and defense compounds (Gershenzon and Dudareva 2007; Crowell and Huizinga 2009; Santner et al. 2009; Nowicka and Kruk 2010; Jain et al. 2014). The biosynthesis of any isoprenoid starts from successive condensations of only two building blocks: isopentenyl diphosphate (IPP) and its isomer dimethyallyl diphosphate (DMAPP). IPP and DMAPP are in turn derived via two distinctive pathways: the mevalonate (MVA) pathway and the methylerythritol 4-phosphate (MEP) pathway (Pérez-Gil and Rodríguez-Concepción 2013) (fig. 1 for MVA pathway and supplementary fig. 1 for MEP pathway, Supplementary Material online). Archaea and eukaryotes are known to possess the MVA pathway. In contrast, most bacteria have the MEP pathway, although some bacteria do harbor the MVA pathway. Photosynthetic eukaryotes are known to have both the MVA and the MEP pathways, and the MEP pathway is inferred to have originated from an ancestral MEP pathway within a symbiotic cyanobacteria that interacted with early eukaryotes (Lichtenthaler 1999).

The known MVA pathways are composed of seven or eight enzymes that combine three molecules of acetyl-CoA and eventually generate IPP (fig. 1). The first enzyme, acetoacetyl-CoA thiolase (AACT), is not specific to the biosynthesis of isoprenoid, but the other six or seven enzymes are dedicated to the MVA pathway. The MVA pathway is well-conserved in the domain Eukaryotes. The MVA pathway in bacteria is mostly homologous to the eukaryotic MVA pathway, while only Chloroflexi harbors an archaeal-type MVA pathway. The MVA pathways discovered in archaea so far have three variations (fig. 1). Sulfolobales has the eukaryotic-type MVA pathway (Nishimura et al. 2013). In contrast, halobacteria (a known archaeal group) have a modified version of the MVA pathway. In eukaryotes, the reaction steps from mevalonate 5-phosphate to IPP proceeds in the order of phosphorylation and decarboxylation catalyzed by phosphomevalonate kinase (PMVK) and diphosphomevalonate decarboxylase (DPMD), respectively (DPMD route; fig. 1). In contrast, halobacterial MVA pathway proceeds via decarboxylation and then phosphorylation catalyzed by mevalonate phosphate decarboxylase (MPD) and by isopentenyl
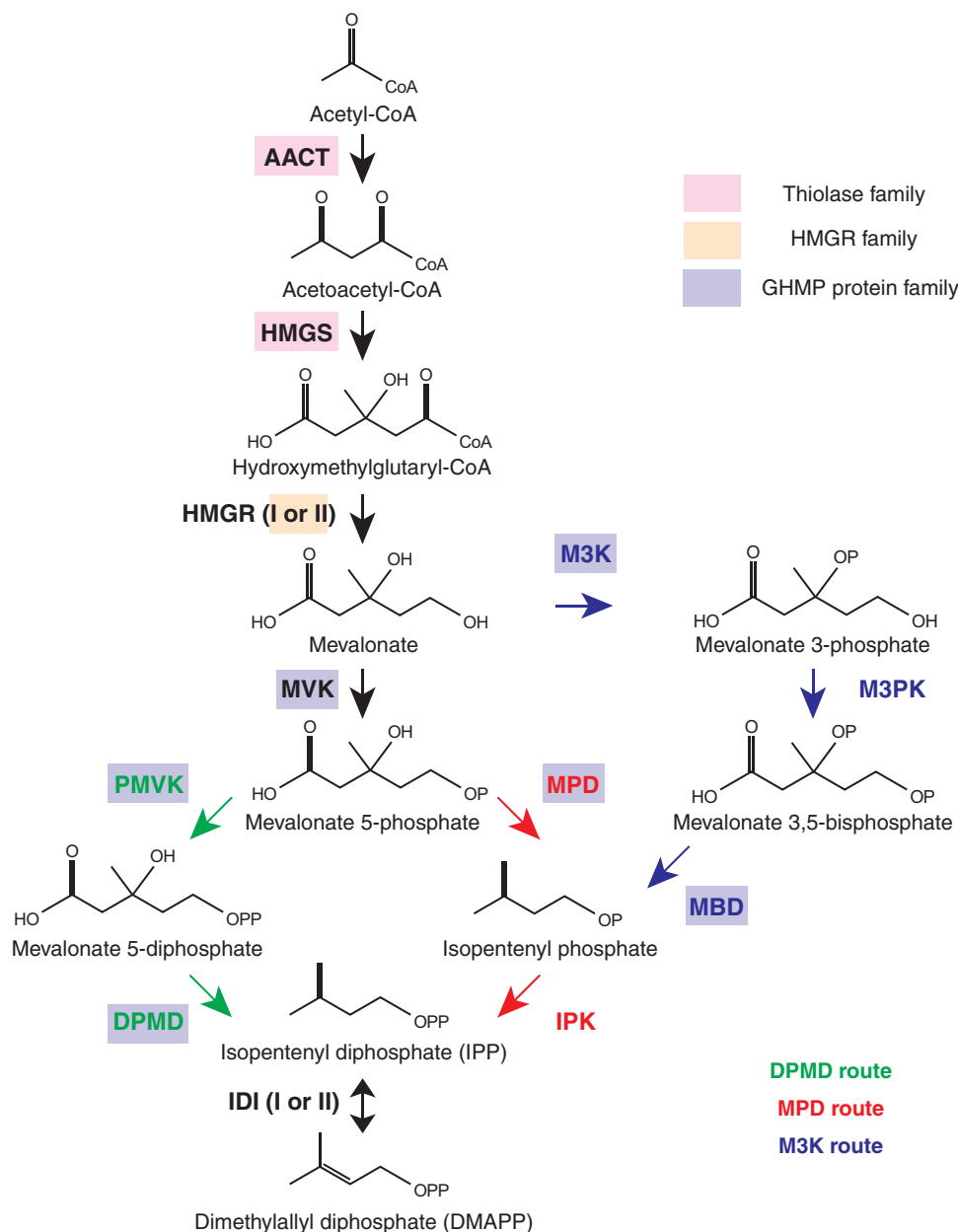
Open Access

Article

Acetyl-CoA

AACT

Acetoacetyl-CoA

HMGS

Hydroxymethylglutaryl-CoA

HMGR (I or II)

Mevalonate

M3K

Mevalonate 3-phosphate

M3PK

MVK

Mevalonate 5-phosphate

PMVK   **MPD**

Mevalonate 3,5-bisphosphate

Mevalonate 5-diphosphate

Isopentenyl phosphate

MBD

DPMD   **IPK**

Isopentenyl diphosphate (IPP)

IDI (I or II)

Dimethylallyl diphosphate (DMAPP)

Thiolase family
HMGR family
GHMP protein family

DPMD route
MPD route
M3K route

Fig. 1. Three variations of the MVA pathway discovered in the three domains of life - DPMD, MPD, and M3K routes. The enzymes with identically colored boxes indicate that they are evolutionarily related. Abbreviations for enzymes: AACT, acetoacetyl-CoA thiolase; HMGS, 3-hydroxy-3-methylglutaryl-CoA synthase; HMGR, 3-hydroxy-3-methylglutaryl-CoA reductase; MVK, mevalonate kinase; PMVK, phosphomevalonate kinase; DPMD, Diphosphomevalonate decarboxylase; IDI, IPP/DMAPP isomerase; MPD, mevalonate phosphate decarboxylase; IPK, isopentenyl phosphate kinase; M3K, mevalonate 3-kinase; M3PK, mevalonate 3-phosphate 5-kinase; MBD, mevalonate 3,5-bisphosphate decarboxylase.

phosphate kinase (IPK), respectively (MPD route). Interestingly, the MPD and IPK steps in halobacteria are in the reverse reaction order as the PMVK and DPMD steps from eukaryotes. The MPD route is currently only confirmed in halobacteria and chloroflexi (Dellas et al. 2013; VanNice et al. 2014). Chloroflexi are the only organisms known to possess an archaeal-type MVA pathway outside the domain Archaea. Furthermore, another modified MVA pathway has been discovered in the order Thermoplasmata (Azami et al. 2014; Vinokur et al. 2014, 2016; Rossoni et al. 2015). Species in this order do not use any of MVK, PMVK, and DPMD, but instead utilize three novel enzymes. Mevalonate 3-kinase (M3K) and mevalonate 3-phosphate kinase (M3PK) catalyze

stepwise mevalonate phosphorylations and then mevalonate 3,5-bisphosphate decarboxylase (MBD) produces isopentenyl phosphate (M3K route; fig. 1). Isopentenyl phosphate is transformed into IPP by IPK as is the case for the MPD route in halobacteria. Outside halobacteria, thermoplasmata and sulfolobales, no homologs of PMVK, DPMD, MPD, M3K, M3PK, and MBD have been discovered in archaea so far, even though most archaea are still inferred to have a similar MVA pathway because they universally possess other necessary enzymes (AACT, HMGS, HMGR, MVK, and IPK).

The third step and the last step of the MVA pathway are each composed of two different enzymes (Type-I and Type-II HMGR and IDI; fig. 1). Both types are functionally equivalent

and they display a complex distribution among organisms depending on genus or even species. In general, Type-I HMGR (HMGR-I) is found more in archaea and eukaryotes, whereas Type-II HMGR (HMGR-II) is typical of bacteria. In addition, Type-I IDI (IDI-I) is common in bacteria and eukaryotes, while Type-II IDI (IDI-II) is found in bacteria and archaea (Lombard and Moreira 2011).

Interestingly, three enzymes of the bacterial/eukaryotic MVA pathway (MVK, PMVK, and DPMD) and three enzymes in the archaeal MVA pathway (MPD, M3K, and MBD) are all distantly related to each other and belong to the GHMP protein superfamily (fig. 1) (Romanowski et al. 2002; Yang et al. 2002; Krepkiy and Miziorko 2004; Dellas et al. 2013; Azami et al. 2014; Vinokur et al. 2016). Furthermore, 4-(cytidine 5′-diphospho)-2-C-methylerythritol kinase (CMK) in the MEP pathway in bacteria is also included in the GHMP superfamily (supplementary fig. 1, Supplementary Material online) (Miallau et al. 2003). Therefore, it is implied that these GHMP family proteins (MVK, PMVK, DPMD, MPD, M3K, MBD, and CMK) involved in the isoprenoid biosynthesis share a common ancestor. Similarly, the first two enzymes of the MVA pathway (AACT and HMGS) are evolutionary related to each other (thiolase family) (Jiang et al. 2008) and two types of HMGR (I and II) are also distant homologs to each other (fig. 1).

A previous phylogenetic study argued that the MVA pathway in the three domains of life was vertically inherited from the last universal common ancestor (LUCA), respectively, but in most bacteria the MVA pathway was later replaced with the MEP pathway (Lombard and Moreira 2011). This inference was based on the presence of three distinct clades seen in phylogenies inferred for each individual enzyme of the MVA pathway. However, the punctate distribution of the MVA pathway in the bacterial domain renders this argument inconclusive.

In our current study, the distribution of the MVA pathway from the three domains of life is re-examined using currently available sequences from public databases and an integrated approach toward phylogenetics (Gaucher et al. 2002, 2004; West et al. 2002). There is no comprehensive phylogenetic study that discusses newly discovered MVA enzymes (MPD, M3K, MBD) in detail. In addition, the potential ability to produce isoprenoids in a recently discovered superphylum called Candidate Phyla Radiation (CPR) has been suggested and the evolutionary link between CPR to bacteria and archaea has garnered attention (Brown et al. 2015; Castelle and Banfield 2018).

Our study suggests that CPR possesses a potential MVA pathway that is homologous to the bacterial/eukaryotic MVA pathway. CPR is possibly a basal member of bacteria and our observation supports the presence of the MVA pathway in the common ancestor of bacteria and CPR. However, CPR is also found to possess additional MVA enzymes that are otherwise only distributed in the archaeal domain and dedicated to the archaeal MVA pathway. MVA enzymes in CPR always cluster close to archaeal sequences and are thus inferred to share a common ancestor. CPR bears the characteristics of both the bacterial/eukaryotic pathway and the archaeal MVA pathway, and provides unique insights into the evolutionary relationship of these two MVA pathways.

Isoprenoids are ubiquitous components of cell membranes in all three domains of life without exception and the evolution of isoprenoids could potentially be as old as the origin of the cell membrane itself. A more precise understanding of isoprenoid biosynthesis evolution could provide beneficial information about the composition of ancestral cell membranes at or even before the era of LUCA, which ultimately could give us clues to the origin of cellular life.

Results and Discussion

Distribution of MVA Pathway in Bacteria, Archaea, and CPR

The distribution of 13 MVA enzymes (AACT, HMGS, HMGR-I/II, MVK, PMVK, DPMD, MPD, M3K, MBD, IPK, and IDI-I/II) was examined for bacteria, archaea and the newly discovered superphylum CPR (fig. 2) (see Materials and Methods for the details of data selection). Although CPR is considered a basal member of bacteria, it has distinct characteristics from other bacteria and is separately discussed in our current study.

Bacteria

Bacteria that have three or more enzymes of the MVA pathway are found in 20 out of 62 bacterial phyla (supplementary fig. 2 for the detailed distribution pattern; Supplementary Material online). Out of these 20, 11 have only a limited number of species that possess MVA enzymes, suggesting a sparse distribution. Thus, genes encoding MVA enzymes (MVA genes hereafter) in these bacterial phyla are likely to be derived by horizontal gene transfer (HGT) from outside these phyla. Besides the inferred HGT-derived MVA genes, most bacterial MVA genes are found in Spirochaetes, δ-Proteobacteria, γ-Proteobacteria, Firmicutes, Actinobacteria, Bacteroidetes, and Chloroflexi. Even for these phyla, MVA gene-carrying species never occupy a majority of the phylum but rather contain the MEP pathway in a ubiquitous manner.

Apart from Chloroflexi and Bacteroidetes, bacterial phyla broadly share a common tree topology and thus are inferred to have a common evolutionary history (figs. 3 and 4). The close evolutionary relationship of MVA genes among bacteria is also implied by the synteny of MVA genes. Most bacterial species, except for chloroflexi, form a gene cluster that includes most MVA genes (fig. 2). Hence, MVA genes are inferred to have been transmitted in bacteria as a syntenic set. The observed tree topologies of MVA genes in the bacterial clade, however, only partially match recently published species trees (e.g., monophyly of Firmicutes and Actinobacteria; figs. 3 and 4 and supplementary figs. 7 and 12, Supplementary Material online) (Toft and Andersson 2010; Raymann et al. 2015; Hug et al. 2016). In most MVA trees, Terrabacteria (Firmicutes and Actinobacteria) is the most derived clade, which conversely is the earliest-
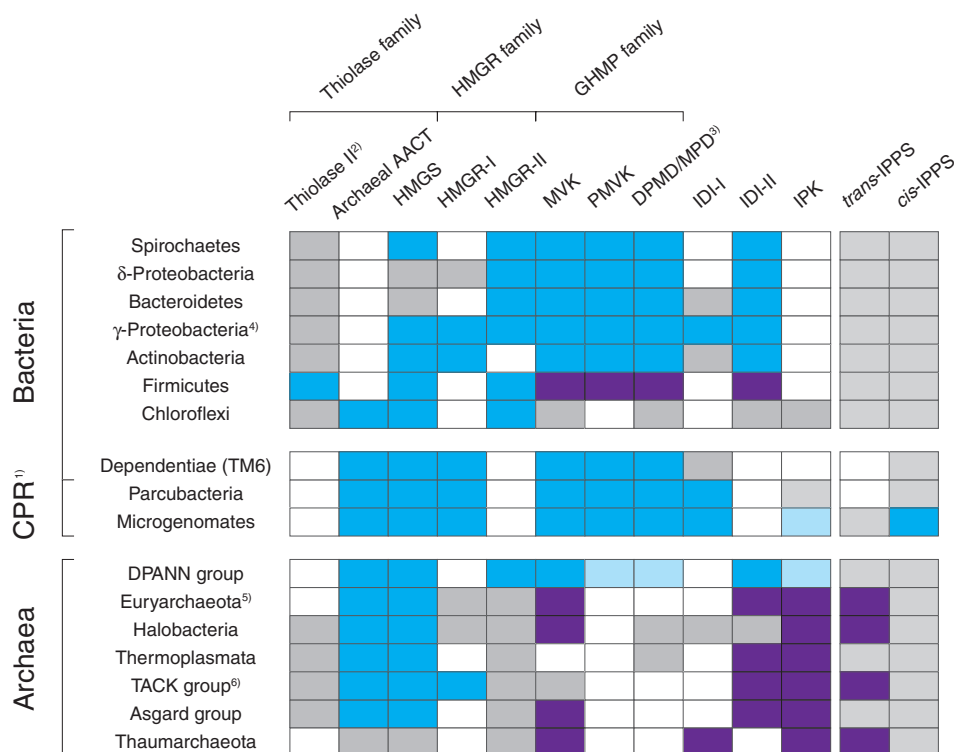
Fig. 2. Distribution of MVA genes and their syntenic relationship in bacteria, CPR and archaea. The colored box (blue or purple) indicates that the majority of MVA gene-carrying species in a phylum (or a superphylum) form a gene cluster in their genomes. Two different colors (blue and purple) within a single row indicate that MVA genes form two separate gene clusters in a genome. Grey box indicates the presence of MVA genes but they are not syntenic with other MVA genes. Light blue color indicates the presence of a gene cluster but there are variations among species (see the main text). Unfilled box indicates the absence of MVA genes in a phylum, or only a very small number of species carry the gene. It should be noted that only MVA gene-carrying species are included in the dataset and not all species in each phylum possess the MVA genes. Particularly in all bacteria and CPR parcubacteria, MVA gene-carrying species never occupy a majority in any phylum (see the main text in detail). Abbreviation: IPPS, isoprenyl diphosphate synthase. See fig. 1 for other abbreviations. Notes: 1) CPR is separately treated from the traditional bacterial domain, 2) Thiolase II is AACT used by bacteria and eukaryotes, 3) DPMD/MPD column encompasses M3K and MBD homologs, 4) Presence of HMGR-I/II and of IDI-I/II depends on species, 5) Halobacteria and Thermoplasmata are shown separately, 6) Thaumarchaeota is shown separately.

branching clade in all species trees. Additionally, the distribution of MVA genes in bacteria is mostly limited to the most-derived class/order in the given phylum. Even though the possibility of vertical transmittance of MVA genes in the bacterial domain is not excluded, the very limited distribution of MVA genes in discrete bacterial taxa more strongly suggests HGT as a major driving force for the dissemination of MVA genes in bacteria. Nearly each bacterial phylum forms its own separate monophyletic clade in all MVA trees, thus such HGT events would have been relatively early in the bacterial evolution. Likewise, a horizontal origin of eukaryotic MVA genes is suggested since the eukaryotic clade mostly clusters with bacteria and CPR (figs. 3 and 4 and supplementary figs. 7 and 11, Supplementary Material online), except for the unresolved HMGR-I tree (supplementary fig. 6, Supplementary Material online). This is not consistent with the general consensus that eukaryotes are more closely related to archaea (Eme et al. 2017). In addition, since only a small number of α-proteobacteria (none from the closest relatives of mitochondria, Rickettsiales

and Pelagibacterales) possess the MVA pathway, a mitochondrial origin of the eukaryotic MVA pathway is unlikely. Therefore, MVA genes in eukaryotes are inferred to be horizontally derived from the bacterial domain, contrary to the previous inference that the eukaryotic MVA pathway originated in LUCA (Lombard and Moreira 2011).

Compared with other bacteria, the phylogenetic position of Bacteroidetes is not consistent for each MVA enzyme tree. This may be due to its high divergence rate relative to other bacteria (figs. 3 and 4) and reflected in the fact that some MVA homologs had not been discovered in this phylum until very recently (MVK and PMVK; Hayakawa et al. 2017). Similarly, there is a group of bacteria carrying MVA enzymes that occupy varying phylogenetic positions (asterisks; figs. 3 and 4 and supplementary figs. 6 and 12, Supplementary Material online). Bacteria in this group always form a single clade, but are composed of several discrete phyla (e.g., γ-proteobacteria, Verrucomicrobia and Lentisphaerae). Hence, MVA genes in this group are inferred to have a single origin, but have been horizontally transferred. In MVK and PMVK trees, this group diverges from CPR independently from other
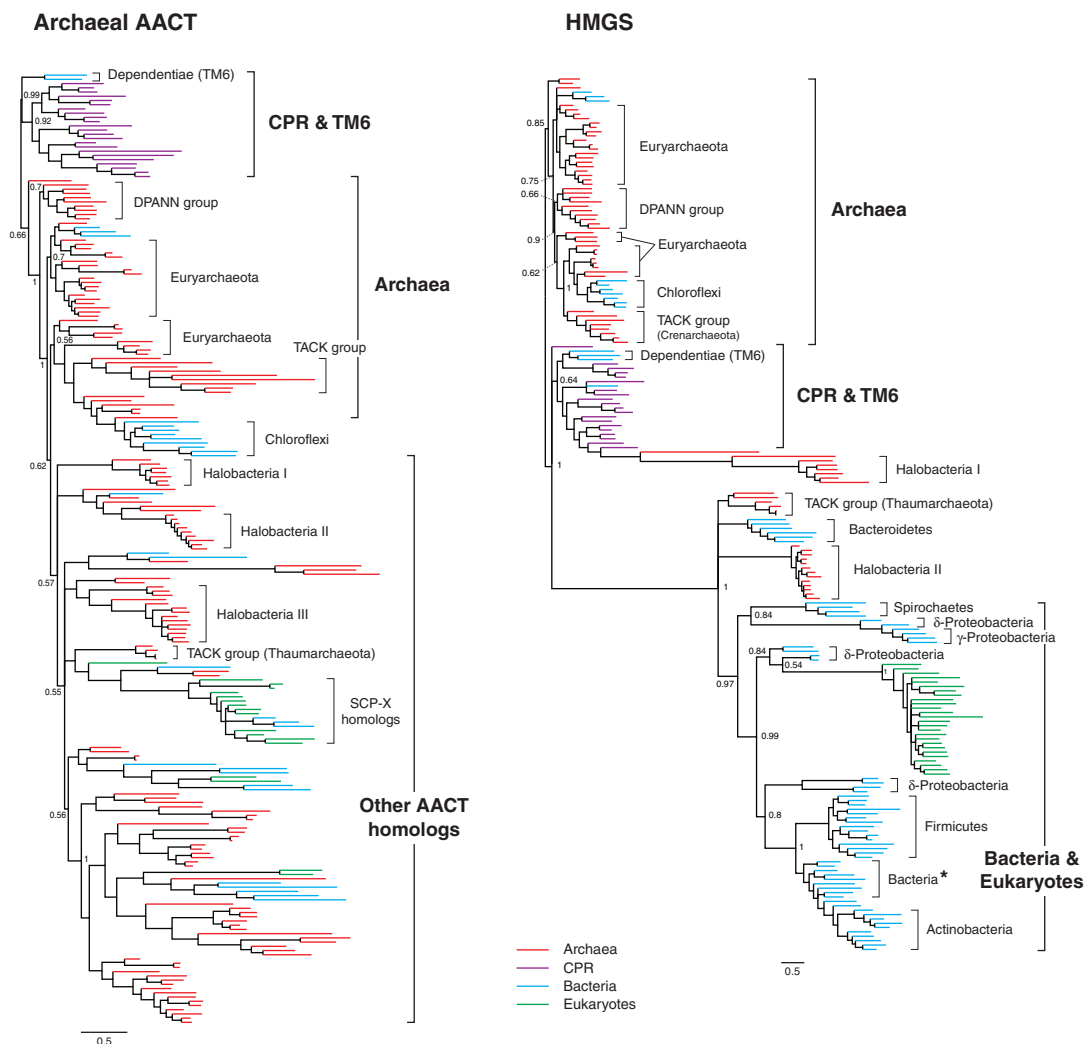
FIG. 3. Bayesian phylogenetic trees of thiolase family enzymes (AACT and HMGS). The complete trees with species annotation are in supplementary figs. 3 and 5, Supplementary Material online. The tree of bacterial/eukaryotic-type AACT (thiolase II) is also in supplementary figure 4, Supplementary Material online. Posterior probabilities are shown for major nodes. Nodes with the posterior probabilities of <0.5 are collapsed. Scale bar represents 0.5 amino acid replacements per site per unit evolutionary time. The bacterial clade with asterisk is composed of specific bacterial species observed in most MVA trees (see the main text). Abbreviation: SCP-X, sterol carrier protein X.

bacteria. This may suggest a separate origin of these two MVA genes in this group.

The MVA enzymes of Chloroflexi seem to have had a different evolutionary history from other bacteria. Many MVA enzymes in Chloroflexi cluster among or near those of archaea (i.e., AACT, HMGS, HMGR-II, MPD, IDI-II, and IPK; figs. 3–5 and supplementary figs. 7 and 12, Supplementary Material online), and a horizontal origin of Chloroflexi MVA enzymes from archaea seems likely. However, for HMGR-II, MPD, and IPK, MVA homologs in Chloroflexi diverge at the root of the archaeal clade and this might imply a very ancient origin of Chloroflexi MVA genes, even if they were indeed horizontally obtained. Chloroflexi is the only bacterial phylum known to possess an archaeal-type MVA pathway utilizing MPD and IPK (MPD route; Dellas et al. 2013) instead of PMVK and DPMD from the bacterial/eukaryotic MVA pathway. The earliest-branching Chloroflexi clade, comprising Dehalococcoidia, Ktedonobacteria and SAR202 cluster, does only contain species having the MEP pathway (Hug et al. 2016; Landry et al.

2017; Shih et al. 2017). This clade forms a sister clade to other chloroflexi in which the MVA pathway is ubiquitous. Therefore, Chloroflexi is made up of the MVA clade and the MEP clade. It is not clear which pathway was present in the common ancestor of Chloroflexi, but in either case the acquisition of the other isoprenoid biosynthesis pathway may have been a critical event for the diversification of Chloroflexi.

In general, bacterial species that possess the MVA pathway lack the MEP pathway and vice versa. Hence, these two pathways are exclusive to each other. Exceptions are seen in Firmicutes and Actinobacteria as MVA-gene carrying species in these groups also have MEP genes (Begley et al. 2004; Dairi 2005). It has been suggested that in these species the MEP pathway is mainly devoted to the production of primary metabolites, whereas the MVA pathway is linked to the secondary metabolite biosynthesis (Dairi 2013). In our current study, the MEP pathway is confirmed in virtually all bacterial phyla that contain at least one sequenced genome (58 out of 59 phyla; data not shown), a sharp contrast to the limited
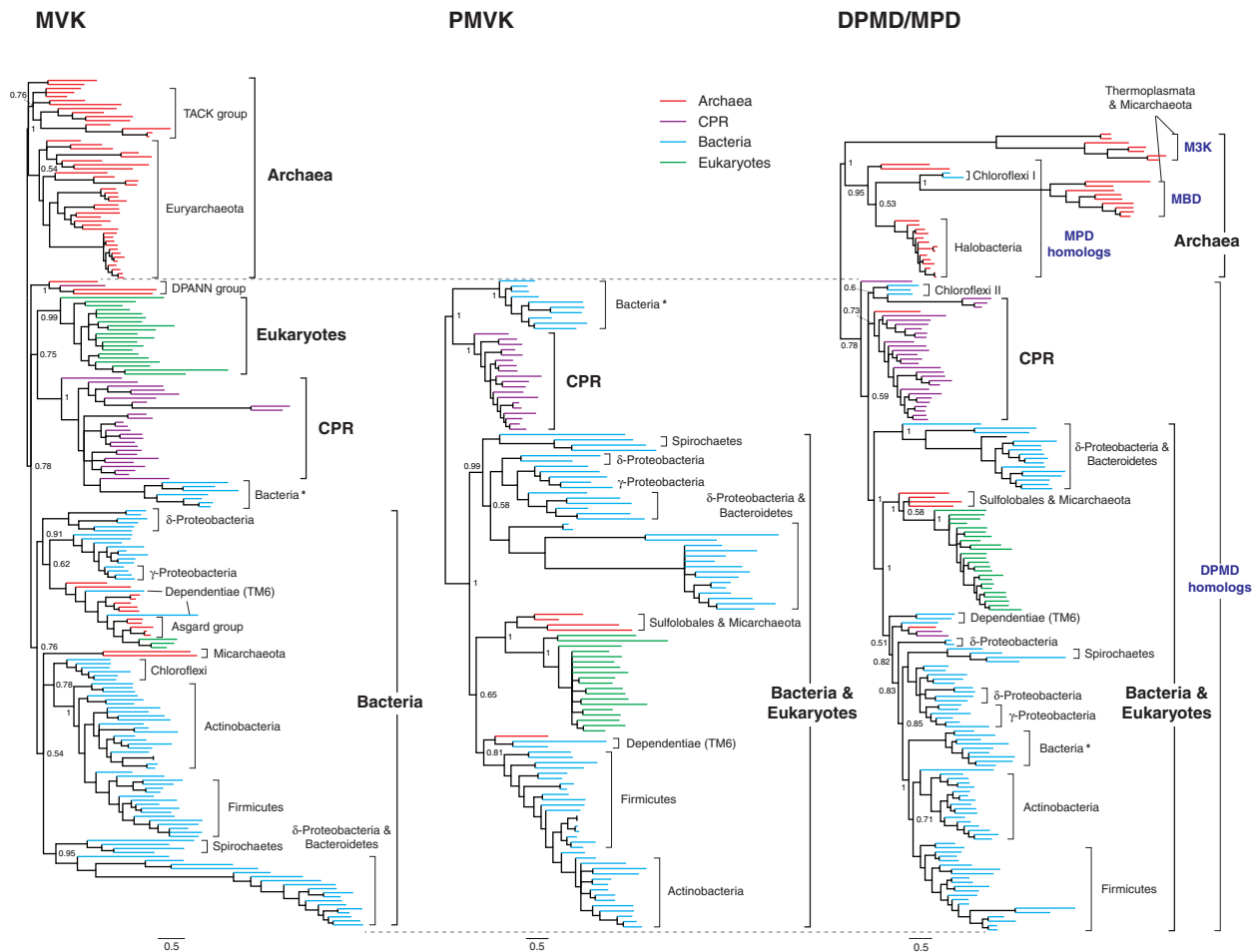
FIG. 4. Bayesian phylogenetic trees of GHMP family enzymes (MVK, PMVK, and DPMD/MPD). The complete trees with species annotations are in supplementary figs. 8–10, Supplementary Material online. Posterior probabilities are shown for major nodes. Nodes with the posterior probabilities of <0.5 are collapsed. Scale bar represents 0.5 amino acid replacements per site per unit evolutionary time.

distribution of the MVA pathway in bacteria. The only exception to this observation is the phylum Dependentiae which completely lacks the MEP pathway and instead solely has the MVA pathway (TM6; Yeoh et al. 2016). Interestingly, some MVA enzymes from this phylum cluster with homologs from the superphylum CPR. This is consistent with the possibility of a close evolutionary relationship between Dependentiae and CPR.

Archaea

Archaea have been inferred to possess an alternative MVA pathway utilizing MPD and IPK as seen in chloroflexi instead of using the PMVK and DPMD enzymes from the bacterial/eukaryotic MVA pathway (fig. 1) (Dellas et al. 2013). The early-step MVA enzymes (AACT, HMGS, and HMGR) and the last-step MVA enzyme (IPK) are universally distributed in archaea, while the middle-step enzymes (GHMP proteins) are largely absent, except for MVK (fig. 2). MPD homologs have only been discovered in Halobacteria (an archaeal class) and Thermoplasmata. Thus, the isoprenoid biosynthesis pathway in most archaea is still not fully understood. The sole exception to the above observations is the order Sulfolobales that

has the bacterial/eukaryotic MVA pathway (Nishimura et al. 2013).

Our current study confirmed the limited distribution of MPD, with a few exceptions. Thermoplasmata and a few Micarchaeota species from the DPANN group are found to possess two distinctive groups of MPD homologs within the archaeal clade (M3K and MBD; fig. 4). Each group contains proteins that are recently confirmed to be involved in the third route of the MVA pathway (M3K route) (Azami et al. 2014; Vinokur et al. 2016). The presence of two divergent groups within the archaeal clade may suggest the gene duplication of an ancestral MPD and a subsequent neofunctionalization in a common ancestor of Thermoplasmata, accompanied with the modification of their MVA pathway. The occurrence of M3K and MBD homologs is limited to Thermoplasmata and a few Micarchaeota species. Thus, the M3K route is inferred to be specific to these archaeal taxa. Indeed, gene transfers between Thermoplasmata and Micarchaeota have recently been suggested (Chen et al. 2018). Thermoplasmata additionally possesses a few more MPD homologs, but their function is not yet known (see Supplementary Material online).

The MVA gene set in the DPANN group varies depending on species. In our current study, MVA enzymes were uncovered in three DPANN phyla (Micarchaeota, Aenigmarchaeota, and Diapherotrites). In Micarcheaota, besides the species that have M3K, MBD, and IPK genes (M3K route), there are a few other species that possess MVK, PMVK, and DPMD genes (DPMD route), which was previously only known to Sulfolobales in the archaeal domain (fig. 4). Species in the other two DPANN phyla mostly only have MVK and IPK similar to the majority of archaea. The DPANN group represents a deep-branching clade of archaea, although its taxonomical position is still debated as its symbiotic lifestyle and accompanying fast gene divergence rate may cause an artifactual deep-branching position (Adam et al. 2017). The phylogenetic positions of MVA enzymes from DPANN archaea are not consistent. For the first three enzymes (AACT, HMGS, and HMGR), all DPANN archaea form a monophyletic clade and cluster together with other archaea (fig. 3 and supplementary figs. 3, 5, and 7, Supplementary Material online). Hence, the archaeal common ancestor is inferred to have possessed these three MVA enzymes. For the fourth enzyme in the pathway (MVK), DPANN archaea do not cluster with other archaea (fig. 4). However, considering the universal distribution of MVK in all archaea, MVK could also have been present in the archaeal common ancestor. In contrast, other GHMP enzymes (PMVK and DPMD) involved in the latter part of the MVA pathway are largely absent (Aenigmarchaeota, Diapherotrites, and most Micarchaeota species) or grouped among the bacterial/eukaryotic clade (a few Micarchaeota species) (fig. 4). Thus, GHMP enzymes in Micarchaeota, except for MVK, are most likely to be attributed to HGT. However, PMVK and DPMD from Micarchaeota cluster together with homologs from Sulfolobales at the root of the eukaryotic clade and thus suggests that the gene transfer to Micarchaeota occurred during the early evolution of eukaryotes. In contrast to the variations in MVA pathways observed for Micarchaeota, archaea from the other two DPANN phyla mostly have the identical set of MVA genes as seen in archaea. The ubiquity of the archaeal-type MVA pathway in the entire archaeal domain suggests the presence of this type of pathway in the archaeal common ancestor.

A few DPANN archaea are also found to possess enzymes from the MEP pathway that are otherwise completely missing from the archaeal domain. The phylogenetic tree of one MEP enzyme suggests that the DPANN group forms a monophyletic clade along with CPR and FCB group bacteria (supplementary fig. 14, Supplementary Material online). In addition, the distribution of MEP enzymes is sparse in the DPANN group. Hence, the MEP enzymes in this group are likely to be attributed to HGT from FCB group bacteria (see Supplementary Material online for a more detailed analysis).

Candidate Phyla Radiation

In our current study, the recently proposed superphylum CPR is found to possess all necessary homologs of MVA enzymes involved in the bacterial/eukaryotic MVA pathway (AACT,

HMGS, HMGR-I, MVK, PMVK, DPMD, and IDI-I; fig. 2). Most MVA enzymes in CPR form a distinct monophyletic clade separate from the three domains of life. The only exception to this observation is for IDI-I. The IDI-I tree is not well separated between bacteria, CPR, and archaea (supplementary fig. 11, Supplementary Material online). CPR is taxonomically divided into two major groups: microgenomates (previously known as OP11) and parcubacteria (OD1) (Rinke et al. 2013; Brown et al. 2015). The majority of microgenomates possess MVA enzymes, whereas only a few parcubacterial phyla possess them. Phylogenies of the early-step MVA enzymes (AACT, HMGS, and HMGR-I) are congruent with the suggested species relationship of CPR (Hug et al. 2016). Phylogenies of the later-step MVA enzymes (MVK, PMVK, DPMD, and IDI-I) is less congruent with the species tree, but CPR mostly form a monophyletic clade for each enzyme, and the distribution of the later-step MVA enzymes in CPR is exactly the same as that of the early-step MVA enzymes in CPR. Thus, it is inferred that all MVA enzymes existed in the common ancestor of CPR. Although the distribution of MVA enzymes in parcubacteria is very limited, several basal parcubacteria (e.g., peregrinibacteria, dojkabacteria) do possess MVA enzymes. Thus, the loss of MVA genes at an early stage of the parcubacteria evolution is inferred. In addition, the first two MVA enzymes (AACT and HMGS) from the bacterial phylum Dependentiae cluster together with CPR homologs. Correspondingly, this bacterial phylum has been suggested to form a sister clade to the CPR superphylum (Brown et al. 2015; Yeoh et al. 2016). However, this is not observed in other MVA enzyme trees. Additional molecular data may provide more insight.

The tree topology of CPR, bacteria, and archaea observed in the MVA enzyme trees matches the species relationship of the domains Bacteria and Archaea (figs. 3 and 4). In most MVA trees, CPR occupies the basal position of the bacterial domain as predicted from the species relationship of CPR and bacteria, except for the unresolved HMGR-I and IDI-I trees (Hug et al. 2016). Therefore, the bacterial common ancestor is inferred to have possessed the bacterial/eukaryotic MVA pathway. As discussed earlier, it is not clear if MVA genes are vertically inherited within bacteria, but in any case, the origin of bacterial MVA genes seems to be within the bacterial/CPR clade. Further, some MVA enzymes found in CPR have a close relationship to those in archaea (Archaeal AACT, HMGS, HMGR, and MVK) (figs. 3 and 4 and supplementary fig. 6, Supplementary Material online). This suggests that the common ancestor of CPR and archaea, or LUCA, may have had those MVA enzymes.

Interestingly, there are differences between the set of MVA genes in CPR and those in bacteria. In addition to having the MVA genes involved in the bacterial/eukaryotic MVA pathway, CPR is also found to possess two MVA genes involved in the archaeal MVA pathway: archaeal AACT and IPK (fig. 2). These genes have only been observed in archaea and chloroflexi, in addition to CPR. Hence, CPR bears the characteristics of both the bacterial/eukaryotic and archaeal MVA pathways. For instance, thiolase II (which is universally distributed in bacteria/eukaryotes) is replaced by the archaeal AACT in CPR (fig. 2). In the other instance, the archaeal IPK homolog
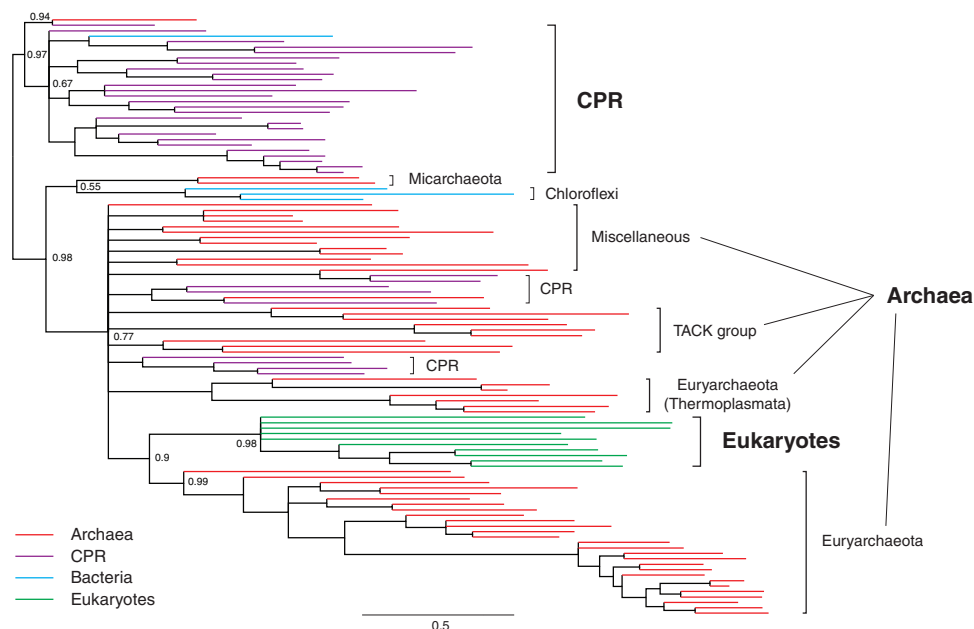
Fig. 5. Bayesian phylogenetic tree of IPK homologs. The complete tree with species annotation is in supplementary figure 13, Supplementary Material online. Posterior probabilities are shown for major nodes. Nodes with the posterior probabilities of <0.5 are collapsed. Scale bar represents 0.5 amino acid replacements per site per unit evolutionary time.

is observed in CPR but its presence seems to be irrespective of other MVA genes for these species. Nevertheless, the IPK gene mostly has a syntenic relationship with other MVA genes in the CPR species that possess both the IPK gene and other MVA genes (fig. 2). Even in CPR species that lack most MVA genes, the IPK gene is typically syntenic with the IDI-I gene that encodes an enzyme to isomerize one product of the MVA pathway into the other product of the pathway (IPP and DMAPP; fig. 1). Additionally, the distribution of the IPK gene in CPR is mostly limited to the microgenomate group as is the case for the distribution of other MVA genes in CPR. In the IPK tree, most homologs in CPR cluster separately from those in archaea (fig. 5), and thus it is inferred that the IPK gene was present in the common ancestor of CPR and archaea (LUCA), and is thus not attributed to a recent HGT from archaea.

MVA genes in CPR further displays a characteristic syntenic relationship with isoprenyl diphosphate synthase (IPPS) genes, which is not observed in bacteria. IPPS is involved in the isoprenoid chain elongation following the MVA/MEP pathway, condensing two end-products of the MVA/MEP pathway (IPP and DMAPP; fig. 1). There are two types of IPPS, depending on the stereochemistry of the reaction product: *trans*-IPPS and *cis*-IPPS (Liang et al. 2002). These two types of IPPS do not have a sequence/structural similarity to each other and have different evolutionary origins. *trans*-IPPS is essential for the biosynthesis of many primary metabolites such as quinones and hemes in the three domains of life (Nowicka and Kruk 2010; Hederstedt 2012) and of archaeal membrane lipids (Jain et al. 2014). Meanwhile, *cis*-IPPS plays a critical role in the N-glycosylation in the three domains of life (Jones et al. 2009). MVA genes in CPR typically form a

syntenic cluster with the *cis*-IPPS gene (fig. 2). Meanwhile, archaeal MVA genes are commonly syntenic with the *trans*-IPPS gene (fig. 2). This could reflect a connection between the MVA pathway and the biosynthesis of isoprenyl lipid membranes unique to the archaeal domain. In contrast, no synteny is observed between *trans*/*cis*-IPPS genes and MVA genes in bacteria. Instead, both *trans*- and *cis*-IPPS genes in bacteria are commonly syntenic with MEP genes. This is in contrast to the connection between IPPS genes and MVA genes for CPR. The MEP pathway is sparsely distributed in CPR (supplementary fig. 14, Supplementary Material online) and it is not clear whether the MEP pathway was present in the common ancestor of CPR or not (see Supplementary Material online for a more detailed analysis). In either case, the connection between IPPS and MEP genes in bacteria is inferred to have been established after the divergence of bacteria from CPR.

Evolution of MVA Enzymes

As discussed earlier, the bacterial/eukaryotic-type MVA pathway and the archaeal-type MVA pathway are inferred to be present in the common ancestor of bacteria/CPR and of archaea, respectively. However, CPR possesses MVA enzymes of both pathways and provides unique information about the composition of the ancestral MVA pathway present in LUCA. One possible evolutionary history of individual MVA enzymes and a hypothetical MVA gene set in LUCA is presented below (fig. 6).

Thiolase Family (AACT and HMGS)

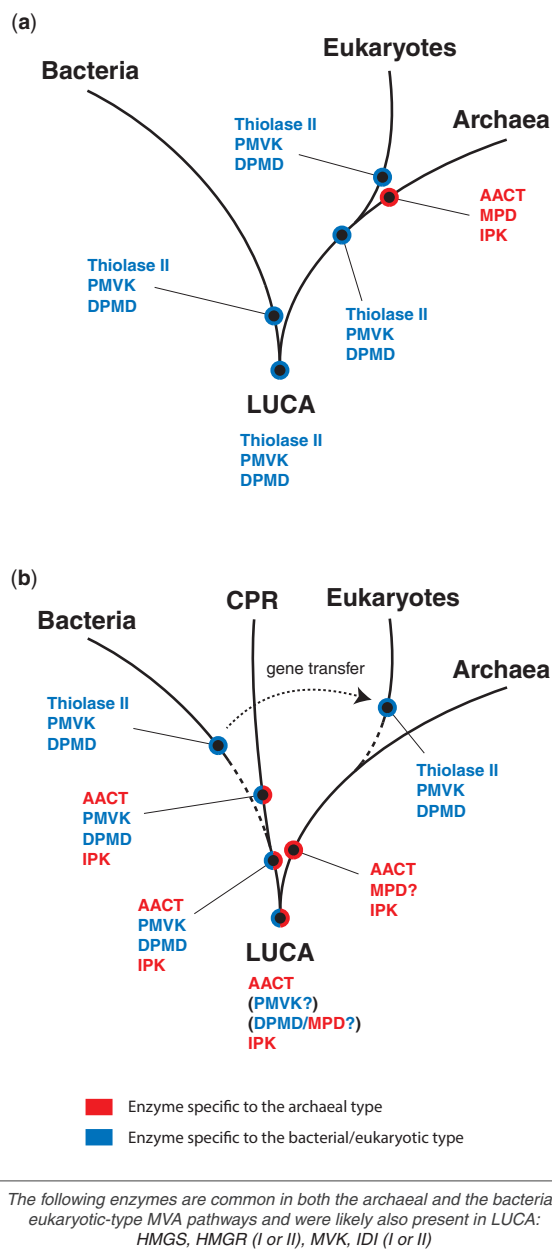The first two enzymes of the MVA pathway, AACT and HMGS, are distantly related to each other (Jiang et al. 2008),

Fig. 6. Hypotheses for the evolution of MVA enzymes. (*a*) Previous hypothesis (Lombard and Moreira 2011). (*b*) New hypothesis proposed in our current study. MVA enzymes common in both the archaeal-type and the bacterial/eukaryotic-type pathways are treated separately at the bottom of the figure. The dashed line indicates that the vertical inheritance of MVA genes along the line is not decisive or unlikely (see the main text).

yet both belong to the thiolase family that is distributed universally among the three domains of life. For both AACT and HMGS, the archaeal/CPR clade and the bacterial/eukaryotic clade are distant from each other, respectively (fig. 3 and supplementary fig. 4, Supplementary Material online). Although MVA enzymes in CPR and bacteria are inferred to share a common ancestor, there is a substantial divergence between CPR and bacteria for both AACT and HMGS. Some HMGS homologs from Halobacteria and Thaumarchaeota cluster near the bacterial clade. This may be due to a

substantial amount of gene transfer from bacteria to these archaea (Nelson-Sathi et al. 2012; Wagner et al. 2017). Within the CPR clade and the early-branching archaeal clades, the tree topology is consistent with the proposed species tree of CPR and archaea. Archaeal species typically have multiple AACT homologs and they form several distinct clades (Other AACT homologs; fig. 3). These clades sometimes contain proteins from bacteria and eukaryotes, including those that are known as sterol carrier proteins (SCP-X; Peretó et al. 2005). These nonarchaeal homologs cluster among archaeal homologs and are generally divergent relative to nearby archaeal homologs. Therefore, these nonarchaeal homologs are inferred to be derived from individual HGT events originated from archaea. The function of AACT homologs outside the early-branching archaeal clades is largely unknown, but many of them lack the conserved catalytic motif (CH motif; Jiang et al. 2008) and thus possibly have different functions (see Supplementary Material online for a more detailed discussion).

In contrast to archaea and CPR, AACT homologs in eukaryotes and bacteria (thiolase II) do not follow the species relationship of eukaryotes and bacteria (supplementary fig. 4, Supplementary Material online). Thiolase II is universally distributed in bacteria, regardless of the presence of the MVA pathway (data not shown). Eukaryotic thiolase II seems to have been transferred from bacteria multiple times independently. In addition, some archaea possess thiolase II homologs, but they cluster among bacterial homologs (supplementary fig. 4, Supplementary Material online). Thus, it is inferred that thiolase II evolved in the bacterial domain irrelevant to isoprenoid biosynthesis. Indeed, thiolase II is not specific to the isoprenoid synthesis in bacteria and is also involved in other biosynthetic pathways such as hydroxybutyric acid synthesis (Peretó et al. 2005). No syntenic relationship between the thiolase II gene and MVA genes exists in most bacteria and this also points to a different origin of the thiolase II gene from MVA genes in bacteria (fig. 2).

Consequently, the common ancestor of archaea and CPR, or LUCA, could have had the archaeal AACT gene and the HMGS gene based on the universal distribution of these genes in both CPR and archaea and the tree topologies consistent with the species tree (fig. 6). In contrast, thiolase II probably has a different origin in the bacterial domain after the divergence from CPR and it would have functionally replaced archaeal AACT.

HMGR Family (HMGR-I and II) and IDI Enzymes (IDI-I and II)

The evolutionary history of HMGR is complicated by the split of HMGR into two distinct types that are distantly related to each other (Type I and II). Since the HMGR-I tree of the CPR clade is consistent with the species tree of CPR (fig. 2) and HMGR-II is completely absent in CPR, the presence of HMGR-I in the common ancestor of CPR is inferred. In contrast, archaea and bacteria possess both types of HMGR. Previously, HMGR enzymes found in archaea were mostly limited to HMGR-I, while bacteria commonly have HMGR-II (Lombard and Moreira 2011). However, our current study detects numerous HMGR-II homologs in archaea, mainly in the DPANN group, Asgard group and TACK group

(supplementary fig. 7, Supplementary Material online). Thus, a clear distinction between archaea and bacteria no longer exists and both HMGR-I and II may have already been present in the common ancestor of archaea and bacteria, or LUCA (fig. 6). The absence of HMGR-II in CPR may be explained by gene loss in the common ancestor of CPR. Alternatively, LUCA may have possessed only one type of HMGR and subsequent gene duplication and horizontal transfers among archaea and bacteria may have led to the current complex distribution of HMGR types. However, in both HMGR-I and II trees, each domain of life is roughly separated from each other and also the HMGR tree topology partially matches other MVA enzyme trees. Hence, the occurrence of gene duplication and gene transfers is inferred to be ancient, even if they occurred.

The evolutionary history of IDI enzymes is similarly complicated by the split of IDI into two nonhomologous types (Type I and II). In CPR, IDI-I is universally distributed, while IDI-II is sparse (fig. 2). Hence, it is inferred that IDI-I was present in the common ancestor of CPR, even though IDI-I homologs in CPR do not form a monophyletic clade (supplementary fig. 11, Supplementary Material online). In contrast, IDI-II is predominant in both bacteria and archaea (fig. 2). Thus, as is the case with HMGR, LUCA may have had only one type, followed by enzyme recruitment for the evolution of the other type, or may have had both types with subsequent selective gene loss (fig. 6).

GHMP Family (MVK, PMVK, DPMD, MPD, M3K and MBD)
Although the first three steps of the MVA pathway (AACT, HMGS, and HMGR) are well conserved among all MVA gene-bearing organisms, the latter steps of the pathway have several variations. These latter steps of the pathway are mostly catalyzed by GHMP family enzymes (fig. 1). Bacteria and eukaryotes possess a common MVA pathway composed of MVK, PMVK, and DPMD (DPMD route), reflecting their shared origin that is distinct from the archaeal MVA pathway. CPR possesses the identical set of GHMP proteins as bacteria and eukaryotes. The only exception in bacteria is Chloroflexi, which has MVK and MPD (MPD route). In archaea, halobacteria also have MVK and MPD, while Thermoplasmata and a few Micarchaeota species have M3K and MBD (M3K route). Sulfolobales and a few other Micarchaeota species carry the eukaryotic version of the MVA pathway (DPMD route). In contrast, no other GHMP proteins, except for MVK, have been discovered in other archaea so far.

The observed variations in the latter part of the MVA pathway suggest a complex evolutionary history for this pathway. While the early part of the MVA pathway (AACT, HMGS, and HMGR) would have been well established in LUCA, the reactions catalyzed by GHMP enzymes may not have been fully developed. Among GHMP enzymes, MVK is unique because it is nearly universally distributed among the three domains of life. The only known exceptions occur in Thermoplasmata and a few Micarchaeota species that instead possess M3K and MBD. However, as discussed earlier, these two enzymes are specific to these taxa, and thus are inferred

to be exceptions. Hence, at least one MVK-like enzyme would have existed in LUCA together with the first three enzymes of the MVA pathway (fig. 6). Other GHMP enzymes, PMVK and DPMD homologs, are virtually absent in the archaeal domain, except for a small number of MPD homologs (fig. 4). An MPD-like enzyme might have existed in the archaeal common ancestor, but it is inconclusive (fig. 6). In contrast, the presence of PMVK and DPMD in both CPR and bacteria suggests that these two enzymes could have existed in the common ancestor of the bacterial domain (fig. 6). An archaeal order Sulfolobales that forms a sister clade to the eukaryotic clade in the PMVK and DPMD trees was previously inferred to be a sole archaeal taxon that retains the bacterial/eukaryotic MVA pathway (DPMD route) descended from LUCA (fig. 4) (Boucher et al. 2004; Lombard and Moreira 2011). However, as discussed earlier, the archaeal and eukaryotic PMVK and DPMD homologs are confined within the bacterial/CPR clade and the horizontal origin of these MVA enzymes is more likely (fig. 6).

Evolution of IPK and Archaeal MVA Pathway
In our current study, CPR is discovered to possess MVA enzymes for both the bacterial/eukaryotic MVA pathway and the archaeal pathway (blue and red colors; fig. 6). The phylogenetic analysis of archaeal AACT and IPK suggests the presence of both enzymes in the common ancestor of CPR and archaea, or LUCA (figs. 3 and 5). These observations imply that the archaeal MVA pathway, or at least the archaeal-type MVA gene set (archaeal AACT, HMGS, HMGR, MVK, IPK, and IDI), may have been present in LUCA (fig. 6), contrary to the previous inference (Lombard and Moreira 2011). In contrast, it is unclear if the bacterial/eukaryotic-specific MVA genes (PMVK and DPMD) were present in LUCA. As eukaryotic MVA genes are likely to have been derived from bacteria, the evidence for the bacterial/eukaryotic-specific MVA genes can only be traced back to the bacterial common ancestor in our current study.

There are two possibilities to explain the current variations of MVA pathways seen in bacteria/eukaryotes, CPR and archaea. First, LUCA may have possessed both bacterial/eukaryotic and archaeal MVA enzymes (PMVK, DPMD, archaeal AACT, and IPK). CPR is considered to retain this ancestral state. In contrast, bacteria lost two archaea-specific MVA enzymes (archaeal AACT and IPK), while archaea lost two bacteria/eukaryotes-specific enzymes (PMVK and DPMD). MPD homologs in some archaea may be descendants of ancestral DPMD homolog in LUCA. It is not clear which type of MVA pathway worked in LUCA. The bacterial/eukaryotic-type and archaeal-type MVA pathways might not have been strictly distinguished and functioned simultaneously. Second, LUCA may have only possessed archaeal MVA enzymes. The common ancestor of bacteria and CPR then acquired PMVK and DPMD and established the bacterial/eukaryotic MVA pathway. PMVK and DPMD may have evolved by gene duplications of MVK. In this scenario, MPD homologs in archaea may have been horizontally acquired from ancestral bacteria or CPR. Currently, the function of IPK

homologs in CPR is uncharacterized. Hence, it is not clear if a functional archaeal MVA pathway is present in modern CPR organisms and was present in the common ancestor of CPR and archaea. Although IPK is not always present in CPR species that possess other MVA enzymes, the synteny of the IPK gene with other MVA genes, particularly with the IDI-I gene is common. The characterization of IPK in CPR would be critical to understand its relationship to other MVA enzymes and to decipher the early evolution of the MVA pathway.

Conclusions

In our current study, the recently discovered superphylum CPR is found to possess a potential MVA pathway. CPR carries enzymes for both the bacterial/eukaryotic-type and the archaeal-type MVA pathways and may retain the characteristics of the ancestral MVA pathway possibly present in LUCA. The modern bacterial/eukaryotic MVA pathway is inferred to have diverged from the CPR lineage. In addition, the MVA pathway in eukaryotes is inferred to have emerged via HGT from bacteria. The first four steps of the MVA pathway are well conserved in both bacteria and archaea and are likely to have already been established in LUCA. In contrast, the later steps catalyzed by GHMP enzymes seem not to have been fully developed in LUCA, reflecting the observed variations in modern bacterial and archaeal MVA pathways. The characterization of a potential MVA pathway in CPR, including the function of enigmatic IPK homologs, could be a key to decipher the early evolution of the MVA pathway.

Materials and Methods

Data Set Construction

Thirteen enzymes known to be involved in the IPP biosynthesis of the MVA pathway were analyzed in this study (AACT, HMGS, HMGR-I, HMGR-II, MVK, PMVK, DPMD, MPD, M3K, MBD, IDI-I, IDI-II, and IPK; fig. 1). AACT, the first enzyme in the MVA pathway, was excluded from previous studies as this enzyme is not specific to the isoprenoid biosynthesis (Boucher et al. 2004; Lombard and Moreira 2011). However, the phylogeny of AACT homologs found in archaea displayed a strong relationship to that of HMGS, so archaeal AACT homologs (archaeal AACT) were also analyzed in this study. In contrast, bacterial and eukaryotic AACT homologs form a distant clade from archaeal AACT. They consist of two groups (thiolase I and II) and only thiolase II is involved in the isoprenoid biosynthesis (Peretó et al. 2005). Thus, only thiolase II homologs were analyzed. Among the other 12 enzymes, eight enzymes are involved in the MVA pathway found in eukaryotes and bacteria (HMGS, HMGR-I/II, MVK, PMVK, DPMD, and IDI-I/II). However, Type I and II HMGR and Type I and II IDI are functionally equal, respectively. Subsequently, the remaining six enzymes are functionally distinct from each other in the bacterial/eukaryotic MVA pathway. As for archaea, HMGS, HMGR-I/II, MVK, MPD/M3K, IPK, and IDI-I/II are six functionally distinctive enzymes. In our current study, organisms that have at least three homologs out of these six enzymes were included in the data set.

Representative sequences for individual enzymes were identified from GenBank (http://www.ncbi.nlm.nih.gov/); AACT (accession number: NP_002970.2 and AAD34967.1), HMGS (EPX58968.1), HMGR-I (EPX55455.1), HMGR-II (EPX62345.1), MVK (EPX62344.1), PMVK (EPX62342.1), DPMD (EPX62343.1), MPD (ABU57050.1), M3K (CAC12426.1), IDI-I (CAO97285.1), and IDI-II (EPX62346.1). Homologous protein sequences for all MVA enzymes, except for AACT, were retrieved from GenBank, using BLASTp (altschul, 1990), with the cutoff threshold of $<1\times10^{-5}$. Only for AACT, the cutoff threshold was $<1\times10^{-20}$ as homologs with the E-value of $>1\times10^{-20}$ only display a similarity in a very limited region. Sequences with the minimum length of 200 amino acids were collected for most enzymes, except for IDI-I with the minimum lengths of 100 amino acids. At least one homolog from an individual taxonomical family was included in the data set. For some taxa, the distribution of the MEP pathway was compared with that of the MVA pathway. Species with at least three out of seven MEP enzymes (supplementary fig. 1, Supplementary Material online) were regarded as having a potential MEP pathway and were included in the analysis. Representative sequences for individual enzymes were identified from GenBank: DXS (BAC43863.1), DXR (BAC43938.1), MCT (BAC44069.1), CMK (BAC44823.1), MDS (BAC44812.1), HDS (BAC44727.1), and HDR (BAC43925.1). Sequences with the minimum length of 200 amino acids were collected. Nonbootstrapped phylogenetic trees were constructed first to identify major clusters for each enzyme and to eliminate spurious sequences. Then, a representative sequence from each cluster was used to further investigate homologous sequences within the cluster. Thiolase I and II homologs were clearly identified by this preliminary tree as they formed two distinctive clades, and thus thiolase I homologs were excluded for further analysis as mentioned earlier.

Phylogenetic Analysis

Sequences were aligned using T-Coffee (Notredame et al. 2000) and Muscle (Edgar 2004). Based on a preliminary alignment, aberrant sequences were omitted. Irregularly long 5′ and 3′ ends and indels were removed when present in only a couple of species, yet indels which seem related to a specific taxonomical clade were retained. In cases where the targeted protein was fused to a different protein, the unrelated protein sequence was identified and removed based on the preliminary alignment. Subsequently, the final alignments were generated. The tree construction was carried out using RAxML v.8.2.10 (Stamatakis 2014) and PhyloBayes v.4.1 (Lartillot et al. 2009). The RAxML analysis was performed with the hillclimbing mode using the gamma substitution model. RAxML was used only for preliminary nonbootstrapped trees to identify major clusters for each enzyme. The Bayesian analysis was conducted using both the single substitution model (LG+G) and the profile mixture model (CAT-GTR+G, C40). For each analysis, two chains were run and the convergence was assessed using the bpcomp and tracecomp programs in PhyloBayes. 20% of sampled points were discarded as burnin. The global tree topology for the three domains of life was

broadly identical between the LG model and the CAT model. There are differences in the branching order within individual domains of life, but they do not affect the interpretation of the data. The Bayesian trees based on the CAT model are shown. Similarly, the global tree topology based on the Muscle alignment was nearly identical to the T-Coffee alignment and trees based on T-Coffee alignments are shown. Three different species trees that were recently published (Toft and Andersson 2010; Raymann et al. 2015; Hug et al. 2016) were used as reference species trees.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

References

Adam PS, Borrel G, Brochier-Armanet C, Gribaldo S. 2017. The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *ISME J.* 11(11):2407.

Azami Y, Hattori A, Nishimura H, Kawaide H, Yoshimura T, Hemmi H. 2014. (R)-mevalonate 3-phosphate is an intermediate of the mevalonate pathway in *Thermoplasma acidophilum. J Biol Chem.* 289(23):15957–15967.

Begley M, Gahan CGM, Kollas A-K, Hintz M, Hill C, Jomaa H, Eberl M. 2004. The interplay between classical and alternative isoprenoid biosynthesis controls $\gamma\delta$ T cell bioactivity of *Listeria monocytogenes. FEBS Lett.* 561(1–3):99–104.

Boucher Y, Kamekura M, Doolittle W. 2004. Origins and evolution of isoprenoid lipid biosynthesis in archaea. *Mol Microbiol.* 52(2):515–527.

Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, Wilkins MJ, Wrighton KC, Williams KH, Banfield JF. 2015. Unusual biology across a group comprising more than 15% of domain bacteria. *Nature* 523(7559):208.

Buckingham J. 2007. Dictionary of natural products on DVD, CRC. FL: Boca Raton.

Castelle CJ, Banfield JF. 2018. Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* 172(6):1181–1197.

Chen L-X, Méndez-García C, Dombrowski N, Servín-Garcidueñas LE, Eloe-Fadrosh EA, Fang B-Z, Luo Z-H, Tan S, Zhi X-Y, Hua Z-S, et al. 2018. Metabolic versatility of small archaea Micrarchaeota and Parvarchaeota. *ISME J.* 12(3):756–775.

Crowell DN, Huizinga DH. 2009. Protein isoprenylation: the fat of the matter. *Trends Plant Sci.* 14(3):163–170.

Dairi T. 2005. Studies on biosynthetic genes and enzymes of isoprenoids produced by Actinomycetes. *J Antibiot.* 58(4):227–243.

Dairi T. 2013. Biosynthetic genes and enzymes of isoprenoids produced by Actinomycetes. In: Bach JT, Rohmer M, editors. Isoprenoid synthesis in plants and microorganisms: new concepts and experimental approaches. New York: Springer New York. p. 29–49.

Dellas N, Thomas ST, Manning G, Noel JP. 2013. Discovery of a metabolic alternative to the classical mevalonate pathway. *eLife* 2:e00672.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32(5):1792.

Eme L, Spang A, Lombard J, Stairs CW, Ettema TJG. 2017. Archaea and the origin of eukaryotes. *Nat Rev Microbiol.* 15(12):711.

Gaucher EA, Das UK, Miyamoto MM, Benner SA. 2002. The crystal structure of eEF1A refines the functional predictions of an evolutionary analysis of rate changes among elongation factors. *Mol Biol Evol.* 19(4):569–573.

Gaucher EA, Graddy LG, Li T, Simmen RC, Simmen FA, Schreiber DR, Liberles DA, Janis CM, Benner SA. 2004. The planetary biology of cytochrome P450 aromatases. *BMC Biol.* 2:19.

Gershenzon J, Dudareva N. 2007. The function of terpene natural products in the natural world. *Nat Chem Biol.* 3(7):408–414.

Hayakawa H, Sobue F, Motoyama K, Yoshimura T, Hemmi H. 2017. Identification of enzymes involved in the mevalonate pathway of Flavobacterium johnsoniae. *Biochem Biophys Res Commun.* 487(3):702–708.

Hederstedt L. 2012. Heme A biosynthesis. *Biochim Biophys Acta Bioenergetics* 1817(6):920–927.

Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, et al. 2016. A new view of the tree of life. *Nat Microbiol.* 1:16048.

Jain S, Caforio A, Driessen AJM. 2014. Biosynthesis of archaeal membrane ether lipids. *Front Microbiol.* 5:641.

Jiang C, Kim SY, Suh D-Y. 2008. Divergent evolution of the thiolase superfamily and chalcone synthase family. *Mol Phylogenet Evol.* 49(3):691–701.

Jones MB, Rosenberg JN, Betenbaugh MJ, Krag SS. 2009. Structure and synthesis of polyisoprenoids used in N-glycosylation across the three domains of life. *Biochim Biophys Acta Gen Subj.* 1790(6):485–494.

Krepkiy D, Miziorko HM. 2004. Identification of active site residues in mevalonate diphosphate decarboxylase: implications for a family of phosphotransferases. *Protein Sci.* 13(7):1875–1881.

Landry Z, Swan BK, Herndl GJ, Stepanauskas R, Giovannoni SJ. 2017. SAR202 genomes from the dark ocean predict pathways for the oxidation of recalcitrant dissolved organic matter. *mBio* 8(2):e00413–17.

Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25(17):2286–2288.

Liang P-H, Ko T-P, Wang AHJ. 2002. Structure, mechanism and function of prenyltransferases. *Eur J Biochem.* 269(14):3339–3354.

Lichtenthaler HK. 1999. The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annu Rev Plant Physiol Plant Mol Biol.* 50:47–65.

Lombard J, Moreira D. 2011. Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Mol Biol Evol.* 28(1):87–99.

Miallau L, Alphey MS, Kemp LE, Leonard GA, McSweeney SM, Hecht S, Bacher A, Eisenreich W, Rohdich F, Hunter WN. 2003. Biosynthesis of isoprenoids: crystal structure of 4-diphosphocytidyl-2C-methyl-d-erythritol kinase. *Proc Natl Acad Sci U S A.* 100(16):9173–9178.

Nelson-Sathi S, Dagan T, Landan G, Janssen A, Steel M, McInerney J, Deppenmeier U, Martin W. 2012. Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. *Proc Natl Acad Sci U S A.* 109(50):20537–20542.

Nishimura H, Azami Y, Miyagawa M, Hashimoto C, Yoshimura T, Hemmi H. 2013. Biochemical evidence supporting the presence of the classical mevalonate pathway in the thermoacidophilic archaeon *Sulfolobus solfataricus. J Biochem.* 153(5):415–420.

Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol.* 302(1):205–217.

Nowicka B, Kruk J. 2010. Occurrence, biosynthesis and function of isoprenoid quinones. *Biochim Biophys Acta Bioenergetics* 1797(9):1587–1605.

Peretó J, López-García P, Moreira D. 2005. Phylogenetic analysis of eukaryotic thiolases suggests multiple proteobacterial origins. *J Mol Evol.* 61(1):65–74.

Pérez-Gil J, Rodríguez-Concepción M. 2013. Metabolic plasticity for isoprenoid biosynthesis in bacteria. *Biochem J.* 452(1):19–25.

Raymann K, Brochier-Armanet C, Gribaldo S. 2015. The two-domain tree of life is linked to a new root for the Archaea. *Proc Natl Acad Sci U S A.* 112(21):6670–6675.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, et al. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499(7459):431–437.

Romanowski M, Bonanno J, Burley S. 2002. Crystal structure of the *Streptococcus pneumoniae* phosphomevalonate kinase, a member of the GHMP kinase superfamily. *Proteins* 47(4):568–571.

Rossoni L, Hall SJ, Eastham G, Licence P, Stephens G. 2015. The putative mevalonate diphosphate decarboxylase from *Picrophilus torridus* is in reality a mevalonate-3-kinase with high potential for bioproduction of isobutene. *Appl Environ Microbiol.* 81(7):2625–2634.

Santner A, Calderon-Villalobos LIA, Estelle M. 2009. Plant hormones are versatile chemical regulators of plant growth. *Nat Chem Biol.* 5(5):301–307.

Shih PM, Ward LM, Fischer WW. 2017. Evolution of the 3-hydroxypropionate bicycle and recent transfer of anoxygenic photosynthesis into the Chloroflexi. *Proc Natl Acad Sci U S A.* 114(40):10749–10754.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.

Toft C, Andersson SGE. 2010. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat Rev Genet.* 11(7):465.

VanNice JC, Skaff DA, Keightley A, Addo JK, Wyckoff GJ, Miziorko HM. 2014. Identification in *Haloferax volcanii* of phosphomevalonate decarboxylase and isopentenyl phosphate kinase as catalysts of the terminal enzyme reactions in an archaeal alternate mevalonate pathway. *J Bacteriol.* 196(5):1055–1063.

Vinokur JM, Cummins MC, Korman TP, Bowie JU. 2016. An adaptation to life in acid through a novel mevalonate pathway. *Sci Rep.* 6:39737.

Vinokur JM, Korman TP, Cao Z, Bowie JU. 2014. Evidence for a novel mevalonate pathway in archaea. *Biochemistry.* 53:4161–4168.

Wagner A, Whitaker RJ, Krause DJ, Heilers J-H, van Wolferen M, van der Does C, Albers S-V. 2017. Mechanisms of gene flow in archaea. *Nat Rev Microbiol.* 15(8):492.

West CM, van der Wel H, Gaucher EA. 2002. Complex glycosylation of Skp1 in *Dictyostelium*: implications for the modification of other eukaryotic cytoplasmic and nuclear proteins. *Glycobiology* 12(2):17R–27R.

Yang D, Shipman LW, Roessner CA, Scott AI, Sacchettini JC. 2002. Structure of the *Methanococcus jannaschii* mevalonate kinase, a member of the GHMP kinase superfamily. *J Biol Chem.* 277(11):9462–9467.

Yeoh YK, Sekiguchi Y, Parks DH, Hugenholtz P. 2016. Comparative genomics of candidate phylum TM6 suggests that parasitism is widespread and ancestral in this lineage. *Mol Biol Evol.* 33(4):915–927.