Origins and Early Evolution of the Mevalonate Pathway of Isoprenoid Biosynthesis in the Three Domains of Life

Jonathan Lombard and David Moreira*

Unité d'Ecologie, Systématique et Evolution, Centre National de la Recherche Scientifique, Université Paris-Sud, Orsay, France

***Corresponding author:** E-mail: david.moreira@u-psud.fr.

Associate editor: Andrew Roger

Abstract

Isoprenoids are a very diverse family of organic compounds widespread in the three domains of life. Although they are produced from the condensation of the same precursors in all organisms (isopentenyl pyrophosphate and dimethylallyl diphosphate), the evolutionary origin of their biosynthesis remains controversial. Two independent nonhomologous metabolic pathways are known: the mevalonate (MVA) pathway in eukaryotes and archaea and the methylerythritol phosphate (MEP) pathway in bacteria and several photosynthetic eukaryotes. The MVA pathway is also found in a few bacteria, what has been explained in previous works by recent acquisition by horizontal gene transfer (HGT) from eukaryotic or archaeal donors. To reconsider the question of the evolutionary origin of the MVA pathway, we have studied the origin and the evolution of the enzymes of this pathway using phylogenomic analyses upon a taxon-rich sequence database. On the one hand, our results confirm the conservation in archaea of an MVA pathway partially different from eukaryotes. This implies that each domain of life possesses a characteristic major isoprenoid biosynthesis pathway: the classical MVA pathway in eukaryotes, a modified MVA pathway in archaea, and the MEP pathway in bacteria. On the other hand, despite the identification of several HGT events, our analyses support that the MVA pathway was ancestral not only in archaea and eukaryotes but also in bacteria, in contradiction with previous claims that the presence of this pathway in bacteria was due to HGT from other domains. Therefore, the MVA pathway is likely an ancestral metabolic route in all the three domains of life, and hence, it was probably present in the last common ancestor of all organisms (the cenancestor). These findings open the possibility that the cenancestor had membranes containing isoprenoids.

Key words: isoprenoids, mevalonate pathway, early evolution, cenancestor.

Introduction

Isoprenoids constitute one of the largest families of biological compounds, encompassing around 30,000 known products in the three domains of life (bacteria, archaea, and eukaryotes). They fulfill diverse biochemical functions as major structural membrane components in archaea, photosynthetic pigments, hormones, quinones acting in electron transport chains, and plant defense compounds (McGarvey and Croteau 1995; Lange et al. 2000). All these molecules are oligomers synthesized by organisms of the three domains of life through successive condensations of two activated forms of isoprene: the isopentenyl pyrophosphate (IPP) and the dimethylallyl diphosphate (DMAPP) (Ruzicka 1953). Although all known organisms seem to employ IPP and DMAPP, these two isoprenoid precursors can be synthesized by two independent and nonhomologous metabolic pathways, namely, the mevalonate (MVA) and the methylerythritol phosphate (MEP) pathways (fig. 1 and supplementary fig. 1, Supplementary Material online).

The MVA pathway (fig. 1) was first discovered in yeasts and animals in the 1950s, and during some time, it was assumed to be responsible for the production of IPP and DMAPP from acetyl-coenzyme A (CoA) in all organisms (McGarvey and Croteau 1995), even if it seemed to be absent in most bacteria (Zhou and White 1991). The MEP pathway, capable of generating IPP and DMAPP from D-glyceraldehyde-3-phosphate (G3P) and pyruvate, was discovered in the 1990s in bacteria and plants (Rodríguez-Concepción and Boronat 2002; Phillips et al. 2008). Today, the MVA pathway is considered to be the major route of IPP and DMAPP synthesis in eukaryotes and archaea (Kuzuyama 2002; Boucher et al. 2004), whereas the MEP pathway would be characteristic of bacteria (Lange et al. 2000). However, several exceptions to this general rule are known: First, some bacteria do possess the MVA pathway (Bochar et al. 1999; Romanowski et al. 2002; Voynova et al. 2004; Steussy et al. 2006), although this has been commonly explained as a horizontal gene transfer (HGT) acquisition from archaeal or eukaryotic donors (Boucher and Doolittle 2000; Wilding et al. 2000). Second, plants, other photosynthetic eukaryotes, and apicomplexa (a clade of parasitic protists derived from a photosynthetic ancestor) have been found to contain the MEP pathway in addition to the MVA one. All these eukaryotic groups carry plastids (either photosynthetic or, in the case of apicomplexa, highly derived), suggesting that they obtained the MEP pathway through the transfer of genes from the original cyanobacterial endosymbiont that originated the plastids (Lange et al. 2000). However, although the phylogenies of some of the enzymes of this pathway support the expected cyanobacterial origin, some others appear to support

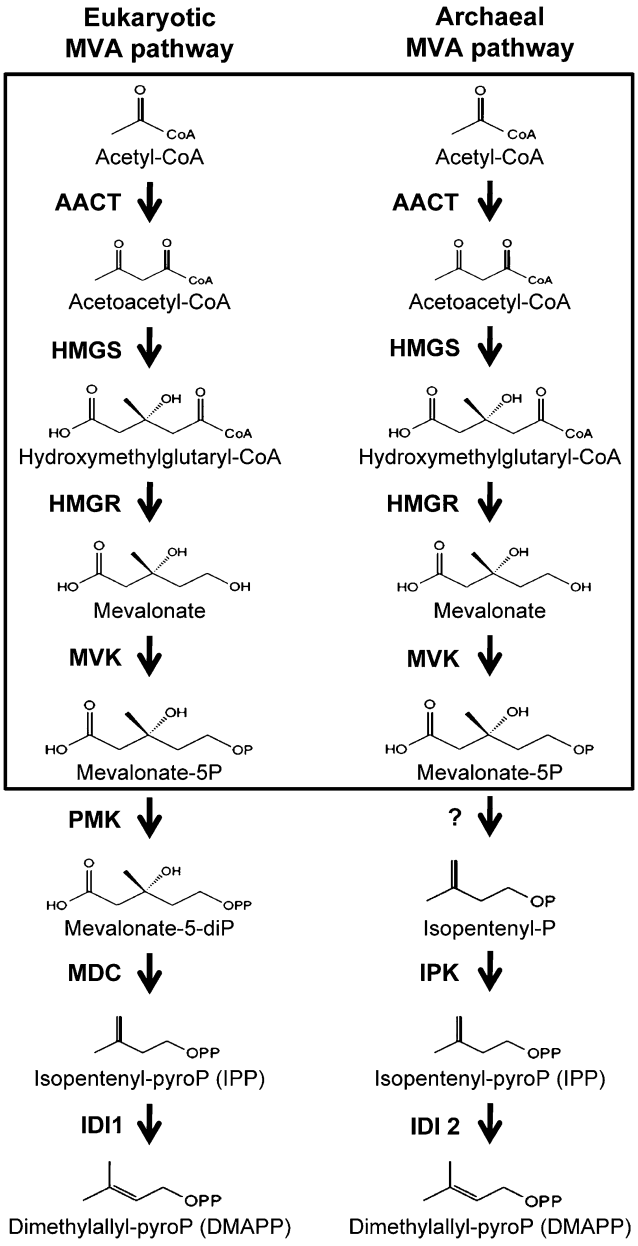

Fig. 1 The partially different mevalonate pathways of eukaryotes and archaea. The classical mevalonate pathway discovered in eukaryotes (left) and the new mevalonate pathway proposed in archaea (right) only share the first four steps (boxed area).

different noncyanobacterial origins (Brinkman et al. 2002; Matsuzaki et al. 2008; Moustafa et al. 2008). Finally, most archaeal species lack the three last enzymes of the classic (eukaryotic-like) MVA pathway, as the analysis of some archaeal genomes already revealed several years ago (Smit and Mushegian 2000; Boucher et al. 2004). These enzymes are the phosphomevalonate kinase (PMK), the mevalonate-5-decarboxylase (MDC), and the isopentenyl diphosphate isomerase (IDI), which are only present in some restricted archaeal clades maybe as a result of proposed HGTs from eukaryotes or bacteria (Boucher et al. 2004).

The modifications of the MVA pathway in archaea have been studied in particular detail. To explain the absence of some MVA pathway enzymes in these organisms, some authors have underlined the existence of examples of nonhomologous enzyme recruitment to replace some steps of the MVA pathway in other groups. This is the case of the nonorthologous PMK similar to viral nucleoside monophosphate kinases used by metazoa instead of the regular gene from the galactokinase, homoserine kinase, mevalonate kinase (MVK), and PMK family (Houten and Waterham 2001; Herdendorf and Miziorko 2006). These findings have led to propose the same kind of nonhomologous replacements in archaea (Smit and Mushegian 2000), and several studies have tried to identify the enzymes carrying out the analogous reactions required to complete the pathway in this domain of life. A nonhomologous enzyme able to catalyze the same reaction as IDI has been discovered in *Streptomyces* sp. and called IDI2 (Kaneda et al. 2001); this IDI2 has not to be mistaken for another enzyme of mammals, also named IDI2, which is a paralogue of the first IDI to have been isolated, IDI1. Subsequently, IDI2 was found to be encoded in the genomes of different archaea and characterized in several of them (Barkley et al. 2004; Dutoit et al. 2008). In addition, when trying to isolate enzymes in *Methanocaldococcus jannaschii* able to phosphorylate the mevalonate-5-phosphate instead of the typical PMK, Grochowski et al. (2006) found a putative alternative mechanism to produce IPP. They identified an enzyme able to phosphorylate the isopentenyl phosphate into IPP (an isopentenyl phosphate kinase, IPK), and they proposed that it could be part of an alternative pathway to produce IPP in *M. jannaschii* in which the order of the phosphorylation and decarboxylation steps would be interchanged (fig. 1).

Despite the ubiquity of isoprenoids, the uneven taxonomic distribution of the two major biosynthetic pathways (MVA in archaea and eukaryotes and MEP in bacteria) opens several questions about their evolutionary origin. Some authors have proposed that the two nonhomologous pathways would have appeared twice independently, one in the bacterial lineage and the other one in the archaeal descent. According to this scenario, the last common ancestor of all organisms (the cenancestor) would have had no isoprenoids (Lange et al. 2000; Martin and Russell 2003). As isoprenoids are essential components of the archaeal membranes, their late emergence might imply, in the opinion of some of these authors (Lange et al. 2000; Martin and Russell 2003), the absence of lipid membranes in the cenancestor. Thereby, the cenancestor has been proposed to be either noncellular (Koga et al. 1998) or limited by a mineral structure instead of a lipid membrane (Martin and Russell 2003).

Some recent phylogenomic analyses have already confirmed the MEP pathway to be restricted to bacteria and plastid-bearing eukaryotes (Matsuzaki et al. 2008). In this work, we have carried out detailed phylogenetic analyses of each enzyme involved in the other major route of isoprenoid biosynthesis, the MVA pathway. Our results complete the data about the existence of an alternative MVA pathway in archaea as proposed by some authors

(Boucher et al. 2004; Grochowski et al. 2006) and suggest that each major pathway for the synthesis of the isoprenoid precursors evolved in each domain of life (the MEP pathway in bacteria and the two MVA pathways in archaea and eukaryotes, respectively). Interestingly, our phylogenies disagree with the hypothesis that all the bacteria bearing the MVA pathway acquired it from archaea or eukaryotes by HGT. On the contrary, this pathway appears to have been present in the last common ancestor of bacteria and, most likely, in the cenancestor. This finding has important implications concerning the nature of lipids in the ancestral membranes and their subsequent evolution during the diversification of the three domains of life.

Materials and Methods

Sequence Retrieval and Alignment

For each domain of life, one protein sequence of each of the MVA pathway enzymes was retrieved from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg), with a few exceptions: The sequences of *Emiliania huxleyi*, *Aureococcus anophagefferens*, and *Naegleria gruberi* were obtained from the Joint Genome Institute (http://genome.jgi-psf.org), the 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR) of *Pseudomonas mevalonii* was obtained from the Protein Data Bank (http://www.rcsb.org/pdb/home/home.do), and the 3-hydroxy-3-methylglutaryl-CoA synthase (HMGS) of *Halobacterium salinarum* and the isopentenyl kinase (IPK) of *M. jannaschii* were obtained from GenBank (http://www.ncbi.nlm.nih.gov/Genbank). Similarity searches with BLASTp (Altschul et al. 1990) were done with these representative sequences as queries against their respective domain of life. In cases where a particular enzyme was missing in KEGG for one domain, we used sequences from the other domains as queries. Similarity searches in archaea and bacteria were done against the completely sequenced genomes available in GenBank (supplementary table 1, Supplementary Material online) and completed with the ongoing genome projects (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi). In eukaryotes, all searches were done against the nonredundant eukaryote-annotated GenBank database.

Sequences found by these searches in the three domains of life were aligned with Muscle 3.6 (Edgar 2004), and alignments were edited and manually refined with the program ED of the MUST package (Philippe 1993). When information was available about functional sites important for the biochemical function of a particular enzyme, it was used in order to identify and remove likely nonorthologous enzymes or enzymes unable to catalyze a given reaction. Redundant and partial sequences were also removed at this step.

Phylogenetic Analyses

The complete sequence data set for each enzyme of the MVA pathway was first analyzed by Neighbor-Joining (Saitou and Nei 1987) using the MUST package (Philippe 1993) to obtain preliminary phylogenetic trees in order to select some representative sequences with which to carry out more detailed maximum likelihood (ML) and Bayesian inference (BI) phylogenetic analyses. ML tree reconstructions were done with the program TREEFINDER (Jobb et al. 2004) with the LG + Γ model (Le and Gascuel 2008) and four rate categories, which was selected as the best-fit model for all our data sets by the model selection tool implemented in TREEFINDER (Jobb et al. 2004). Node support was assessed by 1,000 bootstrap replicates with the same model. BI trees were reconstructed using the program MrBayes v. 3.0b4 (Ronquist and Huelsenbeck 2003) with a mixed substitution model and a Γ distribution of substitution rates with four categories. Searches were run with 4 chains of 1,000,000 generations for which the first 250,000 generations were discarded as "burn in," trees being sampled every 100 generations. Stabilization of the chain parameters was verified using the program TRACER (Rambaut and Drummond 2003).

Results and Discussion

As mentioned above, the MVA pathway, described for the first time in eukaryotes, has traditionally been considered as the main way to produce IPP and DMAPP in eukaryotes and archaea, whereas the MEP pathway has been affiliated to bacteria and plastid-bearing eukaryotes (Lange et al. 2000; Kuzuyama 2002). Nevertheless, two observations raise doubts about this general assumption: archaea lack the three last enzymes of the eukaryotic MVA pathway (Smit and Mushegian 2000; Boucher et al. 2004) and some bacteria, notably some Firmicutes, do possess the genes of the MVA pathway (Boucher and Doolittle 2000; Wilding et al. 2000). In the next sections, we summarize the results of our phylogenetic analyses and propose an alternative view of the evolution of those pathways in the three domains of life.

An Alternative MVA Pathway Is Well Conserved in Archaea

We carried out molecular phylogenetic studies using a selection of representative species for all the enzymes involved in the MVA pathway except for its first enzyme, the acetoacetyl-CoA thiolase (AACT). This enzyme is not specific of this pathway, and it belongs to a large thiolase superfamily with a very complex evolutionary history that has already been studied in detail (Peretó et al. 2005). Our analyses show that all archaeal complete genome sequences encode many of the enzymes needed to accomplish the MVA pathway, with the exception of *Nanoarchaeum equitans* (table 1). This small archaeon lives in an obligatory association with the hyperthermophilic crenarchaeote *Ignicoccus hospitalis* and is known to obtain its lipids from this host (Jahn et al. 2004), which explains the absence of these biosynthetic genes.

Apart from AACT, the first three enzymes of the MVA pathway are shared by eukaryotes and archaea. Their phylogenies showed a general topology congruent with the accepted phylogeny of organisms. As an example, we show the phylogenetic tree of the HMGS responsible for the second step of the pathway (fig. 2). It supported the expected monophyly of groups down to the level of phyla or classes

Table 1. Archaeal Genera Bearing Enzymes of the MVA Pathway.

		HMGS	HMGR	MVK	PMK	MDC	IPK	IDI1	IDI2
Crenarchaeota	***Aeropyrum***	+	I	+	−	−	+	−	+
	Hyperthermus	+	I	+	−	−	+	−	+
	Staphylothermus	+	II	+	−	−	+	−	+
	Ignicoccus	+	I	+	−	−	+	−	+
	Metallosphaera	+	I	+	+	+	+	−	+
	Sulfolobus	+	I	+	+	+	§	−	+
	Caldivirga	+	I	+	−	−	+	−	+
	Pyrobaculum	+	I	+	−	−	+	−	+
	Thermofilum	+	II	+	−	−	+	−	+
	Desulfurococcus	+	II	+	−	−	+	−	+
	Thermoproteus	+	I	+	−	−	+	−	+
Thaumarchaeota	***Cenarchaeum***	+	II	+	−	−	+	+	−
(Crenarchaeota group I)	***Nitrosopumilus***	+	II	+	−	−	+	+	−
Euryarchaeota	***Haloarcula***	+	I	+	−	+	+	+	−
	Halobacterium	+	I	+	−	+	+	+	+
	Haloquadratum	+	I	+	−	+	+	+	−
	Halorubrum	+	I	+	−	+	+	+	+
	Natronomonas	+	I	+	−	+	+	+	+
	Natrialba	+	I	+	−	+	+	+	+
	***Candidatus* Methanoregula**	+	I	+	−	−	+	−	+
	Methanospirillum	+	I	+	−	−	+	−	+
	Methanocorpusculum	+	I	+	−	−	+	−	+
	Methanoculleus	+	I	+	−	−	+	−	+
	Methanococcoides	+	I	+	−	−	+	−	+
	Methanosaeta	+	II	+	−	−	+	−	+
	Methanosarcina	+	I	+	−	−	+	−	+
	Archaeoglobus	+	II	+	−	−	+	−	+
	Methanobrevibacter	+	I	+	−	−	+	−	+
	Methanothermobacter	+	I	+	−	−	+	−	+
	Methanosphaera	+	I	+	−	−	+	−	+
	Methanocaldococcus	+	I	+	−	−	+	−	+
	Methanococcus	+	I	+	−	−	+	−	+
	Pyrococcus	+	I	+	−	−	+	−	+
	Thermococcus	+	I	+	−	−	+	−	+
	Methanopyrus	+	I	+	−	−	+	−	+
	Nanoarchaeum	−	−	−	−	−	−	−	−
	Ferroplasma	+	II	?	−	−	?	−	+
	Picrophilus	+	II	?	−	+	+	−	+
	Thermoplasma	+	II	?	−	+	+	−	+

(+) A homologue of the enzyme can be detected; (−) No detection of homologous sequences; (I) class I HMGR; (II) class II HMGR; (§) our analysis detected IPK homologues in all Sulfolobales but *Sulfolobus acidocaldarius* and *S. tokodaii*; (?) a homologue can be detected but it remains unclear if it is an actual orthologue of the enzyme.

with only two clear exceptions concerning the bacterial phylum Chloroflexi and the bacteroidete *Flavobacterium johnsoniae*, which branched among the archaea, supporting ancient HGT events from this domain. The HMGS phylogenetic tree also allowed the distinction of the mitochondrial and cytosolic paralogues described in metazoa (Ayte et al. 1990) and showed the intriguing very long branch basal to most archaeal HMGS sequences discussed by Boucher et al. (2004) and Jiang et al. (2008). The two following enzymes of the pathway, HMGR (fig. 3) and MVK (fig. 4), also yielded phylogenetic trees with a good global agreement with the accepted phylogeny of organisms. Notably, except for a few clear cases of HGT like the ones commented above, all these trees retrieved a well-supported separation of the eukaryotic, archaeal, and bacterial domains.

The last three enzymes of the classical MVA pathway, PMK, MDC, and IDI 1 (IDI1), did not follow this "three domain" distribution because they are absent from most archaeal species, as noticed in previous works (Smit and Mushegian 2000; Boucher et al. 2004). Our sequence similarity searches (table 1) confirmed their absence in a larger archaeal taxonomic spectrum. The exceptions are Haloarchaea and Thermoplasmatales, which have MDC sequences, and Haloarchaea and Thaumarchaeota, which contain homologues of IDI1. However, these archaeal homologues of MDC and IDI1 branched among the bacteria, suggesting HGT events (Boucher et al. 2004) (supplementary figs. 3 and 4, Supplementary Material online). In contrast, the class Sulfolobales (Crenarchaeota) contains both PMK and MDC homologues branching with a strong support in an intermediate position between the eukaryotic and the bacterial sequences (table 1, supplementary figs. 2 and 3, Supplementary Material online). This position makes it difficult to consider them as the result of an HGT, neither from bacteria nor from eukaryotes, as it was proposed by Boucher et al. (2004). On the contrary, the intermediate position showed by these sequences

FIG. 2 Bayesian phylogenetic tree of the HMGS. This tree was reconstructed using 99 representative sequences and 311 conserved sites. Numbers at nodes are posterior probabilities. Maximum likelihood bootstrap values are also provided for nodes concerning the monophyly of major clades.

opens the possibility for an ancestral presence of PMK and MDC in archaea, subsequently conserved only in the Sulfolobales.

As mentioned above, two enzymes have been proposed to participate in an alternative pathway of conversion of phosphomevalonate to IPP and DMAPP in archaea (Barkley et al. 2004; Grochowski et al. 2006; Dutoit et al. 2008): the IPK and the IDI 2 (IDI2). We observe that archaea containing other genes of the MVA pathway also have orthologues of IPK and IDI2, except for *Sulfolobus tokodaii*, *Cenarchaeum*

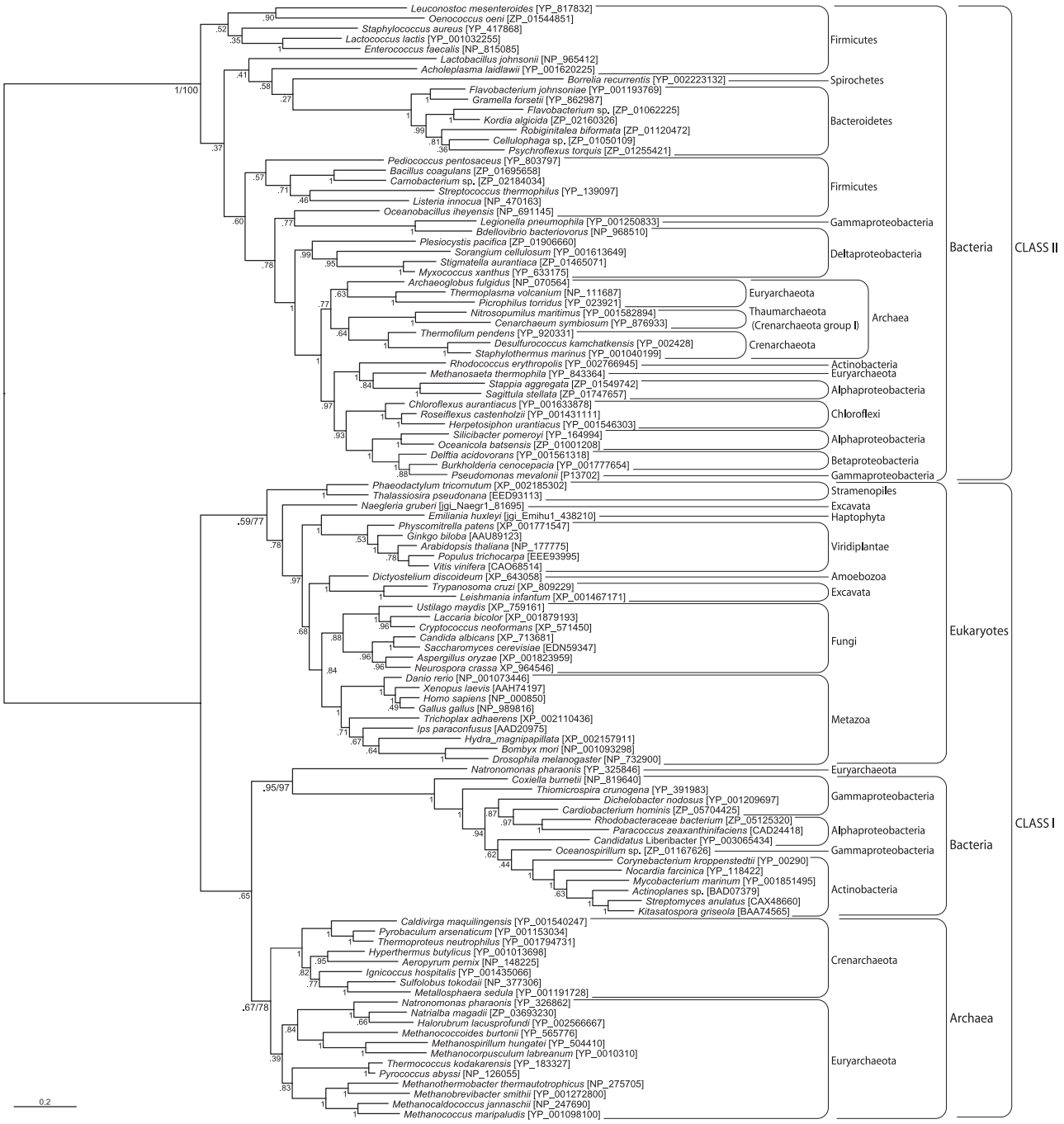

FIG. 3 Bayesian phylogenetic tree of the HMGR. This tree was reconstructed using 110 representative sequences and 274 conserved sites. Numbers at nodes are posterior probabilities. ML bootstrap values are also provided for nodes concerning the monophyly of major clades.

symbiosum, *Nitrosopumilus maritimus* and some halophilic archaea. Interestingly, these archaeal genomes lacking IDI2 have IDI1 instead, suggesting nonhomologous replacement of IDI2. The phylogeny of IDI2 supported the separation between archaeal and bacterial sequences, except for the Firmicutes, which branched among archaea, indicating a likely HGT event from this domain (supplementary fig. 5, Supplementary Material online). With the exception of 14 eukaryotes and 9 bacteria, homologues of the archaeal IPK could not be detected in the genomes of most species from the other domains of life (not shown). IPK sequences were less conserved than those of the other MVA pathway enzymes, in particular in the case of the bacterial and eukaryotic homologues. As a result, phylogenetic trees for this enzyme were poorly supported (low bootstrap values and posterior probabilities) and less congruent with the presumed organismal phylogeny. Removal of the very divergent bacterial and eukaryotic homologues led to better alignments and more supported trees, where most of the different archaeal classes were monophyletic (supplementary fig. 6, Supplementary Material online). These results are consistent with the existence of a modified MVA pathway in archaea (Grochowski

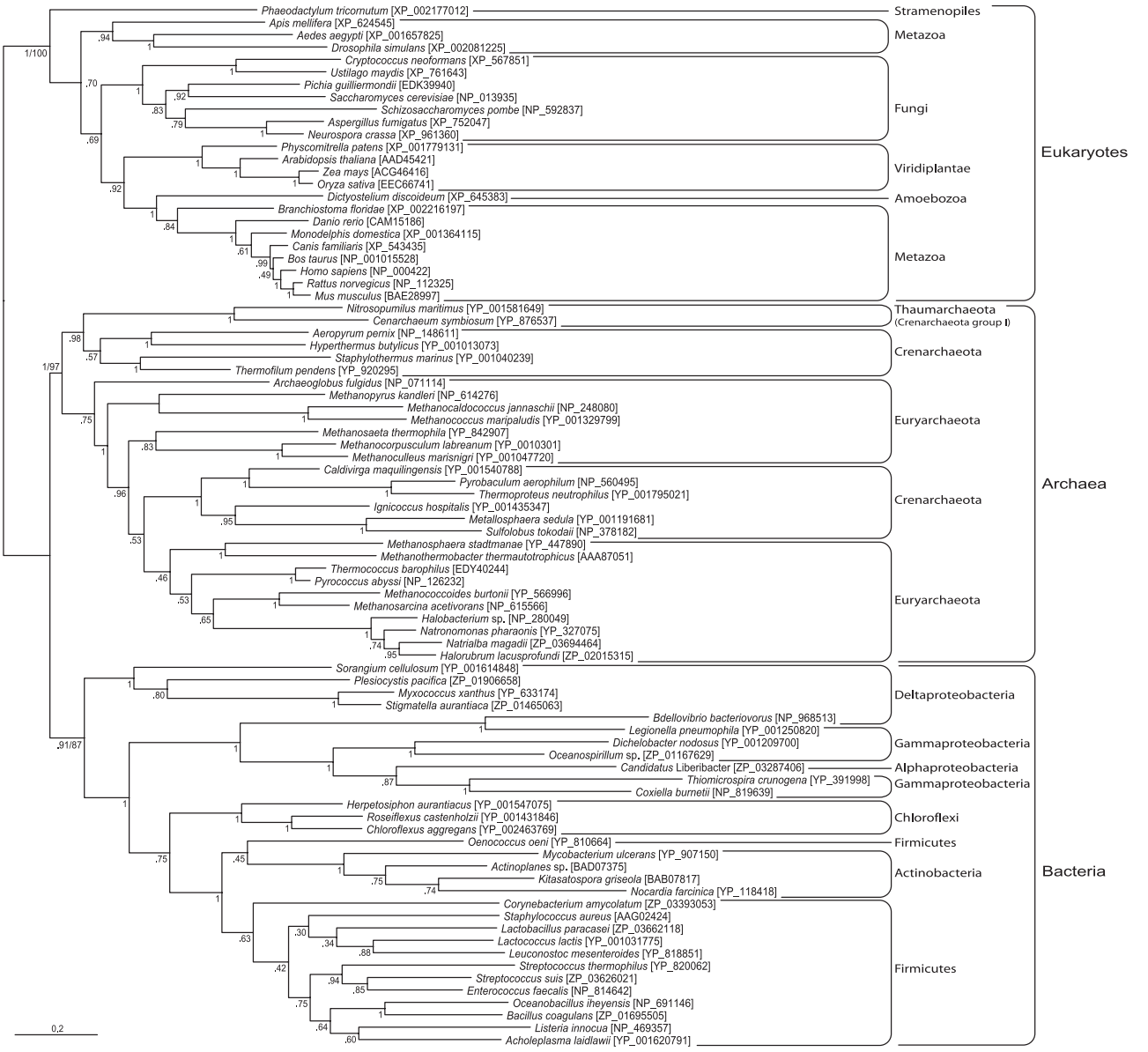

Fig. 4 Bayesian phylogenetic tree of the mevalonate kinase (MVK). This tree was reconstructed using 84 representative sequences and 232 conserved sites. Numbers at nodes are posterior probabilities. Maximum likelihood bootstrap values are also provided for nodes concerning the monophyly of major clades.

et al. 2006; Chen and Poulter 2010), and, even if the phosphomevalonate decarboxylase required to complete this metabolic route remains to be identified, our results support the alternative pathway to be extensive and characteristic of the whole archaeal domain. This suggests that the alternative pathway, including the recently characterized enzymes IPK and IDI2, was ancestral to archaea. This has implications on the classical bipartite vision of the distribution of isoprenoid biosynthesis pathways in the three domains of life. Actually, despite several sporadic cases of HGT, three ways to produce IPP and DMAPP would exist, each of which is widespread within a domain of life: the MEP pathway in bacteria, the classical MVA pathway in eukaryotes, and the alternative MVA pathway in archaea (fig. 1). The characterization of the missing archaeal phosphomevalonate decarboxylase, including its distribution and phylogeny, will be a major help to understand the differentiation of those three pathways in the tree of life. This evolutionary history may have been rather complex because our results concerning the Sulfolobales raise the possibility that the eukaryotic-type MVA pathway was present in ancestral archaea before its replacement in most archaeal groups by the archaeal-type pathway.

An Ancestral MVA Pathway in Bacteria

As previously mentioned, even if it is generally assumed that bacteria do not possess the MVA pathway, this metabolic route has already been described in some Firmicutes (Bochar et al. 1999; Wilding et al. 2000; Romanowski et al. 2002; Voynova et al. 2004; Steussy et al. 2006). It has been recently shown that the MVA pathway may complement defects in the synthesis of IPP involved in iron–sulfur

Table 2. Bacteria Bearing Enzymes of the MVA Pathway.

		HMGS	HMGR	MVK	PMK	MDC	IPK
Actinobacteria	*Corynebacterium amycolatum*	+	I	+	+	+	−
	Gardnerella vaginalis	+	II	+	+	+	−
	Mycobacterium marinum	+	I	+	+	+	−
	Nocardia farcinica	+	I	+	+	+	−
	Streptomyces griseoflavus	+	I	+	+	+	−
Bacteroidetes	*Croceibacter atlanticus*	−	II	−	−	+	−
	Dokdonia donghaensis	−	II	−	−	+	−
	Flavobacteria bacterium	+	II	−	−	+	−
	Flavobacterium johnsoniae	+	II	−	−	+	−
	Gramella forsetii	−	II	−	−	+	−
	Kordia algicida	+	II	−	−	+	−
	Leeuwenhoekiella blandensis	−	II	−	−	+	−
	Polaribacter irgensii	−	II	−	−	+	−
	Psychroflexus torques	−	II	−	−	+	−
	Robiginitalea biformata	−	II	−	−	+	−
	Unidentified eubacterium	−	II	−	−	+	−
Firmicutes	*Bacillus coagulans*	+	II	+	+	+	−
	Gemella haemolysans	+	II	+	+	+	−
	Listeria grayi	+	II	+	+	+	−
	Oceanobacillus iheyensis	+	II	+	+	+	−
	Staphylococcus capitis	+	II	+	+	+	−
	Catonella morbi	+	II	+	+	+	−
	Carnobacterium **sp.**	+	II	+	+	+	−
	Enterococcus faecalis	+	II	+	+	+	−
	Lactobacillus johnsonii	+	II	+	+	+	−
	Lactococcus lactis	+	II	+	+	+	−
	Leuconostoc citreum	+	II	+	+	+	−
	Oenococcus oeni	+	II	+	+	+	−
	Pediococcus pentosaceus	+	II	+	+	+	−
	Streptococcus agalactiae	+	II	+	+	+	−
	Acholeplasma laidlawii	+	II	+	+	+	−
Chloroflexi	*Chloroflexus aggregans*	+	II	+	−	+	+
	Herpetosiphon aurantiacus	+	II	+	−	+	+
	Roseiflexus castenholzii	+	II	+	−	+	+
Proteobacteria	**Rhodobacteraceae bacterium**	+	I	+	+	+	−
	Candidatus **Liberibacter**	+	I	+	+	+	−
	Bdellovibrio bacteriovorus	−	II	+	+	+	−
	Myxococcus xanthus	+	II	+	+	+	−
	Plesiocystis pacifica	+	II	+	−	+	−
	Sorangium cellulosum	+	II	+	+	+	−
	Stigmatella aurantiaca	+	II	+	+	+	−
	Coxiella burnetii	+	I	+	+	+	−
	Dichelobacter nodosus	+	I	+	+	+	−
	Thiomicrospira crunogena	+	I	+	+	+	−
Spirochetes	*Borrelia burgdorferi*	+	II	+	?	+	−

This table summarizes the 93 species where homologous sequences of MVA pathway enzymes were systematically detected in 43 bacterial representatives of their respective genera. (+) A homologue of the enzyme can be detected; (−) no detection of homologous sequences; (I) class I HMGR; (II) class II HMGR; (?) a homologue can be detected but it remains unclear if it is an actual orthologue of the enzyme.

protein biogenesis in certain bacteria (Vinella et al. 2009), but the major role of this pathway in Firmicutes is related to isoprenoid biosynthesis (Wilding et al. 2000; Voynova et al. 2004). Contrary to previous studies where the presence of this pathway seemed to be limited to a very reduced range of bacterial taxa, especially Firmicutes (Lange et al. 2000; Boucher et al. 2004), the current availability of a larger sampling of complete genome sequences has allowed us to find that its occurrence appears far from being restricted to this bacterial phylum. In fact, we found homologues of the MVA pathway enzymes in 93 bacterial species belonging to 43 different genera of 6 phyla (table 2). In particular, Firmicutes, Actinobacteria, Proteobacteria, and the Spirochete genus *Borrelia* encode in their genomes most of the enzymes of this pathway. In addition, Bacteroidetes possess only some of the enzymes, and the Chloroflexi contain IPK instead of PMK. The case of the HMGR will be discussed below, but we can highlight the existence of two different classes of enzymes, one present in some Proteobacteria and Actinobacteria and the other widespread in the six MVA pathway–bearing bacterial phyla.

The distribution of the isopentenyl isomerases IDI1 and IDI2 is much more complex. The reaction carried out by the IDI enzymes is necessary in the MVA pathway to produce DMAPP from IPP, but it is not compulsory in the MEP pathway because the last enzyme of this pathway (HMDR—4-hydroxy-3-methylbut-2-enyl diphosphate reductase) catalyzes the production of both IPP and DMAPP

at the same time (Rohdich et al. 2001, 2002). However, IDI genes are present in the genomes of a variety of bacteria, probably because they may allow these organisms to isomerize IPP and DMAPP according to their metabolic requirements at a given moment. Thus, although IDI1 and IDI2 are characteristic of eukaryotes and archaea, respectively, the distribution of these enzymes among bacteria is much less clear-cut. Bacteroidetes, most actinobacteria, and some proteobacterial species have IDI1, whereas Firmicutes, Cyanobacteria, Chlorobi, Chloroflexi, *Deinococcus–Thermus*, Spirochetes, Deltaproteobacteria, and some other proteobacterial and actinobacterial species contain IDI2 (supplementary figs. 4 and 5, Supplementary Material online).

One of the most intriguing phylogenies is the one of HMGR. There are two different homologous classes of this enzyme (Scher and Rodwell 1989; Bochar et al. 1999; Friesen and Rodwell 2004; Hedl et al. 2004). The class I HMGR is present in eukaryotes, archaea, and several proteobacteria and actinobacteria, whereas the class II is characteristic of most bacteria and a small number of archaea. Biochemical studies of HMGR showed that both classes share similar active sites and form homodimers. However, the respective function of each monomer is different from one class to the other, the two classes do not use exactly the same cofactors, and they are differently inhibited by statins. Because class II HMGR is widespread and characteristic of many bacterial phyla, it may be considered as ancestral to the bacterial domain. The clade of class II HMGR also contains a few archaeal representatives, which were proposed to have received their genes by HGT from bacteria (Boucher et al. 2001). With respect to class I bacterial sequences, they clearly branch within the class I clade despite their divergence (fig. 3), which was interpreted as a complex history of HGTs and specialization leading these sequences to be present in several MVA pathway–lacking organisms (Gophna et al. 2006). Nevertheless, these bacterial sequences did not branch within the eukaryotic or archaeal clades, as could be expected if they had been acquired by recent HGT from one of these domains. Intriguingly, a second HMGR copy found in the haloarchaeon *Natronomonas pharaonis* branched sister to this bacterial clade far from the other haloarchaeal sequences (which include a second copy of this same species).

In summary, the phylogenetic analyses of the eukaryotic-like MVA pathway enzymes in a large taxonomic sampling produced topologies supporting the monophyly of major groups (figs. 2–4 and supplementary figs. 2–5, Supplementary Material online). In particular, this includes the emergence of the bacterial sequences as a monophyletic group distinct from archaea and eukaryotes (i.e., the three domains topology). In fact, for each enzyme, the vast majority of bacterial sequences form an independent monophyletic group, as, for example, in the phylogenies of HMGS (fig. 2) and MVK (fig. 4). If, as mentioned above, bacteria had obtained their MVA pathway enzymes by HGT as traditionally assumed (Boucher and Doolittle 2000; Wilding et al. 2000), all the bacterial sequences should branch among those of another domain of life and share similar biochemical properties with their donors. On the contrary, most bacterial sequences for each enzyme form monophyletic groups separated from the archaeal and eukaryotic clades, and, when well characterized biochemically, they have their own sequence signatures and biochemical characteristics. Our phylogenetic analyses only revealed the expected pattern of HGTs in the case of the HMGS of Chloroflexi and *F. johnsoniae* and the IDI2 of Firmicutes branching within the archaeal clade (fig. 2 and supplementary fig. 5, Supplementary Material online), which supports that these bacteria obtained at least part of their MVA pathway from archaea. Conversely, we also observed cases of HGT from bacteria to archaea, for example, the class II HMGR of certain euryarchaeota and thaumarchaeota, which emerged well nested within the bacterial clade (fig. 3). Phylogenetic trees constructed using only the bacterial sequences yielded topologies supporting the monophyly of several bacterial phyla, which argued against the hypothesis of multiple HGT events among distantly related bacteria (data not shown).

Three different scenarios can be suggested to explain the segregation of bacterial sequences from those of archaea and eukaryotes observed in the phylogenetic trees of the MVA pathway enzymes. The first one would entail that an ancient bacterial lineage obtained all the enzymes by HGT from another domain of life, archaea or eukaryotes, followed by a period of fast evolution that would be responsible of the high divergence of the bacterial enzymes from the original ones. Then, new HGT events would have occurred independently from this bacterial lineage toward all the current MVA pathway–bearing bacterial classes (fig. 5*A*). The other two possibilities would imply that the MVA pathway was ancestral to bacteria, which would be the consequence either of the acquisition by an ancestral bacterium of the pathway from another domain before the radiation of different bacterial phyla (fig. 5*B*) or of the direct vertical inheritance of an MVA pathway already present in the last common ancestor of all organisms, the cenancestor (fig. 5*C*). These two latter hypotheses imply a massive loss of the MVA pathway in many bacterial lineages, but this can easily be explained by the functional redundancy with the MEP pathway, which is present in most bacteria.

Conclusions: The Early Origin of the MVA Pathway

The distribution of the different isoprenoid biosynthesis pathways in the three domains of life is very complex but displays a clear preferential relationship: the MEP pathway in bacteria, the classical MVA pathway in eukaryotes, and the alternative MVA pathway in archaea. Our work supports that each route is ancestral to one of the domains, but it is not enough to decide whether these pathways were present in the cenancestor or not. If we exclude the extremely improbable possibility that the pathway was assembled in the three domains by independent recruitment of homologous enzymes carrying out other functions, three other possibilities concerning this question can

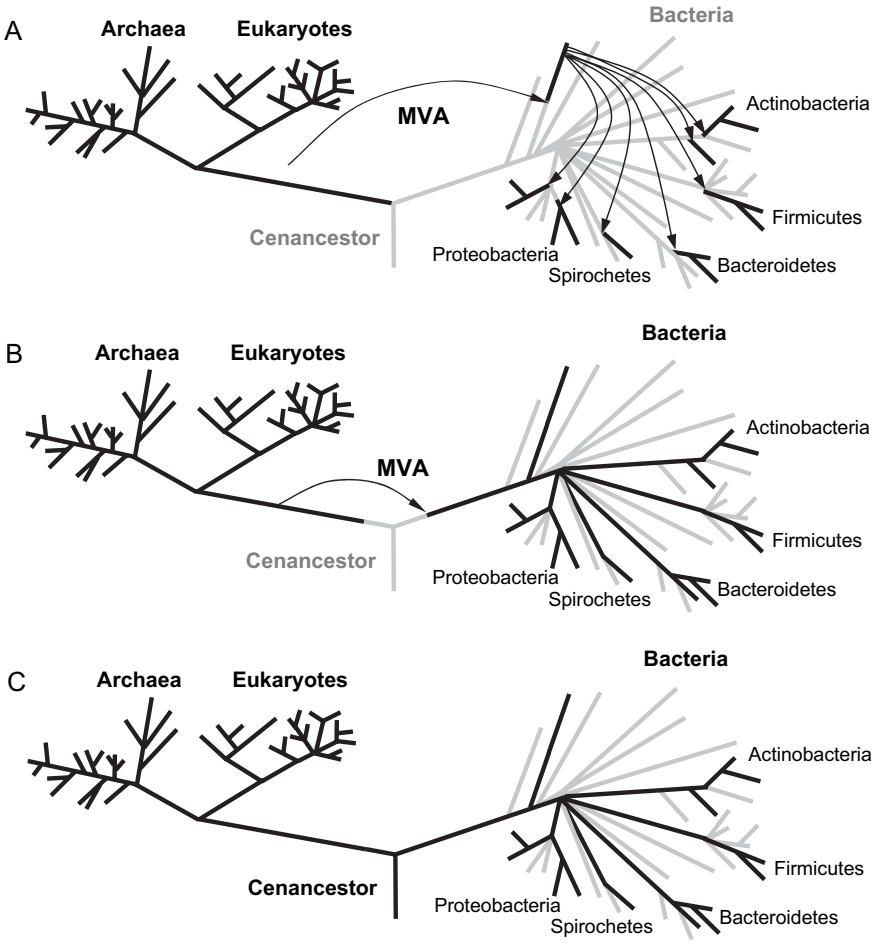

Fig. 5 Hypotheses for the origin of the MVA pathway of isoprenoid synthesis in bacteria. (*A*) Transfer of the complete MVA pathway from eukaryotic or archaeal donors to a bacterial lineage followed by multiple transfers from this lineage to all the bacterial groups currently possessing the MVA pathway (black branches). (*B*) Ancient transfer of the complete MVA pathway to a bacterial ancestor before the divergence of the contemporary bacterial phyla. (*C*) Inheritance of the MVA pathway from the cenancestor. Scenarios (*B* and *C*) entail multiple losses of the MVA pathway in several bacterial phyla (gray branches). For convenience, the classical bacterial rooting of the tree of life has been adopted.

be proposed. The first one would be the lack of any isoprenoid biosynthesis pathway in the cenancestor (fig. 6*A*), an idea that has already been advanced and used by some authors as an argument to discard the presence of lipid membranes in the cenancestor (Martin and Russell 2003). This is, however, at odds with the present-day ubiquitous presence of isoprenoids in the three domains of life. The mirror hypothesis would be the presence of ancestral forms of the three pathways in the cenancestor followed by selective losses in each domain of life (fig. 6*B*). Nevertheless, the MEP pathway is completely absent in archaea and in all non–plastid-bearing eukaryotes, so we can assume that it was not present in the ancestors of these two domains.

In the light of our results, a more parsimonious third hypothesis emerges: the classical MVA pathway would have been present in the cenancestor and inherited not only by archaea and eukaryotes but also by bacteria, which would have replaced it by the MEP pathway in a variety of bacterial phyla (fig. 6*C*). As we said previously, it is even possible that the last common ancestor of archaea also had the entire classical MVA pathway, including the last enzymes (PMK and MDC) that have been lost in the vast majority of archaeal species, but not in Sulfolobales. Thus, we can speculate that the classical MVA pathway was present in the cenancestor and replaced or modified in most bacteria and archaea by their respective current characteristic pathways. If this hypothesis turned out to be correct, the MVA pathway would represent a rare case where a metabolic pathway would have retained more ancestral characteristics in eukaryotes than in bacteria and most archaea (the Sulfolobales being an exception among the available archaeal genomes, as mentioned above).

The possibility of an isoprenoid biosynthesis pathway operating in the cenancestor has interesting implications on the open question about the presence and nature of the membrane of this ancestral organism. Indeed, the differences between the membranes of archaea, on the one hand, and those shared by eukaryotes and bacteria, on the other hand, have led to a hot controversy on the nature of the cenancestral membrane. The major differences are the presence of fatty acid lateral chains bound by ester links to *sn*-glycerol-3-phosphate (G3P) in eukaryotes and bacteria, in contrast with archaeal membranes containing isoprenoid lateral chains and *sn*-glycerol-1-phosphate (G1P)

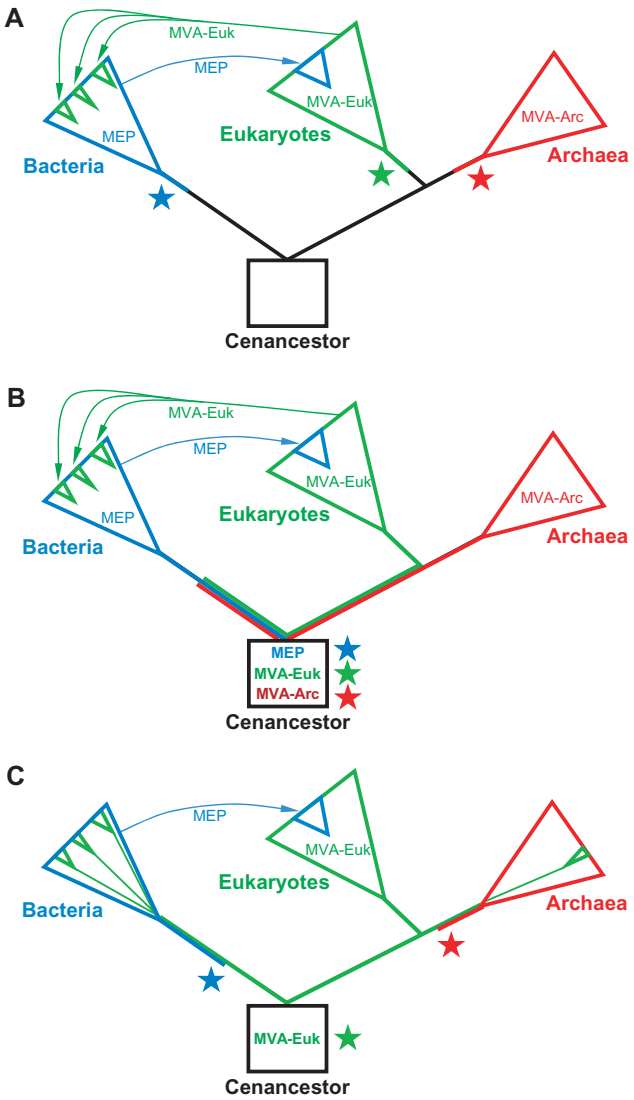

Fig. 6 Hypotheses for the evolution of the different isoprenoid biosynthesis pathways on a schematic three domains of life tree. The MEP pathway is indicated in blue, and the eukaryotic and archaeal MVA (MVA-Euk and MVA-Arc) pathways are indicated in green and red, respectively. Stars filled with these colors indicate the origin of the respective pathways. (*A*) Each pathway would have appeared in the ancestors of each domain of life. Late HGT events would explain the presence of MEP and MVA-Euk pathways in several eukaryotes and bacteria, respectively. (*B*) The three pathways would have been present in the cenancestor, but only one pathway would have been conserved in each domain of life. Presence of supplementary pathways would be explained by HGT acquisitions. (*C*) The classical MVA pathway would have been present in the cenancestor and replaced in most bacteria and archaea by the MEP and MVA-Arc pathways, respectively. This figure shows the classical rooting for convenience, but the hypothesis remains valid if eukaryotes are considered as derived from a symbiotic event between archaea and bacteria.

bound by ether links. Although there are exceptions concerning the nature of the lateral chains (fatty acids or isoprenoids) and the type of link (ester or ether), the glycerol-phosphate stereoisomer is always different between archaea (G1P) and bacteria and eukaryotes (G3P). This, together with the apparent nonhomologous biosynthesis of isoprenoids in archaea and bacteria, has led some authors to propose either that the cenancestor was acellular, namely, that it was not bounded by any type of membrane (Koga et al. 1998), or that it was surrounded by mineral membranes instead of lipid ones (Martin and Russell 2003; Mulkidjanian et al. 2009). Nevertheless, fatty acids have been shown to be present in archaea in variable proportions (Gattinger et al. 2002) and homologues of the bacterial and eukaryotic genes encoding the enzymes involved in fatty acid biosynthesis and degradation are present in most archaeal genome sequences, suggesting that they are ancestral (Peretó et al. 2004). In addition, Peretó et al. (2004) also showed that the dehydrogenases involved in the synthesis of G1P and G3P in archaea and bacteria evolved from universally spread enzymatic families that were most likely present in the cenancestor. The different glycerol stereoisomers would have been selected afterward in the different lineages. Our results, supporting the possibility of an ancestral MVA pathway, provide an additional argument for the hypothesis that all the major components of contemporary membranes (phospholipids composed of glycerol-phosphate bound to lateral chains of fatty acids and/or isoprenoids) existed in the cenancestor and that the specialization observed in current cellular membranes is most likely secondary and linked to the divergence of the three domains of life.

Supplementary Material

Supplementary table 1 and figures 1–6 are available at *Molecular Biology and Evolution* online http://mbe.oxfordjournals.org/.

Acknowledgments

This work was supported by the interdisciplinary programs Origine des Planètes et de la Vie and InTerrVie (Interactions Terre—Vie) of the French Centre National de la Recherche Scientifique and Institut National des Sciences de l'Univers. J.L. is recipient of a PhD fellowship of the French Research Ministry. We thank Puri López-García for critical reading of the manuscript and three anonymous reviewers for helpful comments.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. 215:403–410.

Ayte J, Gil-Gomez G, Haro D, Marrero PF, Hegardt FG. 1990. Rat mitochondrial and cytosolic 3-hydroxy-3-methylglutaryl-CoA synthases are encoded by two different genes. *Proc Natl Acad Sci U S A*. 87:3874–3878.

Barkley SJ, Cornish RM, Poulter CD. 2004. Identification of an archaeal type II isopentenyl diphosphate isomerase in *Methanothermobacter thermautotrophicus*. *J Bacteriol*. 186:1811–1817.

Bochar DA, Stauffacher CV, Rodwell VW. 1999. Sequence comparisons reveal two classes of 3-hydroxy-3-methylglutaryl coenzyme A reductase. *Mol Genet Metab*. 66:122–127.

Boucher Y, Doolittle WF. 2000. The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Mol Microbiol*. 37:703–716.

Boucher Y, Huber H, L'Haridon S, Stetter KO, Doolittle WF. 2001. Bacterial origin for the isoprenoid biosynthesis enzyme HMG-CoA reductase of the archaeal orders Thermoplasmatales and Archaeoglobales. *Mol Biol Evol*. 18:1378–1388.

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

Brinkman FS, Blanchard JL, Cherkasov A, et al. (14 co-authors). 2002. Evidence that plant-like genes in *Chlamydia* species reflect an ancestral relationship between Chlamydiaceae, cyanobacteria, and the chloroplast. *Genome Res*. 12:1159–1167.

Chen M, Poulter CD. 2010. Characterization of thermophilic archaeal isopentenyl phosphate kinases. *Biochemistry* 49:207–217.

Dutoit R, de Ruyck J, Durisotti V, Legrain C, Jacobs E, Wouters J. 2008. Overexpression, physicochemical characterization, and modeling of a hyperthermophilic *Pyrococcus furiosus* type 2 IPP isomerase. *Proteins* 71:1699–1707.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32:1792–1797.

Friesen JA, Rodwell VW. 2004. The 3-hydroxy-3-methylglutaryl coenzyme-A (HMG-CoA) reductases. *Genome Biol*. 5:248.

Gattinger A, Schloter M, Munch JC. 2002. Phospholipid etherlipid and phospholipid fatty acid fingerprints in selected euryarchaeotal monocultures for taxonomic profiling. *FEMS Microbiol Lett*. 213:133–139.

Gophna U, Thompson JR, Boucher Y, Doolittle WF. 2006. Complex histories of genes encoding 3-hydroxy-3-methylglutaryl-coenzymea reductase. *Mol Biol Evol*. 23:168–178.

Grochowski LL, Xu H, White RH. 2006. *Methanocaldococcus jannaschii* uses a modified mevalonate pathway for biosynthesis of isopentenyl diphosphate. *J Bacteriol*. 188:3192–3198.

Hedl M, Tabernero L, Stauffacher CV, Rodwell VW. 2004. Class II 3-hydroxy-3-methylglutaryl coenzyme A reductases. *J Bacteriol*. 186:1927–1932.

Herdendorf TJ, Miziorko HM. 2006. Phosphomevalonate kinase: functional investigation of the recombinant human enzyme. *Biochemistry* 45:3235–3242.

Houten SM, Waterham HR. 2001. Nonorthologous gene displacement of phosphomevalonate kinase. *Mol Genet Metab*. 72:273–276.

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

Jiang C, Kim SY, Suh DY. 2008. Divergent evolution of the thiolase superfamily and chalcone synthase family. *Mol Phylogenet Evol*. 49:691–701.

Jobb G, von Haeseler A, Strimmer K. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol*. 4:18.

Kaneda K, Kuzuyama T, Takagi M, Hayakawa Y, Seto H. 2001. An unusual isopentenyl diphosphate isomerase found in the mevalonate pathway gene cluster from *Streptomyces* sp. strain CL190. *Proc Natl Acad Sci U S A*. 98:932–937.

Koga Y, Kyuragi T, Nishihara M, Sone N. 1998. Did archaeal and bacterial cells arise independently from noncellular precursors? A hypothesis stating that the advent of membrane phospholipid with enantiomeric glycerophosphate backbones caused the separation of the two lines of descent. *J Mol Evol*. 46:54–63.

Kuzuyama T. 2002. Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. *Biosci Biotechnol Biochem*. 66:1619–1627.

Lange BM, Rujan T, Martin W, Croteau R. 2000. Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc Natl Acad Sci U S A*. 97:13172–13177.

Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol Biol Evol*. 25:1307–1320.

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

Matsuzaki M, Kuroiwa H, Kuroiwa T, Kita K, Nozaki H. 2008. A cryptic algal group unveiled: a plastid biosynthesis pathway in the oyster parasite *Perkinsus marinus*. *Mol Biol Evol*. 25:1167–1179.

McGarvey DJ, Croteau R. 1995. Terpenoid metabolism. *Plant Cell*. 7:1015–1026.

Moustafa A, Reyes-Prieto A, Bhattacharya D. 2008. Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. *PLoS ONE*. 3:e2205.

Mulkidjanian AY, Galperin MY, Koonin EV. 2009. Co-evolution of primordial membranes and membrane proteins. *Trends Biochem Sci*. 34:206–215.

Peretó J, López-García P, Moreira D. 2004. Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem Sci*. 29:469–477.

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

Philippe H. 1993. MUST, a computer package of Management Utilities for Sequences and Trees. *Nucleic Acids Res*. 21:5264–5272.

Phillips MA, Leon P, Boronat A, Rodríguez-Concepción M. 2008. The plastidial MEP pathway: unified nomenclature and resources. *Trends Plant Sci*. 13:619–623.

Rambaut A, Drummond AJ. 2003. Tracer: a program for analysisng results from Bayesian MCMC programs such as BEAST & MrBayes. Distributed by the authors. Edinburgh (United Kingdom) Institute of Evolutionary Biology, University of Edinburgh. Available from: http://evolve.zoo.ox.ac.uk/software.html?id=tracer.

Rodríguez-Concepción M, Boronat A. 2002. Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. *Plant Physiol*. 130:1079–1089.

Rohdich F, Hecht S, Gartner K, Adam P, Krieger C, Amslinger S, Arigoni D, Bacher A, Eisenreich W. 2002. Studies on the nonmevalonate terpene biosynthetic pathway: metabolic role of IspH (LytB) protein. *Proc Natl Acad Sci U S A*. 99:1158–1163.

Rohdich F, Kis K, Bacher A, Eisenreich W. 2001. The non-mevalonate pathway of isoprenoids: genes, enzymes and intermediates. *Curr Opin Chem Biol*. 5:535–540.

Romanowski MJ, Bonanno JB, Burley SK. 2002. Crystal structure of the *Streptococcus pneumoniae* phosphomevalonate kinase, a member of the GHMP kinase superfamily. *Proteins* 47:568–571.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.

Ruzicka L. 1953. The isoprene rule and the biogenesis of terpenic compounds. *Experientia* 9:357–367.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 4:406–425.

Scher DS, Rodwell VW. 1989. 3-Hydroxy-3-methylglutaryl coenzyme A lyase from *Pseudomonas mevalonii*. *Biochim Biophys Acta*. 1003:321–326.

Smit A, Mushegian A. 2000. Biosynthesis of isoprenoids via mevalonate in Archaea: the lost pathway. *Genome Res*. 10:1468–1484.

Steussy CN, Robison AD, Tetrick AM, Knight JT, Rodwell VW, Stauffacher CV, Sutherlin AL. 2006. A structural limitation on enzyme activity: the case of HMG-CoA synthase. *Biochemistry* 45:14407–14414.

Vinella D, Brochier-Armanet C, Loiseau L, Talla E, Barras F. 2009. Iron-sulfur (Fe/S) protein biogenesis: phylogenomic and genetic studies of A-type carriers. *PLoS Genet.* 5:e1000497.

Voynova NE, Rios SE, Miziorko HM. 2004. *Staphylococcus aureus* mevalonate kinase: isolation and characterization of an enzyme of the isoprenoid biosynthetic pathway. *J Bacteriol.* 186:61–67.

Wilding EI, Brown JR, Bryant AP, Chalker AF, Holmes DJ, Ingraham KA, Iordanescu S, So CY, Rosenberg M, Gwynn MN. 2000. Identification, evolution, and essentiality of the mevalonate pathway for isopentenyl diphosphate biosynthesis in gram-positive cocci. *J Bacteriol.* 182:4319–4327.

Zhou D, White RH. 1991. Early steps of isoprenoid biosynthesis in *Escherichia coli*. *Biochem J.* 273:627–634.