Mosaic Evolution of the Phosphopantothenate Biosynthesis Pathway in Bacteria and Archaea

Luc Thomès and Alain Lescure *

Architecture et Réactivité de l'ARN, CNRS, UPR9002, Université de Strasbourg, France

*Corresponding author: E-mail: a.lescure@cnrs-ibmc.unistra.fr

Accepted: 11 December 2020

Abstract

Phosphopantothenate is a precursor to synthesis of coenzyme A, a molecule essential to many metabolic pathways. Organisms of the archaeal phyla were shown to utilize a different phosphopantothenate biosynthetic pathway from the eukaryotic and bacterial one. In this study, we report that symbiotic bacteria from the group *Candidatus poribacteria* present enzymes of the archaeal pathway, namely pantoate kinase and phosphopantothenate synthetase, mirroring what was demonstrated for *Picrophilus torridus*, an archaea partially utilizing the bacterial pathway. Our results not only support the ancient origin of the coenzyme A pathway in the three domains of life but also highlight its complex and dynamic evolution. Importantly, this study helps to improve protein annotation for this pathway in the *C. poribacteria* group and other related organisms.

Key words: phosphopantothenate pathway, coenzyme A, *Candidatus poribacteria*.

Significance

Here, we describe the unusual presence of an archaeal pathway for biosynthesis of phosphopantothenate, a precursor of coenzyme A, in a group of unclassified bacteria, *Candidatus poribacteria*. This observation was unexpected, since the use of the two alternative pathways has been proposed as a phylogenetic marker to distinguish between bacteria and archaea. Moreover, the presence of the archaea enzymes appeared to be limited to the subgroup of poribacteria living in symbiosis with sponges, as free-living poribacteria were shown to use the classical bacterial pathway, suggesting that the utilization of the two metabolic routes might be part of the sponge/bacteria interaction.

Introduction

Coenzyme A (CoA) is a ubiquitous molecule participating in many biosynthetic pathways, and more than 400 enzyme-catalyzed reactions are known to involve acyl-CoA as substrate. In most bacteria and eukaryotes, synthesis of one of the first intermediates in this pathway, phosphopantothenate (PP), is achieved in a two-step reaction: synthesis of pantothenate by condensation of pantoate with β-alanine, followed by pantothenate phosphorylation. Interestingly, it was shown that archaea utilize an alternative pathway, where the two consecutive reactions are exchanged, with the phosphorylation step occurring first, followed by addition of β-alanine (see fig. 1) (Yokooji et al. 2009; Ishibashi et al. 2012; Tomita et al. 2012; Katoh et al. 2013). This difference was proposed as an intrinsic characteristic that distinguishes bacterial and archaeal phyla. In comparative genome analyses of a group of bacteria, *Candidatus poribacteria*, we made the striking observation that the enzymes corresponding to the PP synthesis pathway were not correctly annotated. *Candidatus poribacteria* refers to an unclassified group of marine bacteria, evolutionarily related to the superphylum *Planctomycetes–Verrucomicrobia–Chlamydia* (Fieseler et al. 2004; Kamke et al. 2014). These bacteria were originally identified as members of the bacterial community living in symbiosis with diverse sponge species, including *Aplysina aerophoba*. *Candidatus poribacteria* present the peculiarity of sharing several eukaryotic-like features,

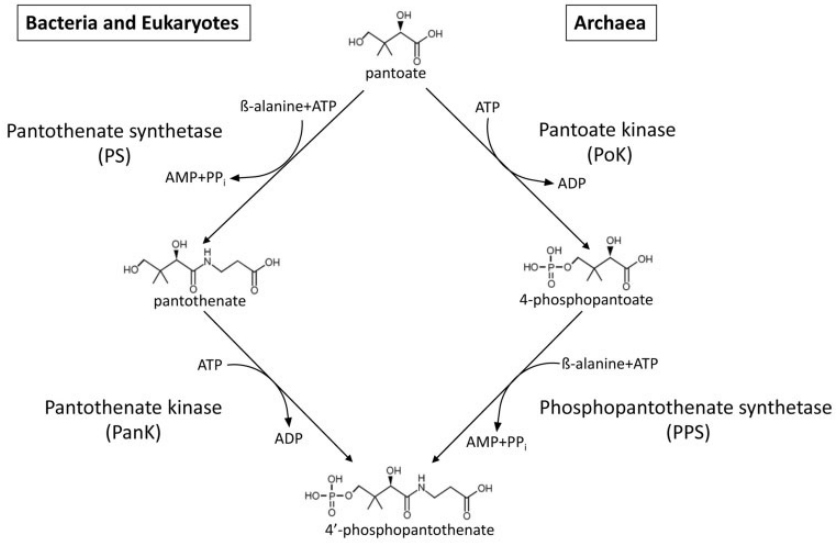

Fig. 1.—Bacterial and archaeal phosphopantothenate biosynthetic pathways. Most bacteria, like eukaryotes, use PS and PanK to synthetize 4′-phosphopantothenate from pantoate. The alternative pathway utilized by most archaea involve PoK and PPS enzymes that catalyze similar reactions but in the reverse reaction order.

such as complex inner membrane structures similar to eukaryotic intracellular compartments and a nucleoid-like structure. Interestingly, a recent metagenomic study identified additional strains of the *C. poribacteria* group living as free-living organisms present in seawater, defining two distinct subgroups characterized according to their lifestyle, and designated Entoporibacteria for the sponge-associated and Pelagiporibacteria for the free-living ones. Genomic analyses revealed a high level of inner divergence between the two groups, indicating a different evolutionary history (Podell et al. 2019). Ontological analysis of the gene sets specific to each subgroup predicted that a large part of the genes specific to the Entoporibacteria group contribute to the host–symbiont interaction.

Results

In bacteria, pantothenate synthetase (PS) is an enzyme responsible for condensation of *β*-alanine and D-pantoate resulting in D-pantothenate. Subsequently, pantothenate kinase (PanK) phosphorylates D-pantothenate to D-4′-phosphopantothenate (fig. 1). D-4′-Phosphopantothenate enzymes can be classified into three different types based on their sequences: PanKs of type I and type III are found in a wide range of bacteria, whereas type II is mostly present in eukaryotes, but has also been identified in *Staphylococci*. Intriguingly, a search for PanK and PS genes in the symbiotic *C. poribacteria* genomes failed to identify homologs for these enzymes. Based on multiple sequence alignments and phylogenetic tree construction, we determined that the proteins annotated as GHMP kinase (GHMPK) and PP/PS in these genomes are similar to the archaeal enzymes pantoate kinase (PoK) that phosphorylates D-pantoate, and phosphopantothenate synthetase (PPS), responsible for condensation of D-4-phosphopantoate with *β*-alanine, respectively (fig. 2*A* and *B*, and supplementary figs. 1 and 2, Supplementary Material online). The *C. poribacteria* proteins displayed 31% or 44% identity with *Methanospirillum hungatei* PoK and PPS (supplementary table 2, Supplementary Material online). Important PPS residues for substrates binding, deduced from the 3D structure (Kim et al. 2013), appeared to be conserved (supplementary fig. 2, Supplementary Material online). Among the 38 symbiotic *C. poribacteria* strains sequenced so far as part of metagenomic analyses, 31 presented a typical PoK gene and 34 displayed a characteristic PPS gene (table 1). As neither PanK nor PS gene were identified in the Entoporibacteria genomes missing the PoK or PPS gene, it supports the notion that a lack of detection of homologous genes in these genomes resulted from incomplete sequencing of the corresponding strains. This observation suggests the presence of the archaeal pathway in the symbiotic poribacteria. Consequently, some ambiguous protein annotations can be resolved, because GHMPK and PP/PS are orthologs of PoK and PPS enzymes, respectively, based on reciprocal best-hit BLAST searches. Multiple sequence alignment and phylogenetic analyses of *C. poribacteria* and representative archaeal PoK and PPS protein sequences revealed that the *Entoporibacteria* sequences clustered together, forming an individual group, more closely related to the euryarchaeota group (supplementary fig. 3, Supplementary Material online). The use of the archaeal rather than the bacterial reaction order for the synthesis of the CoA intermediate D-4′-phosphopantothenate in the symbiotic Entoporibacteria

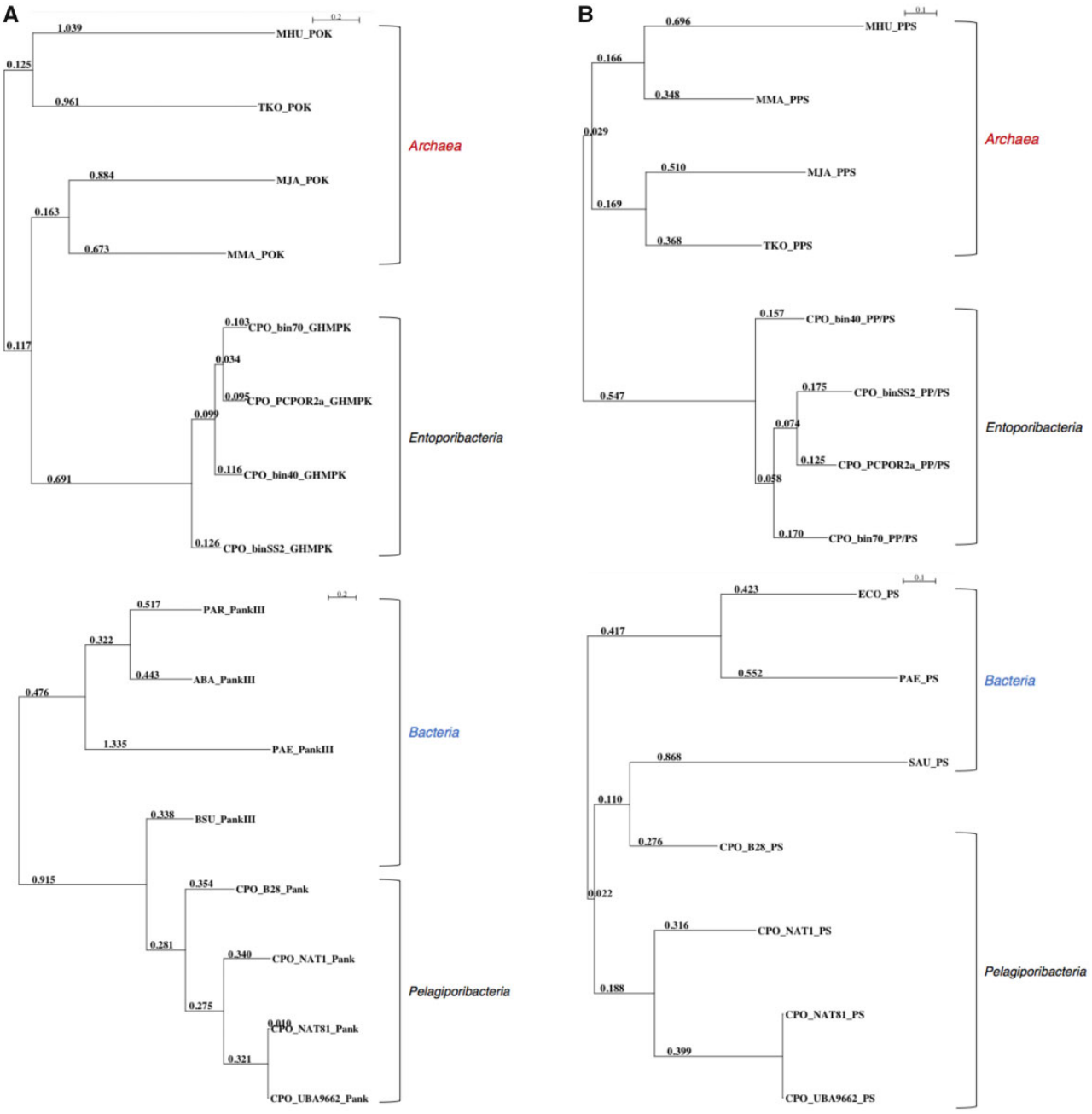

Fig. 2.—Bayesian phylogenetic trees of enzymes involved in phosphopantothenate pathway in archaeal and bacterial groups. The obtained trees show the distribution of the symbiotic (Entoporibacteria) and free-living (Pelagiporibacteria) *Candidatus poribacteria* groups according to the use of (*A*) pantoate kinase (PoK) and pantothenate kinase (PanK) enzymes, and (*B*) phosphopantothenate synthetase (PPS) and pantothenate synthetase (PS) enzymes (upper and lower panels, respectively). Branch lengths are shown for major nodes. Scale bar represents 0.2 and 0.1 amino acid replacements per site per unit evolutionary time on panels *A* and *B*, respectively. Abbreviation: CPO, *Candidatus poribacteria*; MHU, *Methanospirillum hungatei*; TKO, *Thermococcus kodakarensis*; MJA, *Methanocaldococcus jannaschii*; MMA, *Methanosarcina mazei*; ECO, *Escherichia coli*; SAU, *Staphylococcus aureus*; PAR, *Psychrobacter arcticus*; ABA, *Acinetobacter baumannii*; PAE, *Pseudomonas aeruginosa*; BSU, *Bacillus subtilis*.

group mirrors what has been shown in the archaea *Picrophilus torridus*, in which an enzyme much closer to the bacterial PanK than to the canonical archaeal PoK was found and has been annotated as "archaeal PanK" (Takagi et al. 2010; Shimosaka et al. 2016).

Strikingly, the archaeal enzymes of PP pathway present in Entoporibacteria appeared to be absent from the free-living Pelagiporibacteria that possess the classical bacterial genes coding for type-III PanK and PS (fig. 2*A* and *B*, and supplementary figs. 4 and 5, Supplementary Material online). For 19

Table 1
Occurrence of the Enzymes Catalyzing the Eight Reactions of the CoA Biosynthetic Pathway in Symbiotic and Free-Living *Candidatus poribacteria* Groups

Candidatus	Strain ID	KPHMT	KPR	KARI	POK	PPS	PS	Pank III	PPCS/PPCDC (bifunctional protein CoaBC)	PPAT	DPCK
Entoporibacteria	WGA-A3										
	WGA-4G										
	WGA-4E										
	WGA-4C										
	WGA-4CII										
	WGA-3G										
	bin70										
	bin44										
	MSPOR6										
	AGPOR5										
	PCPOR2a										
	PCPOR2										
	PCPOR2b										
	PCPOR1										
	DGPOR9										
	PCPOR4										
	PNGco_C_bin3										
	PNGco_C_binSS2										
	SB0668_bin_40										
	SB0672_bin_19										
	SB0664_bin_42										
	SB0665_bin_33										
	SB0677_bin_25										
	SB0678_bin_11										
	SB0670_bin_34										
	SB0662_bin_35										
	SB0668_bin_37										
	SB0668_bin_36										
	SB0676_bin_15										
	SB0662_bin_49										
	SB0661_bin_50										
	SB0662_bin_50										
	SB0663_bin_6										
	SB0670_bin_33										
	SB0664_bin_28										
	SB0661_bin_29										
	SB0669_bin_10										
	SB0675_bin_22										
Pelagiporibacteria	ARS1035										
	ARS87										
	NAT81										
	NAT79										
	MED599										
	ARS61										
	NAT1										
	SAT10										
	SAT1451										
	SP142										
	RS423										
	NP41										
	NP60										

(continued)

Table 1 Continued

Candidatus	Strain ID	KPHMT	KPR	KARI	POK	PPS	PS	Pank III	PPCS/PPCDC (bifunctional protein CoaBC)	PPAT	DPCK
	UBA9662										
	B28_G17		*								
	AG-410-M18										
	HyVt-180		*								
	SpSt-310										
	SpSt-970		*								

NOTE.—Identification of homologous genes is depicted by a gray box. Absence of homologous gene (white box) must be interpreted with caution, as it can result from incomplete sequencing of the related strain. KPR genes marked by a star correspond to homologous genes identified with low-similarity scores. For more details see supplementary table 3, Supplementary Material online. KPHMT, ketopantoate hydroxymethyltransferase; KPR, ketopantoate reductase; KARI, ketol-acid reductoisomerase; PoK, pantoate kinase; PPS, phosphopantothenate synthetase; PS, pantothenate synthetase; PanK, pantothenate kinase type-III; PPCS, phosphopantothenoylcysteine synthetase; PPCDC, phosphopantothenoylcysteine decarboxylase; PPAT, phosphopantetheine adenylyltransferase; DPCK, dephospho-CoA kinase.

sequenced Pelagiporibacteria, 10 and 16 of them present PanK or PS homologous genes, respectively (table 1), and *C. poribacteria* proteins presented 26% or 47% identity with *Pseudomonas aeruginosa* type-III PanK and *Escherichia coli* PS (supplementary table 2, Supplementary Material online). Importantly, PanK and PS residues involved in substrates binding deduced from the 3D structure (von Delft et al. 2001; Yang et al. 2006) appeared also to be conserved (supplementary figs. 4 and 5, Supplementary Material online). This observation implies that utilization of the alternative PP pathway is dictated by the bacteria interaction with its environment and likely contributes to the holobiont interaction.

Another proposed gene to discriminate between archaeal and bacterial/eukaryotic CoA biosynthetic pathways is the one coding for dephospho-CoA kinase (DPCK). It was demonstrated in *Thermococcus kodakarensis*, and next extended by inference to others archaea groups, that archaea DPCK is not homologous to the classical bacterial and eukaryotic enzymes, but instead is distantly related to proteins of the thiamine pyrophosphokinase family (Shimosaka et al. 2019). Therefore, we searched for a DPCK homologous genes in the two *C. poribacteria* groups. Multiple sequence alignment showed that both Ento- and Pelagi-poribacteria genomes contain a typical bacterial DPCK protein (supplementary fig. 6, Supplementary Material online), presenting an average 30.5% identity with *E. coli* DPCK. This observation stressed that only limited part of the archaeal CoA biosynthetic pathway is represented in the poribacteria group.

Next, we investigated the occurrence of all additional genes coding for enzymes of the CoA biosynthetic pathway in the different *C. poribacteria* genomes. Genes homologous to *E. coli* ketopantoate hydroxymethyltransferase (KPHMT), the bifunctional phosphopantothenoylcysteine synthetase (PPCS)—phosphopantothenoylcysteine decarboxylase (PPCDC), and phosphopantetheine adenylyltransferase (PPAT) were identified (table 1). Surprisingly, no gene encoding a protein with significant similarity to the classical bacterial KPR was detected in the *C. poribacteria* genomes. However, in the two bacteria *Corynebacterium glutamicum* and *Thermotoga maritima* that also lack a classical KPR, it has been demonstrated that this activity is performed by the ketol-acid reductoisomerase (KARI) encoded by the ilvC gene, an enzyme classically involved in the biosynthesis of branched-chain amino acids (Merkamm et al. 2003; Shimosaka et al. 2016). Homology search identified a gene coding for a typical bacterial KARI in most *C. poribacteria* genomes (table 1; supplementary fig. 7, Supplementary Material online). This is the first report of absence of a KPR gene in an organism presenting the PoK and PPS genes.

Discussion

This study revealed that the origin of PP biosynthesis is more complex than anticipated and that what was initially defined as an archaeal pathway is also used in some bacterial groups. It also suggests a high degree of evolutive and functional plasticity in the biosynthesis of the metabolic intermediates of CoA. Interestingly, a similar mosaic evolution utilizing alternative routes in different bacteria and archaea were identified for the mevalonate pathway, a biosynthetic process utilizing acetyl-CoA to form isoprenoid precursors (Lombard and Moreira 2011; Hoshino and Gaucher 2018). Of note, despite the similarities in the catalyzed reactions, multiple sequence alignment showed no common domain between PanK and PoK on one side, nor PS and PPS on the other side. This observation clearly indicated that these enzymes originate from different ancestral genes.

Despite the ubiquity of the CoA pathway, the uneven taxonomic distribution of the two routes for PP synthesis raises several questions about their evolutionary origin. Two alternative but nonexclusive explanations for this phylogenic plasticity can be proposed. On the one hand, the exceptions to the phyla-specific synthesis pathways were acquired by distinct archaeal or bacterial groups through horizontal gene transfers. Our results support this hypothesis, since we showed that, from the pool of sequenced *C. poribacteria* genomes, only symbiotic Entoporibacteria use the archaeal

pathway, as the free-living Pelagiporibacteria use the bacterial pathway. Entoporibacteria are part of a large microbial community that colonizes the sponge mesohyl, constituting an ancestral form of microbiota (Webster and Thomas 2016; Pita et al. 2018). This community includes both bacteria and archaea in close proximity, a condition favorable for interindividual gene transfer. Accordingly, analysis of *C. poribacteria* genomes revealed the presence of many genes coding for eukaryote-like proteins, which were predicted to be involved in mediating host–microbe interactions (Kamke et al. 2014; Podell et al. 2019). Alternatively, the two pathways may have originally coexisted in a common ancestor and one or another of the two branches were then positively or negatively selected according to metabolic properties or toxic effects of the possible metabolic intermediates. However, so far coexistence of the two pathways has not been described in any organism, suggesting that they are exclusive to each other. It has been proposed by Atomi and co-authors that the CoA pathway evolved from an ancestral set of two reactions, a condensation between a cysteine residue and a carboxylic acid followed by removal of the cysteine-derived carboxy group, generating a simplified CoA equivalent (Atomi et al. 2013). Our phylogenetic profiling study not only reinforces the notion of a mosaic of orthologous relationships of CoA biosynthetic genes between bacteria and archaea as originally proposed by Genschel (2004), but also highlights the importance of alternative pathways as an adaptation to environmental constrains.

Materials and Methods

Sequence Retrieval and Multiple Alignment Construction

Reference protein sequences for each enzyme of interest (type I/II/III PanKs, PoK, PS, PPS) were retrieved from the UniprotKB database (https://www.uniprot.org/). Accession numbers for the selected enzymes are shown in supplementary table 1, Supplementary Material online. For these sequences, conserved protein domains were extracted from the CDD database (https://www.ncbi.nlm.nih.gov/Structure/cdd) and conserved regions were identified in the multiple alignment representative of each domain (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Then, three additional sequences from other species were retrieved for each enzyme of interest based on their reviewed Uniprot annotations (see supplementary table 1, Supplementary Material online). For each enzyme, a multiple alignment of the retrieved sequences was constructed using Muscle and conserved regions were manually compared with the expected regions identified in the conserved protein domains.

We then used each reference protein sequence as a query for a TBlastN search of the *C. poribacteria* genomes in the WGS database (https://www.ncbi.nlm.nih.gov/genbank/wgs/), since there are more genomic assemblies (64 on mai 2020) than proteomes available (22). This allowed us to retrieve four protein sequences from pelagic *C. poribacteria* using the reference type-III PanK protein as a query, and four protein sequences from symbiotic *C. poribacteria* using the reference PoK protein as a query (supplementary table 2, Supplementary Material online). Genome assemblies and contigs used are shown in supplementary tables 1–3, Supplementary Material online. Finally, multiple alignments of the retrieved PoK enzymes including Entoporibacteria misannotated sequences and of the retrieved type-III PanK enzymes including Pelagiporibacteria sequences were constructed using Muscle 3.8.31 (Edgar 2004) (supplementary figs. 1 and 4, Supplementary Material online).

A similar approach was used to obtain PS and PPS sequences. First, we retrieved PS sequences for the three reference bacteria previously determined and four PPS sequences for the four PoK-coding archaea from UniprotKB. Accession numbers for the selected enzymes are shown in supplementary table 1, Supplementary Material online. For these sequences, conserved protein domains were identified in the CDD database. For each enzyme, a multiple alignment was constructed and conserved regions were compared with the expected regions from the conserved protein domains. Using the *E. coli* and the *M. hungatei* sequences, we then performed a TBlastN search of the WGS database to retrieve the four protein sequences from the previously considered Pelagiporibacteria and Entoporibacteria. Genome assemblies and contigs used are presented in supplementary table 2, Supplementary Material online. Finally, multiple alignments of the retrieved PPS enzymes (including Entoporibacteria misannotated sequences) and of the retrieved PS enzymes (including Pelagiporibacteria sequences) were constructed using Muscle (supplementary figs. 2 and 5, Supplementary Material online).

Phylogenetic Analyses

The tree construction was carried out using PhyloBayes v.4.1 (Lartillot et al. 2009) for 1) the set of type-III PanKs or 2) PoK proteins, and 3) the set of PS or 4) PPS proteins (supplementary figs. 1 and 4 or 2 and 5, Supplementary Material online, respectively). For each set, two Bayesian analyses were performed using either the single substitution model (LG) or the profile mixture model (CAT-GTR). Each analysis was performed in duplicate, and the convergence was assessed using the bpcomp function provided by PhyloBayes. For each analysis, 100 sampled points were removed as burn-in. The tree topologies obtained for each set using both models are almost identical, and therefore only the CAT-GTR trees are shown here (fig. 2).

Supplementary Material

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

Acknowledgments

We are indebted to Julie Thompson for fruitful discussions and comments on the results of this study. This work was supported by Centre National de la Recherche Scientifique (CNRS), the Meyer Fund and ADISSEO FRANCE SAS (Laboratoire Commun DiagnOxi to A.L.), and a CIFRE fellowship (Grant No. 2017/0063) from the Association Nationale Recherche et Technologie to L.T. supervised by CNRS and ADISSEO FRANCE SAS.

Data Availability

The data underlying this article are available in the article and in its online supplementary material.

Literature Cited

Atomi H, Tomita H, Ishibashi T, Yokooji Y, Imanaka T. 2013. CoA biosynthesis in archaea. Biochem Soc Trans. 41(1):427–431.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–1797.

Fieseler L, Horn M, Wagner M, Hentschel U. 2004. Discovery of the novel candidate phylum "poribacteria" in marine sponges. Appl Environ Microbiol. 70(6):3724–3732.

Genschel U. 2004. Coenzyme A biosynthesis: reconstruction of the pathway in archaea and an evolutionary scenario based on comparative genomics. Mol Biol Evol. 21(7):1242–1251.

Hoshino Y, Gaucher EA. 2018. On the Origin of Isoprenoid Biosynthesis. Mol Biol Evol. 35(9):2185–2197.

Ishibashi T, et al. 2012. A detailed biochemical characterization of phosphopantothenate synthetase, a novel enzyme involved in coenzyme A biosynthesis in the archaea. Extremophiles 16(6):819–828.

Kamke J, et al. 2014. The candidate phylum poribacteria by single-cell genomics: new insights into phylogeny, cell-compartmentation, eukaryote-like repeat proteins, and other genomic features. PLoS ONE 9(1):e87353.

Katoh H, Tamaki H, Tokutake Y, Hanada S, Chohnan S. 2013. Identification of pantoate kinase and phosphopantothenate synthetase from *Methanospirillum hungatei*. J Biosci Bioeng. 115(4):372–376.

Kim M-K, An YJ, Cha S-S. 2013. The crystal structure of a novel phosphopantothenate synthetase from the hyperthermophilic archaea, *Thermococcus onnurineus* NA1. Biochem Biophys Res Commun. 439(4):533–538.

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

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.

Merkamm M, Chassagnole C, Lindley ND, Guyonvarch A. 2003. Ketopantoate reductase activity is only encoded by ilvC in *Corynebacterium glutamicum*. J Biotechnol. 104(1–3):253–260.

Pita L, Rix L, Slaby BM, Franke A, Hentschel U. 2018. The sponge holobiont in a changing ocean: from microbes to ecosystems. Microbiome 6(1):46.

Podell S, et al. 2019. Pangenomic comparison of globally distributed Poribacteria associated with sponge hosts and marine particles. ISME J. 13(2):468–481.

Shimosaka T, Makarova KS, Koonin EV, Atomi H. 2019. Identification of dephospho-coenzyme A (Dephospho-CoA) kinase in *Thermococcus kodakarensis* and elucidation of the entire CoA biosynthesis pathway in archaea. mBio 10(4):e01146–19.

Shimosaka T, Tomita H, Atomi H. 2016. Regulation of coenzyme A biosynthesis in the hyperthermophilic bacterium *Thermotoga maritima*. J Bacteriol. 198(14):1993–2000.

Takagi M, et al. 2010. Pantothenate kinase from the thermoacidophilic archaeon *Picrophilus torridus*. J Bacteriol. 192(1):233–241.

Tomita H, Yokooji Y, Ishibashi T, Imanaka T, Atomi H. 2012. Biochemical characterization of pantoate kinase, a novel enzyme necessary for coenzyme A biosynthesis in the archaea. J Bacteriol. 194(19):5434–5443.

von Delft F, et al. 2001. The crystal structure of *E. coli* pantothenate synthetase confirms it as a member of the cytidylyltransferase superfamily. Structure 9(5):439–450.

Webster NS, Thomas T. 2016. The sponge hologenome. mBio 7(2):e00135–16.

Yang K, et al. 2006. Crystal structure of a type III pantothenate kinase: insight into the mechanism of an essential coenzyme A biosynthetic enzyme universally distributed in bacteria. J Bacteriol. 188(15):5532–5540.

Yokooji Y, Tomita H, Atomi H, Imanaka T. 2009. Pantoate kinase and phosphopantothenate synthetase, two novel enzymes necessary for CoA biosynthesis in the archaea. J Biol Chem. 284(41):28137–28145.

Associate editor: Maria Costantini