environmental microbiology

Environmental Microbiology (2012) **14**(3), 691–701 doi:10.1111/j.1462-2920.2011.02621.x

Evolution of the biosynthesis of di-*myo*-inositol phosphate, a marker of adaptation to hot marine environments

Luís G. Gonçalves,[1] Nuno Borges,[1] François Serra,[2] Pedro L. Fernandes,[3] Hernán Dopazo[2]* and Helena Santos[1]*

[1]*Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Av. da República-EAN, Apartado 127, 2780-157 Oeiras, Portugal.*

[2]*Evolutionary Genomics Laboratory, Bioinformatics & Genomics Department, Centro de Investigación Príncipe Felipe, Valencia 46012, Spain.*

[3]*Instituto Gulbenkian de Ciência, R. Quinta Grande 6, P-2780-156 Oeiras, Portugal.*

Summary

The synthesis of di-*myo*-inositol phosphate (DIP), a common compatible solute in hyperthermophiles, involves the consecutive actions of inositol-1-phosphate cytidylyltransferase (IPCT) and di-*myo*-inositol phosphate phosphate synthase (DIPPS). In most cases, both activities are present in a single gene product, but separate genes are also found in a few organisms. Genes for IPCT and DIPPS were found in the genomes of 33 organisms, all with thermophilic/hyperthermophilic lifestyles. Phylogeny of IPCT/DIPPS revealed an incongruent topology with 16S RNA phylogeny, thus suggesting horizontal gene transfer. The phylogenetic tree of the DIPPS domain was rooted by using phosphatidylinositol phosphate synthase sequences as out-group. The root locates at the separation of genomes with fused and split genes. We propose that the gene encoding DIPPS was recruited from the biosynthesis of phosphatidylinositol. The last DIP-synthesizing ancestor harboured separated genes for IPCT and DIPPS and this architecture was maintained in a crenarchaeal lineage, and transferred by horizontal gene transfer to hyperthermophilic marine *Thermotoga* species. It is plausible that the driving force for the assembly of those two genes in the early ancestor is related to the acquired advantage of DIP producers to cope with high temperature. This work corroborates the view that Archaea were the first hyperthermophilic organisms.

Received 27 July, 2011; revised 14 September, 2011; accepted 17 September, 2011. *For correspondence. E-mail santos@itqb.unl.pt; Tel. (+351) 214469541; Fax (+351) 214428766; E-mail hdopazo@cipf.es; Tel. (+34) 963289680; Fax (+34) 963289701.

Introduction

Inositol and inositol-containing compounds have multiple physiological roles in the three Domains of Life. Inositol derivatives are ubiquitous in Eukarya and widely distributed in Archaea, but rarely encountered in Bacteria (for useful reviews see Michell, 2008; 2011). Several stereoisomeric forms of inositol are found in biological systems, *myo*-inositol being by far the most common. This stereoisomer is present in a variety of metabolites, such as inositol monophosphates, inositol polyphosphates, phosphatidylinositols, phosphatidylinositides, glycosylphosphatidylinositols, CDP-inositol, di-*myo*-inositol-1,3′-phosphate, glycero-phospho-inositol, and mannosyl-di-*myo*-inositol-1,3′-phosphate (Martins *et al*., 1996; Lamosa *et al*., 2006; Michell, 2008; Rodrigues *et al*., 2009).

Inositol-containing compounds are part of the heat stress response in a variety of halophilic and halotolerant prokaryotes that thrive in hot environments. Therefore, their role in thermoprotection of cells and cellular components has been often proposed (Scholz *et al*., 1992; Borges *et al*., 2010; Santos *et al*., 2011). Among these stress compounds, di-*myo*-inositol-1,3′-phosphate (DIP) is closely restricted to (hyper)thermophilic Archaea and Bacteria.

The biosynthetic pathway of DIP proceeds from glucose-6-phosphate via four steps: (i) conversion of glucose-6-phosphate into *myo*-inositol-1-phosphate, in a reaction catalysed by *myo*-inositol-1-phosphate synthase (Neelon *et al*., 2005); (ii) activation of *myo*-inositol-1-phosphate into CDP-inositol, catalysed by inositol-1-phosphate cytidylyltransferase (IPCT); (iii) condensation with another molecule of inositol-1-phosphate into di-*myo*-inositol-1,3′-phosphate phosphate (DIPP) by the action of DIPP synthase (DIPPS); and (iv) dephosphorylation of di-*myo*-inositol-phosphate phosphate into di-*myo*-inositol-1,3′-phosphate by a yet unknown DIPP phosphatase (Borges *et al*., 2006; Rodrigues *et al*., 2007) (Fig. 1). The genes encoding IPCT and DIPPS activities have been characterized and homologous sequences are present in all DIP-accumulating organisms with known genomes

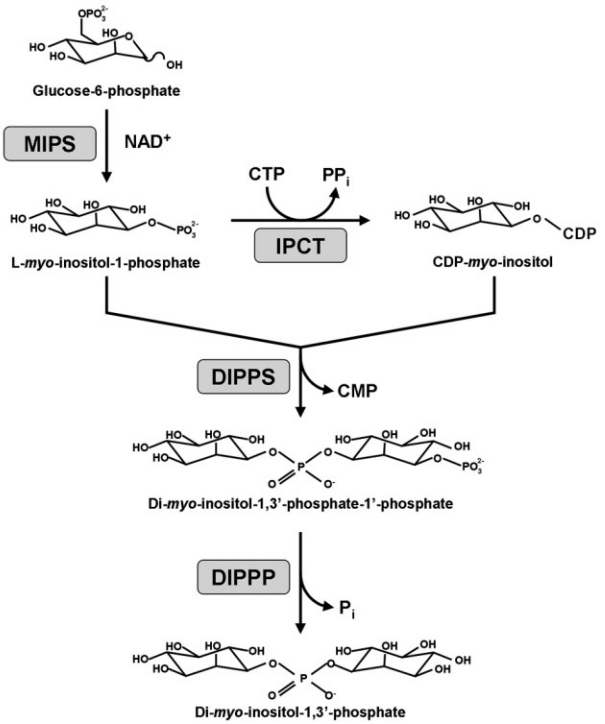

Fig. 1. Pathway for the synthesis of di-*myo*-inositol-1,3′-phosphate. MIPS, *myo*-inositol-1-phosphate synthase; IPCT, inositol-1-phosphate cytidylyltransferase; DIPPS, di-*myo*-inositol-1,3′-phosphate-1′-phosphate synthase; DIPPP, di-*myo*-inositol-1,3′-phosphate-1′-phosphate phosphatase.

(Rodionov *et al*., 2007; Rodrigues *et al*., 2007). In most cases, these two activities (IPCT and DIPPS) are present in a single gene product, but separate genes are found in a few hyperthermophiles, namely, the crenarchaeota *Aeropyrum pernix*, *Hyperthermus butylicus*, *Pyrolobus fumarii* and several *Thermotoga* species.

The DIPPS domain shows absolute specificity to inositol-1-phosphate, but some plasticity in regard to the CDP-alcohol, being able to recognize also CDP-glycerol, which in combination with inositol-1-phosphate, produces the phosphorylated form of glycero-phospho-inositol (Rodrigues *et al*., 2007). Recently, the first three-dimensional structure of an IPCT domain was resolved by X-ray. Surprisingly, the structure bears homology with glucose-1-phosphate thymidylyl/uridylyl-transferases and is less related to the cytidylyltransferases used for activation of polyol-phosphates (Brito *et al*., 2011). Regrettably, the structure of DIPPS has not been resolved due to difficulties in obtaining good-quality crystals of this membranar protein.

The accumulation of DIP has been detected within hyperthermophilic archaea, in members of the genera *Methanotorris*, *Thermococcus*, *Pyrococcus*, *Pyrodictium*, *Aeropyrum*, *Archaeoglobus*, *Stetteria*, *Hyperthermus* and *Pyrolobus*; it is present, along with other inositol derivatives, in bacteria belonging to the hyperthermophilic genera *Thermotoga* and *Aquifex*; finally, DIP is a minor solute in two thermophilic bacteria, *Rubrobacter xylanophilus* and *Persephonella marina* (Santos *et al*., 2011). The occurrence of DIP has never been reported in the Domain Eukarya. The distribution of DIP in the Tree of Life prompts obvious questions about the origin of DIP synthesis and the nature of the evolutionary events that led to its dissemination. In an effort to answer these questions we performed the phylogenetic analysis of this biosynthetic pathway. An evolutionary scenario for the biosynthetic enzymes, IPCT and DIPPS, as well as for the respective gene organization is proposed.

Results and discussion

Homologues of IPCT and DIPPS

In *Archaeoglobus fulgidus*, the IPCT and DIPPS domains are fused in a single polypeptide chain, while they are separated in *Thermotoga maritima* (Rodionov *et al*., 2007; Rodrigues *et al*., 2007). The IPCT/DIPPS proteins from these two organisms were used as templates for performing BLAST searches in the NCBI and JGI protein databases. Thirty-three IPCT/DIPPS homologues were retrieved and clustered into three distinct domain architecture groups (Table 1): (i) fused genes encoding the IPCT and DIPPS domains (22 organisms); (ii) separated genes for the IPCT and DIPPS domains (eight organisms); and (ii) genes containing IPCT, DIPPS and an extra C-terminal domain (three organisms).

Fused genes were found in several members of the *Euryarchaeota* (*A. fulgidus, Archaeoglobus profundus, Archaeoglobus veneficus*, *Ferroglobus placidus*, *Thermococcus kodakarensis*, *Thermococcus barophilus*, *Thermococcus onnurineus*, *Thermococcus* sp. AM4, *Thermococcus sibiricus*, *Thermococcus gammatolerans, Pyrococcus furiosus*, *Pyrococcus abyssi*, *Pyrococcus horikoshii*, *Methanocaldococcus infernus*), and also in some bacteria (*R. xylanophilus*, *Hydrogenivirga* sp. 128–5-R1-1, *Aquifex aeolicus*, *P. marina*, *Thermobaculum terrenum*, *Thermodesulfatador indicus*, *Thermoaerobacter subterraneus* and *Thermoaerobacter marianensis*). Interestingly, from all the *Methanocaldococcus* spp. sequenced thus far (*Methanocaldococcus fervens*, *Methanocaldococcus jannaschii*, *Methanocaldococcus* sp. FS406-22 and *Methanocaldococcus vulcanius*), *M. infernus* is the only one that possesses the genes for DIP synthesis (IPCT/DIPPS). In addition, *M. infernus* is the only member of the genus *Methanocaldococcus* that possesses the gene for the synthesis of *myo*-inositol-1-phosphate, a substrate for the first reaction in DIP synthesis.

Two organisms belonging to the *Crenarchaeota* (*Ignisphaera aggregans* and *Thermofilum pendens*) and one

Table 1. Organisms containing genes coding for inositol-1-phosphate cytidylyltransferase (IPCT) and di-*myo*-inositol-1,3′-phosphate-1′-phosphate synthase (DIPPS).

Organism	Domain; phylum; order	Habitat	Top (°C)	DIP accum.
IPCT and DIPPS genes fused				
Thermococcus kodakarensis	*Archaea; Euryarchaeota; Thermococcales*	M	86	+
Thermococcus barophilus	*Archaea; Euryarchaeota; Thermococcales*	M	85	n.s.
Thermococcus onnurineus	*Archaea; Euryarchaeota; Thermococcales*	M	80	n.s.
Thermococcus sp. AM4	*Archaea; Euryarchaeota; Thermococcales*	M	82	n.s.
Thermococcus sibiricus	*Archaea; Euryarchaeota; Thermococcales*	M	78	n.s.
Thermococcus gammatolerans	*Archaea; Euryarchaeota; Thermococcales*	M	88	n.s.
Pyrococcus furiosus	*Archaea; Euryarchaeota; Thermococcales*	M	100	+
Pyrococcus abyssi	*Archaea; Euryarchaeota; Thermococcales*	M	96	n.s.
Pyrococcus horikoshii	*Archaea; Euryarchaeota; Thermococcales*	M	98	+
Archaeoglobus fulgidus	*Archaea; Euryarchaeota; Archaeoglobales*	M	83	+
Archaeoglobus veneficus	*Archaea; Euryarchaeota; Archaeoglobales*	M	75	+
Archaeoglobus profundus	*Archaea; Euryarchaeota; Archaeoglobales*	M	85	+
Ferroglobus placidus	*Archaea; Euryarchaeota; Archaeoglobales*	M	85	n.s.
Methanocaldococcus infernus	*Archaea; Euryarchaeota; Methanococcales*	M	85	n.s.
Persephonella marina	*Bacteria; Aquificae; Aquilicales*	M	73	+
Hydrogenivirga sp. 128–5-R1-1	*Bacteria; Aquificae; Aquilicales*	M	73	n.s.
Aquifex aeolicus	*Bacteria; Aquificae; Aquilicales*	M	85	+
Rubrobacter xylanophilus	*Bacteria; Actinobacteria; Rubrobacterales*	M	55	+
Thermobaculum terrenum	*Bacteria; unclassified Bacteria; Thermobaculum*	S	67	n.s.
Thermodesulfatator indicus	*Bacteria; Thermodesulfobacteria; Thermodesulfobacteriales*	M	70	n.s.
Thermaerobacter marianensis	*Bacteria; Firmicutes; Clostridiales*	M	75	n.s.
Thermaerobacter subterraneus	*Bacteria; Firmicutes; Clostridiales*	S	70	n.s.
IPCT and DIPPS genes separated				
Pyrolobus fumarii	*Archaea; Crenarchaeota; Desulfurococcales*	M	106	+
Hyperthermus butylicus	*Archaea; Crenarchaeota; Desulfurococcales*	M	99	+
Aeropyrum pernix	*Archaea; Crenarchaeota; Desulfurococcales*	M	90	+
Thermotoga neapolitana	*Bacteria; Thermotogae; Thermotogales*	M	85	+
Thermotoga maritima	*Bacteria; Thermotogae; Thermotogales*	M	80	+
Thermotoga petrophila	*Bacteria; Thermotogae; Thermotogales*	M	80	n.s.
Thermotoga sp. RQ2	*Bacteria; Thermotogae; Thermotogales*	M	80	n.s.
Thermotoga naphthophila	*Bacteria; Thermotogae; Thermotogales*	M	80	n.s.
IPCT and DIPPS genes fused with an extra C-terminal domain				
Candidatus Korarchaeum cryptofilum	*Archaea; Korarchaeota; Candidatus Korarchaeum*	M	85	n.s.
Ignisphaera aggregans	*Archaea; Crenarchaeota; Desulfurococcales*	FW	92	n.s.
Thermofilum pendens	*Archaea; Crenarchaeota; Thermoproteales*	M	88	n.s.

Top, optimum temperature for growth; M, marine; S, soil; FW, fresh water; +, DIP-accumulating organism; n.s., DIP accumulation was not studied.

representative of the *Korarchaeota* (*Candidatus Korarchaeum cryptofilum*) possess genes encoding proteins with high similarity to the fused IPCT/DIPPS (around 24% of identity), but contain an extra C-terminal domain with unknown function. These genes were considered to encode putative IPCT/DIPPSs, although there is lack of experimental data to confirm the presence of DIP or the IPCT/DIPPS activities. Actually, the eventual presence of DIP synthesis in *I. aggregans* is curious, because the presence of compatible solutes is rare in salt-sensitive organisms (Niederberger *et al*., 2006; Santos *et al*., 2011).

Separated genes are present in three *Crenarchaeota* (*A. pernix*, *H. butylicus* and *P. fumarii*), and in several hyperthermophilic marine organisms classified as *Thermotoga* spp. (*T. maritima, Thermotoga neapolitana, Thermotoga naphtophila, Thermotoga* sp. RQ2 and *Thermotoga petrophila*). The gene encoding IPCT is located immediately downstream of the gene encoding DIPPS except for *P. fumarii*, in which the order of the two genes is inverted (Fig. 2). In all these organisms, the two genes overlap to a certain extent: 25 nucleotides in *H. butylicus*, 16 in *A. pernix*, 7 in *Thermotoga* spp. and 4 in *P. fumarii*.

Thermotoga spp. isolated from freshwater (*Thermotoga thermarum* and *Thermotoga lettingae*) do not possess candidate genes to code for IPCT/DIPPS, in line with previous results showing the absence of compatible solutes in non-halotolerant or non-halophilic (hyper)thermophiles (Lamosa *et al*., 1998). The DIP biosynthetic genes are also absent in the genomes of other sequenced *Thermotogales*, like *Fervidobacterium nodosum*, *Thermosipho melanesiensis*, *Thermosipho africanus*, *Petrotoga mobilis* and *Kosmotoga olearia*. Considering that within the order *Thermotogales*, only the genus *Thermotoga* comprises hyperthermophilic members, it is reasonable to propose DIP as a hallmark of hyperthermophily in *Thermotoga* spp. isolated from marine environments.

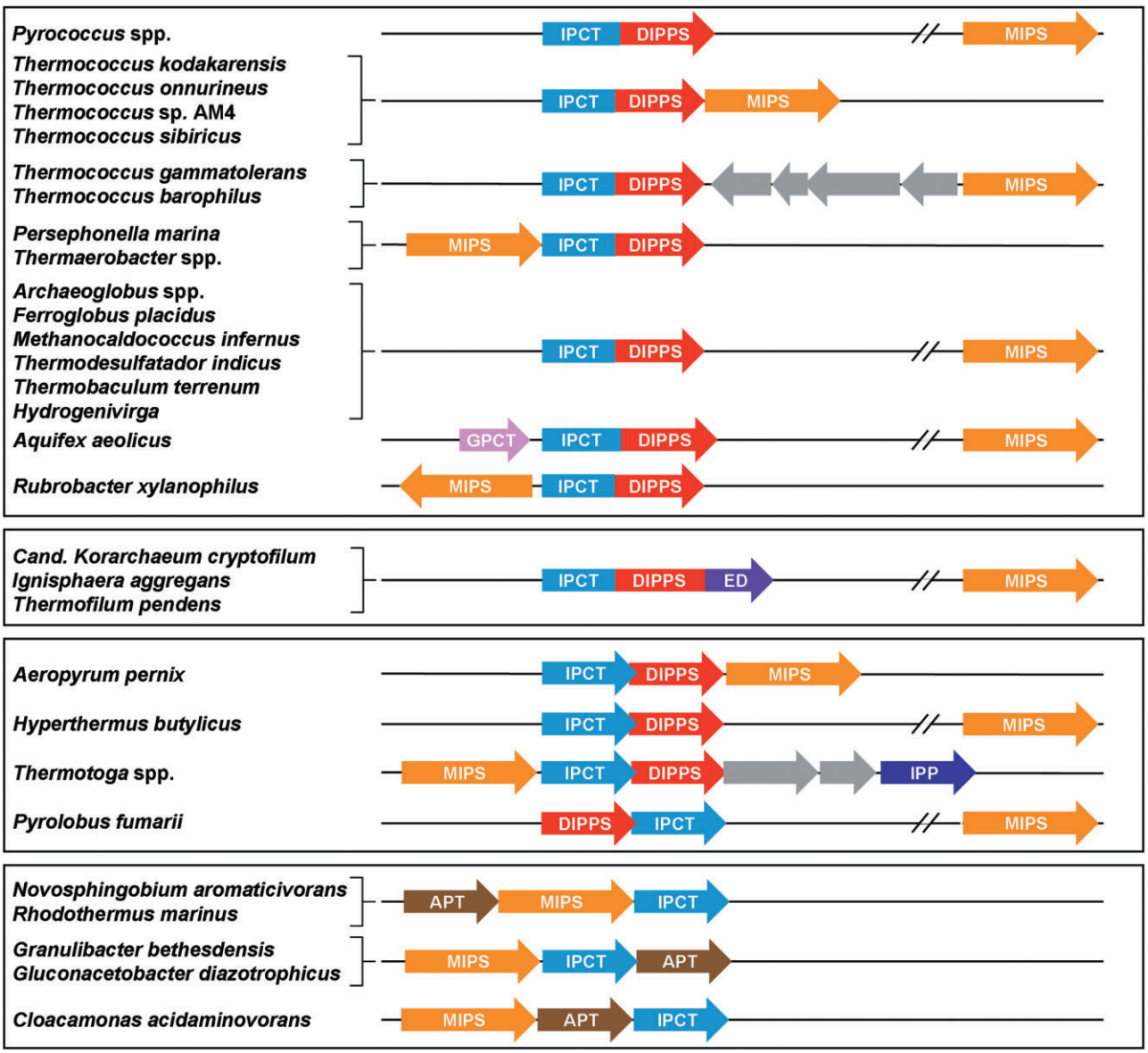

Fig. 2. Schematic organization of the genes encoding the enzymes involved in the synthesis of di-*myo*-inositol-1,3′-phosphate in several organisms. MIPS, *myo*-inositol-1-phosphate synthase (orange arrows); IPCT, inositol-1-phosphate cytidylyltransferase (light blue arrows), DIPPS, di-*myo*-inositol-1,3′-phosphate-1′-phosphate synthase (red arrows), IPP, *myo*-inositol-1-phosphate phosphatase (dark blue arrow); ED, extra C-terminal domain (purple arrow); GPCT, glycerol-phosphate cytidylyltransferase (pink arrow); APT, CDP-alcohol phosphatidyltransferase (brown arrows); unspecified genes (grey arrows).

The absolute correlation between the presence of genes for DIP synthesis and the gene encoding *myo*-inositol-1-phosphate synthase (MIPS) in the *Thermotogales* and in the *Methanocaldococcales* supports the view that the role of MIPS in these organisms is exclusively to provide the substrate required for the synthesis of DIP. Moreover, in the *Thermotoga* spp., genes coding for MIPS, IPCT and DIPPS are organized in an operon-like structure together with *myo*-inositol-1-phosphate phosphatase, which has been proposed to act in the dephosphorylation of DIPP (Fig. 2) (Rodionov *et al.*, 2007).

The phylogenetic distribution of the enzymes implicated in the synthesis of DIP is limited to organisms with an optimal temperature of growth above 55°C, which reveals a strong correlation between this compatible solute and adaptation to high temperatures. The presence of DIP was confirmed experimentally in many marine hyperthermophilic archaea, in a few hyperthermophilic bacteria belonging to the *Thermotogae* (*T. neapolitana* and *T. maritima*) and the *Aquificae* (*A. aeolicus* and *Aquifex pyrophilus*). DIP was also detected in *R. xylanophilus*, a member of the *Actinobacteria*, but only in vestigial amounts (Table 1 and Santos *et al.*, 2011). The present analysis

reveals that DIP synthesis is probably more widespread than previously thought. For example, in the Domain Bacteria, genes for DIP synthesis are predicted in the *Firmicutes* (*T. marianensis* and *T. subterraneus*), the *Thermodesulfobacteria* (*T. indicus*) and an unclassified bacterium (*T. terrenum*). Interestingly, thermophily is apparently the only trait common to these organisms.

Curiously, five bacterial members, *Rhodothermus marinus*, *Granulibacter bethesdensis*, *Candidatus Cloacamonas acidaminovorans*, *Gluconacetobacter diazotrophicus* and *Novosphingobium aromaticivorans*, possess putative IPCTs but not DIPPS. The predicted function was confirmed for *R. marinus* by cloning and heterologous expression of the respective gene (L.G. Gonçalves, N. Borges, C. Jorge, H. Santos, unpubl. data). In all cases the gene for IPCT is part of an operon-like structure with genes coding for MIPS and a putative CDP-alcohol phosphatidyltransferase (Fig. 2). This genomic organization suggests the involvement of IPCT in a yet unknown pathway in which CDP-inositol is used for a metabolic purpose other than DIP synthesis. Among the bacteria containing putative IPCTs, *R. marinus* has been extensively studied in respect to osmo- and thermoadaptation strategies and the presence of DIP was never observed. In fact, this thermophilic bacterium uses primarily mannosylglycerate as osmolyte for thermo- and osmo-adaptation (Nunes *et al*., 1995; Silva *et al*., 1999). Therefore, the role in *R. marinus* of CDP-inositol, the product of the IPCT activity, remains obscure.

Phylogeny of IPCT/DIPPS

A phylogenetic analysis of the proteins involved in the synthesis of DIP was performed to unravel their evolutionary history. The analysis was based on the sequences of the two domains (IPCT and DIPPS) by using both maximum likelihood and Bayesian methods. For the alignments, the IPCT and DIPPS sequences were concatenated whenever these activities were encoded by separated genes; the extra C-terminal domain present in some of the sequences was excluded (Fig. S1). Both inference methods led to topologies comprising four well-supported groups (bootstrap values higher than 80): one group includes the organisms that possess the IPCT and DIPPS domains separated; a second group is composed by bacterial proteins with the IPCT/DIPPS domains fused; the third one comprises fused IPCT/DIPPS domains, primarily from the phyla *Euryarchaeota* and *Aquificales*; finally, the fourth group comprises fused IPCT/DIPPS domains with an extra C-terminal extension (Figs 3 and S2).

The phylogenetic tree of IPCT/DIPPS does not show the dichotomy between Archaea and Bacteria observed in the 16S rRNA based phylogenetic tree (Fig. S3). There is a single branch that is monophyletic for Bacteria (composed by the sequences of *T. marianensis*, *T. subterraneus*, *R. xylanophilus* and *T. terrenum*); however, the majority of the bacterial representatives appear in branches highly populated by archaeal proteins, suggesting the occurrence of horizontal gene transfer (HGT) events in the evolution and dissemination of the IPCT/DIPPS genes. Interestingly, the bacterial sequences belonging to *Thermotogales* and *Aquificales* cluster with sequences from *Crenarchaeota* and *Euryarchaeota* respectively. This pattern is not surprising, as it is accepted that 20% of the genes in *A. aeolicus* and *T. maritima* have an archaeal origin (Aravind *et al*., 1998; Nelson *et al*., 1999). This view is reinforced by phylogenetic analyses of several proteins from *Aquificales* and *Thermotogales*, which revealed archaeal ancestors (Logsdon and Faguy, 1999; Forterre *et al*., 2000; Nesbø *et al*., 2001; Majumder *et al*., 2003; Nanavati *et al*., 2006; Boussau *et al*., 2008; Noll *et al*., 2008). The massive occurrence of HGT from hyperthermophilic archaea to those two bacterial phyla has been related to adaptation of *Aquificales* and *Thermotogales* to extremely hot environments (Aravind *et al*., 1998; Nelson *et al*., 1999; Koonin *et al*., 2001). This association emerges from the presence, in these bacteria, of archaeal genes typical of a hyperthermophilic lifestyle, such as the gene encoding reverse gyrase. This gene is present in all the hyperthermophilic organisms studied thus far, and it is regarded as a marker of hyperthermophily (Kikuchi and Asai, 1984; Shibata *et al*., 1987; Forterre *et al*., 2000). Several studies support an archaeal origin of reverse gyrase and the subsequent transference to hyperthermophilic bacteria, but the evolutionary history of this gene is not fully understood (Brochier-Armanet and Forterre, 2007; Heine and Chandra, 2009). The IPCT/DIPPS genes seem to be another example of an adaptive feature to (hyper)thermophily that was transferred between Archaea and Bacteria. In this case, the direction of the transference events could not be definitively determined (codon usage analysis and BLAST searches using flanking genes as templates gave no clues); however, the IPCT/DIPPS-based phylogenetic tree displays a division between *Euryarchaeota* and *Crenarchaeota*, and the major incongruence with the 16S rRNA-based topology concerns the bacterial sequences (Fig. S3). Thus, it is tempting to propose that the HGT events occurred from Archaea to Bacteria. This view is further supported by the distribution of inositol in the Tree of Life: this hexol is widespread in Archaea, commonly found as a moiety of phospholipid headgroups in archaeal membranes, and rarely found in Bacteria (Koga and Nakano, 2008; Michell, 2008). With this in mind, an archaeal origin of a pathway that involves inositol, such as DIP synthesis, seems to be more plausible than a bacte-

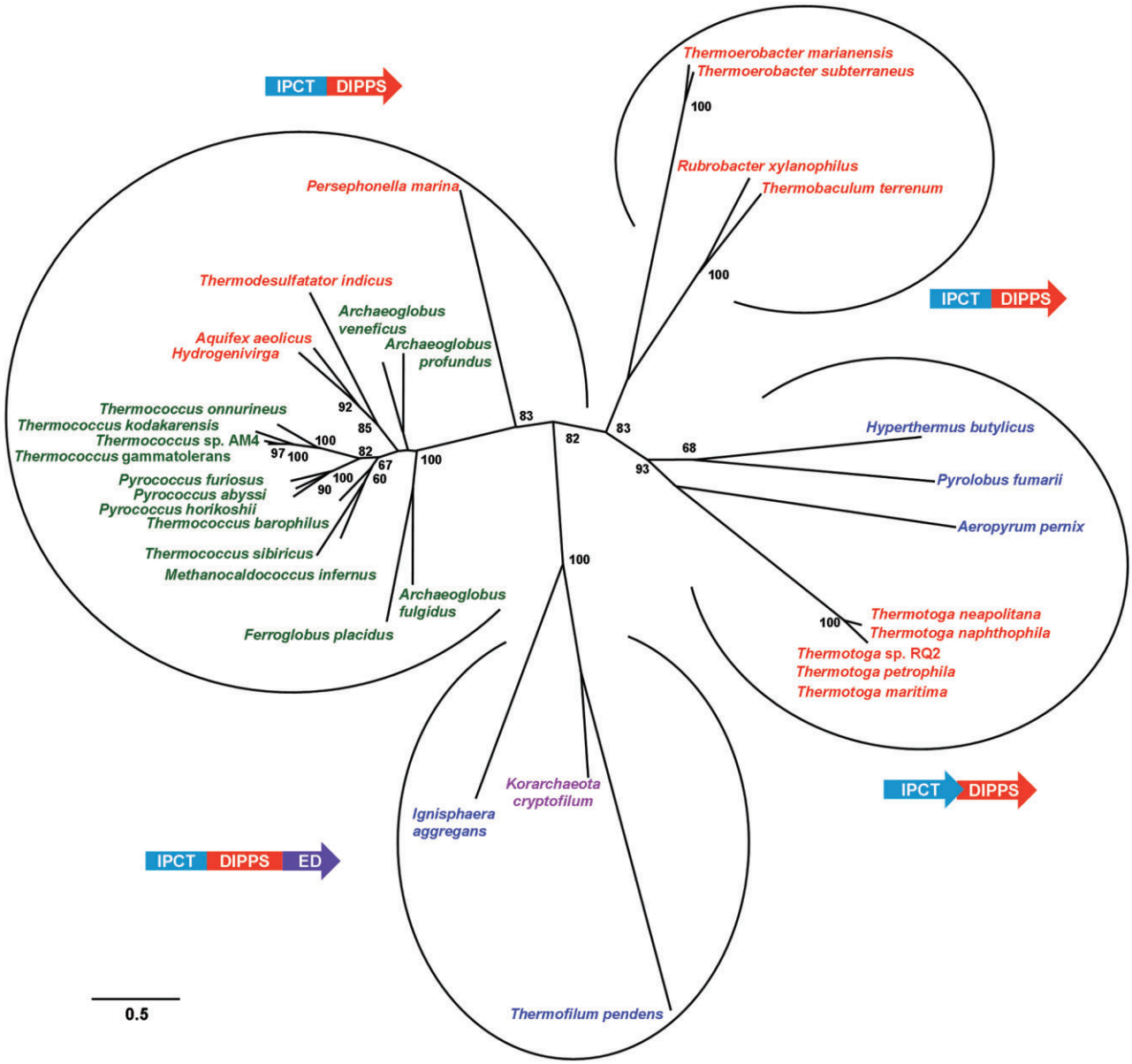

Fig. 3. Maximum likelihood phylogenetic unrooted tree of the IPCT/DIPPS proteins, constructed on the basis of amino acid sequences. Bootstrap values were calculated from 1000 replicates by PHYML. Only bootstrap values greater than 60 are shown. The scale bar represents the number of substitutions per site. *Bacteria* (red), *Euryarchaeota* (green), *Crenarchaeota* (blue), and *Korarchaeota* (purple). and indicate that the genes encoding IPCT and DIPPS are fused and separated respectively. The genes containing an extra domain are indicated by . IPCT and DIPPS GenPept accession numbers are given in Fig. 3S.

rial origin. In summary, our analysis suggests the occurrence of multiple HGT events from archaeal to bacterial organisms during the evolutionary history of IPCT/DIPPS.

It is interesting to note that the topology of the IPCT/DIPPS-based phylogenetic tree appears to be primarily determined by gene organization, that is, all proteins encoded by separate genes form a single cluster, sequences with an extended domain form a second cluster, and sequences encoding a bifunctional protein (fused genes) are found in the two remaining clusters (Fig. 3). Generally, the phylogenetic trees of multi-domain proteins reflect the differences in domain architecture (Yanai *et al.*, 2002). In most cases, the number of events (gene fission or fusion) leading to different domain architectures is small. This trend is also observed in the IPCT/DIPPS-based phylogenetic tree as only a few gene fusion/fission events are necessary to explain the observed clustering: one event led to the division of separated domains and fused domains;

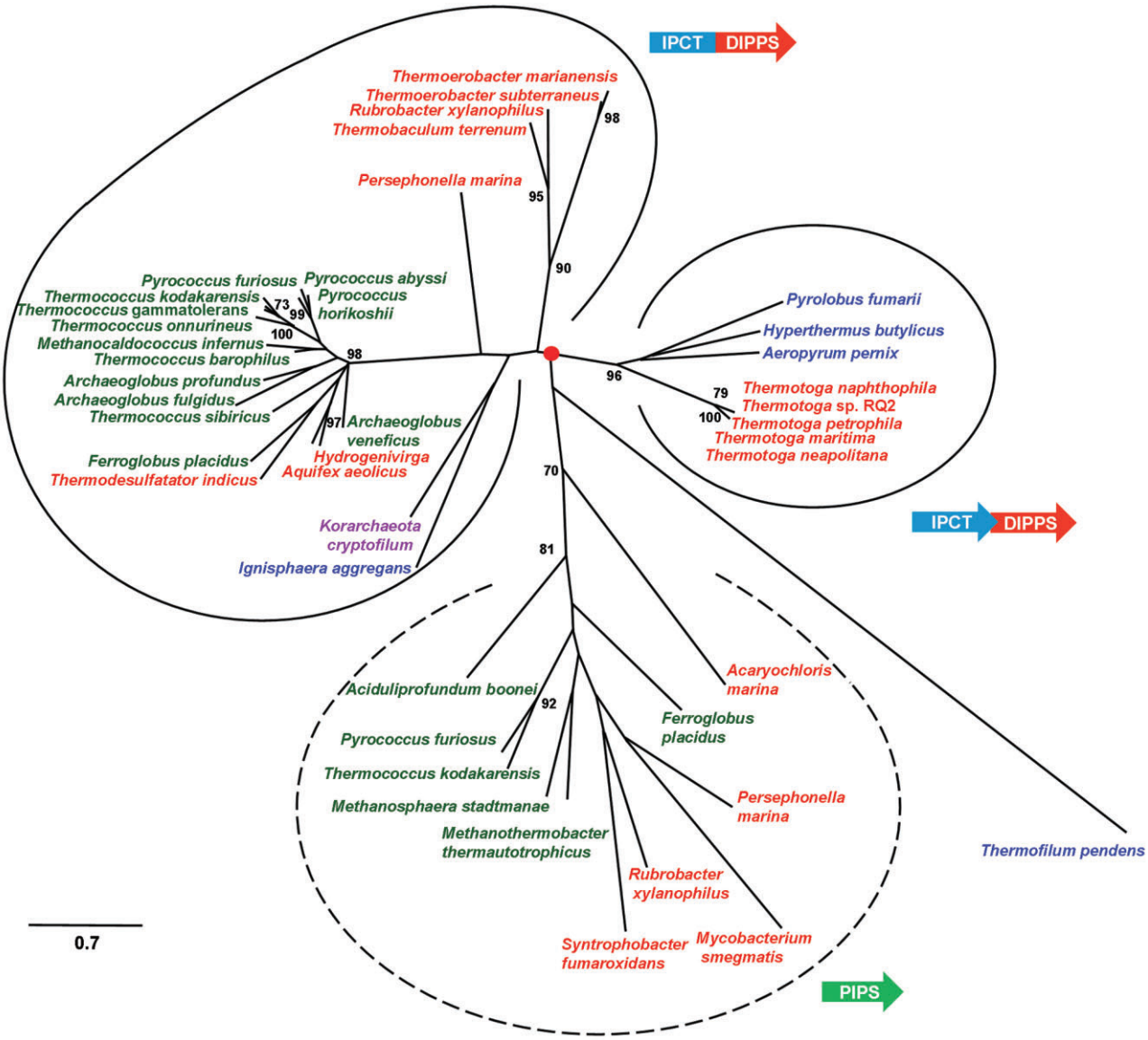

Fig. 4. Maximum likelihood phylogenetic tree of the DIPPS domain based on the amino acid sequences with eleven phosphatidylinositol-phosphate synthase (PIPS) sequences as out-group. Bootstrap values were calculated from 100 replicates by PHYML. Only bootstrap values greater than 60 are shown. Scale bar, and clade references as in Fig. 3.

another event resulted in the divergence of the branch composed by the sequences with an extra-domain.

Phylogeny of the DIPPS domain

Due to the lack of a suitable out-group sequence we were unable to root the IPCT/DIPPS-based phylogenetic tree (Fig. 3). Therefore, we decided to perform the phylogenetic analysis for each domain (IPCT and DIPPS) independently. The chemical structures of DIP and phosphatidylinositol are fairly related insofar as both molecules contain inositol residues engaged in a phosphodiester bond. Interestingly, the synthesis of DIP, like that of phosphatidylinositol in prokaryotes, proceeds via a two-step pathway involving a synthase and a phosphatase (Borges *et al*., 2006; Morii *et al*., 2009; 2010) (Fig. S4). Therefore, there is the formation of a phosphorylated intermediate, in contrast with the synthesis of phosphatidylinositol in Eukarya, which involves free inositol as substrate.

The phosphatidylinositol-phosphate synthase (PIPS) sequences are the closest relative of DIPPS (around 30% sequence similarity). Hence, we constructed a DIPPS phylogenetic tree using 11 PIPS from archaeal and bacterial sources as out-group (Fig. 4). The topology of the DIPPS phylogenetic tree is similar to that of the concat-

enated IPCT/DIPPS, with the exception of the sequence from *T. pendens* (compare Figs 3 and 4), which clusters together with PIPS sequences, though with a high divergence (around 6 substitutions per site). The explanation for this divergence in relation to the IPCT/DIPPS-based phylogenetic tree is the low similarity of the *T. pendens* DIPPS domain (at most 25%) with the other DIPPS sequences, while the coupled IPCT showed at least 48% similarity with the other IPCT sequences. Despite the low bootstrap value (23), the root of the DIPPS phylogenetic tree was identified at the node that separates the group with the IPCT/DIPPS fused genes from the group composed by separated genes (Fig. 4). Therefore, it is plausible that the event leading to the two different domain architectures (fused or separated genes), occurred only once in the evolutionary history of these proteins.

Interestingly, the IPCT/DIPPS is a very rare example of a bifunctional protein combining the domains for cytidylyltransferase and CDP-alcohol transferase in a single polypeptide chain. In fact, a search in the genome databases revealed a sole other example combining the domains for cytidylyltransferase and CDP-alcohol phosphatidyltransferase, presumably involved in the synthesis of polar lipids in the uncultured organism *Candidatus Nitrospira defluvii* (YP_003796430.1). The synthesis of lipid polar heads proceeds invariably via activation of the phosphatidyl group with CTP by a cytidylyltransferase, and subsequent transference to an alcohol group acceptor. A similar reaction sequence is used for the synthesis of many other cellular components, such as teichoic acids (Koga and Morii, 2007; Swoboda *et al*., 2010), but the two activities are never present in a single polypeptide chain. In this context, we infer that the most likely scenario in the evolutionary history of IPCT/DIPPS is the occurrence of a fusion event, rather than fission of an ancestral fused gene.

Phylogeny of the IPCT domain

For the construction of the phylogenetic tree of the IPCT domain, the homologous sequences from all the organisms depicted in Fig. 3, whose genomes also include a DIPPS homologue, were considered. In addition, the 'putative' IPCT sequences from three *Alphaproteobacteria* (*G. bethesdensis*, *G. diazotrophicus* and *N. aromaticivorans*), *R. marinus*, and *Candidatus Cloacamonas acidaminovorans* were included. The genomes of these organisms do not possess DIPPS genes, therefore the designation 'putative' IPCT is used herein. The IPCT phylogenetic tree showed a topology broadly similar to that of the DIPPS domain (Fig. 5). The division into four groups of sequences was apparent: IPCT/DIPPS fused genes; separated IPCT and DIPPS genes; IPCT/DIPPS fused genes with an extra C-terminal domain; and bacterial IPCT/DIPPS fused genes. The so-called 'putative IPCT' sequences clustered together with the bacterial IPCT/DIPPS fused sequences (*Thermaerobacter* spp., *R. xylanophilus*, and *T. terrenum*). This feature suggests that the putative IPCTs evolved from bacterial, fused IPCT/DIPPS with the loss of the second domain (Fig. 5).

When the BLAST searches for the IPCT were less stringent, *E*-values below 1, most of the sequences retrieved belonged to the glucose-1-phosphate thymidylyl-/uridylyl-transferases protein family. These sequences exhibit around 20% similarity with IPCTs. In this context, it is interesting that the X-ray structure of the *A. fulgidus* IPCT domain shows a high structural identity with glucose-1-phosphate thymidylyl-/uridylyltransferases (Brito *et al*., 2011). In contrast, alcohol nucleotidyltransferases, such as glycerol-, ribitol- or methylerythritol-cytidylyltransferase, do not share any structural or sequence similarity with IPCTs. These observations led us to postulate that IPCT, an enzyme that catalyses the activation of inositol, evolved from sugar nucleotidyltransferases rather than from alcohol nucleotidyltransferases.

Conclusions

Herein, we propose the first evolutionary model for the origin and evolution of key enzymes involved in the synthesis of DIP, a compatible solute closely restricted to (hyper)thermophiles (Fig. 6). The ancestral of enzyme DIPPS is found among phosphatidylinositol-phosphate synthases, revealing an evolutionary link between the synthesis of DIP and that of phosphatidylinositol, a major component of membrane lipids. On the other hand, IPCT evolved from sugar nucleotidyltransferases.

Most probably the early ancestor for DIP synthesis thrived in a marine-like, hot environment, and possessed the two enzyme activities encoded by separate genes; this gene architecture was maintained in a crenarchaeal lineage, and transferred by HGT to hyperthermophilic marine *Thermotoga* species.

In a different descendant lineage, fusion of the genes coding for the IPCT and DIPPS domains occurred, and this new gene architecture was propagated to Bacteria and *Euryarchaea*. Finally, fission of the gene encoding the bifunctional IPCT/DIPPS, and loss of the DIPPS domain in the early bacterial lineage, gave rise to bacterial members carrying only the IPCT domain. Presumably, the product of this IPCT activity, CDP-inositol, is a protagonist in as yet unknown biosynthetic pathways.

To summarize our present understanding on the origin and propagation of DIP synthesis, we propose that the last ancestor possessed independent genes for IPCT and DIPPS, and gene fusion was a later event during evolution. Finally, an evolutionary link between the synthesis of phosphatidylinositol and DIP synthesis became apparent.

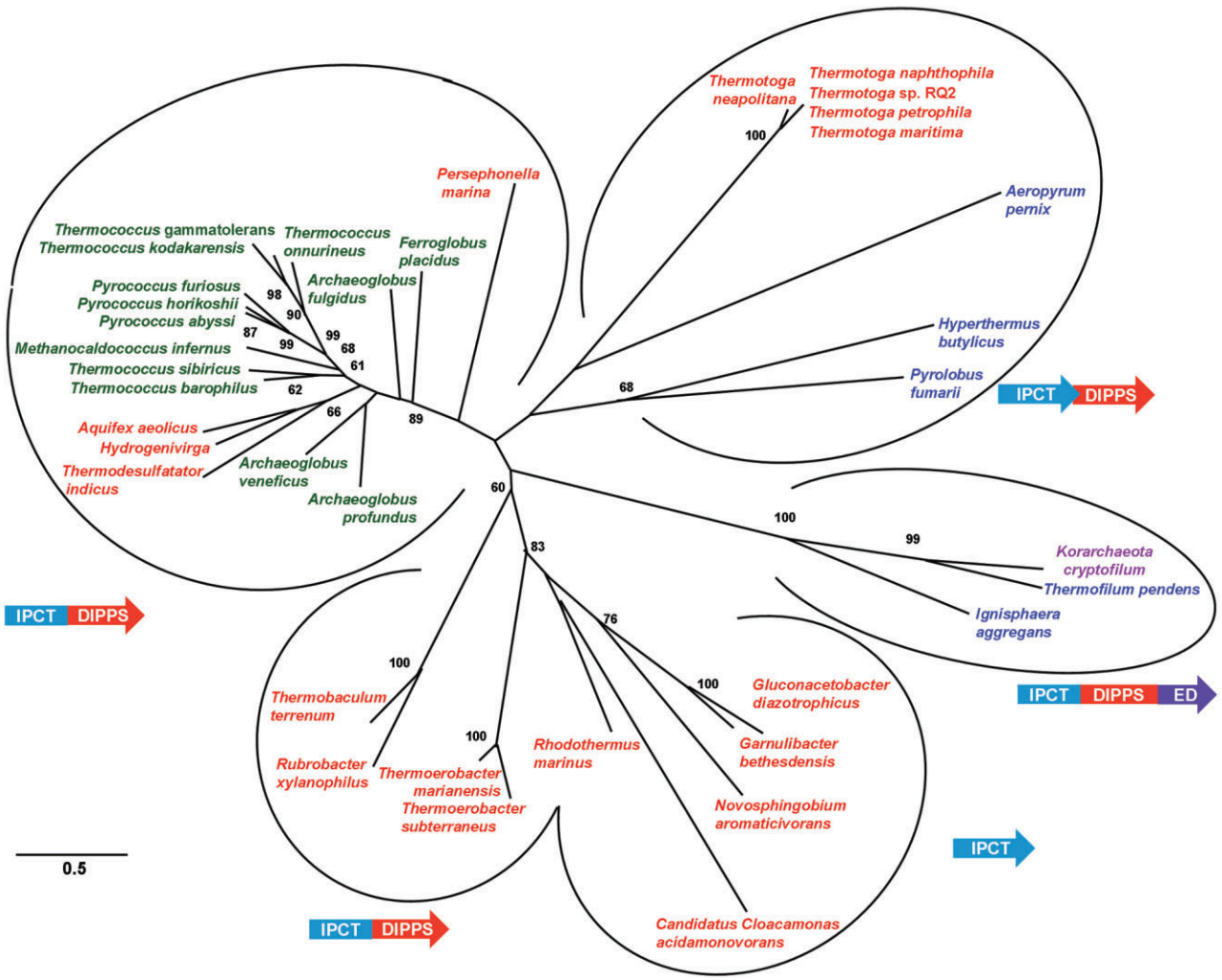

Fig. 5. Maximum likelihood unrooted phylogenetic tree of the IPCT domain based on the amino acid sequences obtained from public databases. Only bootstrap values greater than 60 are shown. Scale bar, and clade references as in Fig. 3.

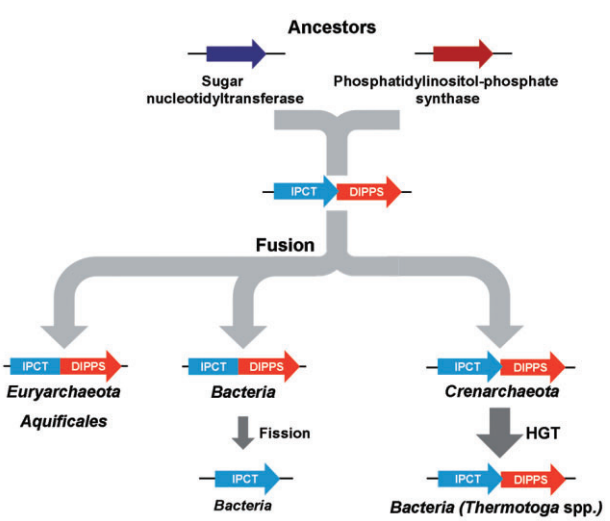

Fig. 6. Evolutionary model for the origin and evolution of DIP biosynthesis.

Experimental procedures

Construction of the datasets

The sequences of IPCT/DIPPS were retrieved from the non-redundant (nr) database at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov) and the Joint Genomic Institute (JGI) (http://www.jgi.doe.gov/), by BLASTP searches on February 2011. The sequences of the IPCT/DIPPS of *A. fulgidus*, and IPCT and DIPPS of *T. maritima* were used as seeds. The hits from uncultured organisms were not considered in this study. For the phylogenetic analysis, the amino acid sequences of IPCT and DIPPS encoded by separated genes (*Thermotoga* spp., *H. butylicus*, *P. fumarii* and *A. pernix*), were artificially fused.

The amino acid sequences were aligned with the Muscle program using the default parameters (Edgar, 2004), and the alignment was refined manually and using the TrimAl program (Capella-Gutierrez *et al.*, 2009). The same approach was used to reconstruct the phylogenetic analysis of the two domains separately. The 16S rRNA sequences used in the

construction of the phylogenetic tree were retrieved from SILVA (http://www.arb-silva.de/) (Pruesse *et al*., 2007).

Phylogenetic analyses

The model of protein evolution that best fits the data was determined using ProtTest 2.4 (Abascal *et al*., 2005). Maximum likelihood (ML) phylogenetic analyses were performed by PHYML 3.0 (Guindon and Gascuel, 2003) using a LG model with a gamma correction (eight discrete classes, an estimated amino acid frequency, with a proportion of invariant sites). Bootstrap analyses (1000 replicates) were performed using PHYML 3.0. Bayesian analyses were performed using MrBayes v.3.0b4 (Ronquist and Huelsenbeck, 2003) with a mixed amino acid model of substitution and gamma correction (eight discrete categories plus invariant sites), to take into account among-site rate variations. MrBayes run with four chains for one million generations and trees were sampled every 100 generations. Consensus tree was built after the first 1500 trees were discarded.

The codon usage of the IPCT/DIPPS sequences was compared with the codon usage table of the organism, when available, using the graphical codon usage analyser (http://gcua.schoedl.de/).

Acknowledegments

This work was funded by the Fundação para a Ciência e a Tecnologia (FCT), POCTI, Portugal, PTDC/BIA-MIC/71146/2006. This work was partially supported by Grant BFU2009-13409-C02-01 to H.D. from the Spanish Ministry of Science and Innovation. L.G.G. acknowledges the grant from FCT, Portugal (SFRH/BPD/26905/2006).

References

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

Aravind, L., Tatusov, R.L., Wolf, Y.I., Walker, D.R., and Koonin, E.V. (1998) Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends Genet* **14:** 442–444.

Borges, N., Gonçalves, L.G., Rodrigues, M.V., Ventura, R., Maycock, C., Lamosa, P., and Santos, H. (2006) Biosynthetic pathways of inositol and glycerol phosphodiesters used for stress adaptation in *Archaeoglobus fulgidus*. *J Bacteriol* **188:** 8128–8135.

Borges, N., Matsumi, R., Imanaka, T., Atomi, H., and Santos, H. (2010) *Thermococcus kodakarensis* mutants deficient in di-*myo*-inositol phosphate use aspartate to cope with heat stress. *J Bacteriol* **192:** 191–197.

Boussau, B., Guéguen, L., and Gouy, M. (2008) Accounting for horizontal gene transfers explains conflicting hypotheses regarding the position of Aquificales in the phylogeny of Bacteria. *BMC Evol Biol* **8:** 272.

Brito, J.A., Borges, N., Vonrhein, C., Santos, H., and Archer, M. (2011) Crystal structure of *Archaeoglobus fulgidus* CTP : inositol-1-phosphate cytidylyltransferase, a key enzyme for di-*myo*-inositol-phosphate synthesis in (hyper)-thermophiles. *J Bacteriol* **193:** 2177–2185.

Brochier-Armanet, C., and Forterre, P. (2007) Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene transfers. *Archaea* **2:** 83–93.

Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009) TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25:** 1972–1973.

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

Forterre, P., Bouthier, D.L., Tour, C., Philippe, H., and Duguet, M. (2000) Reverse gyrase from hyperthermophiles: probable transfer of a thermoadaptation trait from archaea to bacteria. *Trends Genet* **16:** 152–154.

Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52:** 696–704.

Heine, M., and Chandra, S.B. (2009) The linkage between reverse gyrase and hyperthermophiles: a review of their invariable association. *J Microbiol* **47:** 229–234.

Kikuchi, A., and Asai, K. (1984) Reverse gyrase-a topoisomerase which introduces positive superhelical turns into DNA. *Nature* **309:** 677–681.

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

Koga, Y., and Nakano, M. (2008) A dendrogram of archaea based on lipid component parts composition and its relationship to rRNA phylogeny. *Syst Appl Microbiol* **31:** 169–182.

Koonin, E.V., Makarova, K.S., and Aravind, L. (2001) Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* **55:** 709–742.

Lamosa, P., Martins, L.O., da Costa, M.S., and Santos, H. (1998) Effects of temperature, salinity, and medium composition on compatible solute accumulation by *Thermococcus* spp. *Appl Environ Microbiol* **64:** 3591–3598.

Lamosa, P., Gonçalves, L.G., Rodrigues, M.V., Martins, L.O., Raven, N.D., and Santos, H. (2006) Occurrence of 1-glyceryl-1-*myo*-inosityl phosphate in hyperthermophiles. *Appl Environ Microbiol* **72:** 6169–6173.

Logsdon, J.M., and Faguy, D.M. (1999) *Thermotoga* heats up lateral gene transfer. *Curr Biol* **9:** R747–R751.

Majumder, A.L., Chatterjee, A., Ghosh Dastidar, K., and Majee, M. (2003) Diversification and evolution of L-*myo*-inositol 1-phosphate synthase. *FEBS Lett* **553:** 3–10.

Martins, L.O., Carreto, L.S., da Costa, M.S., and Santos, H. (1996) New compatible solutes related to di-*myo*-inositol-phosphate in members of the order *Thermotogales*. *J Bacteriol* **178:** 5644–5651.

Michell, R.H. (2008) Inositol derivatives: evolution and functions. *Nat Rev Mol Cell Biol* **9:** 151–161.

Michell, R.H. (2011) Inositol and its derivatives: Their evolution and functions. *Adv Enzyme Regul* **51:** 84–90.

Morii, H., Kiyonari, S., Ishino, Y., and Koga, Y. (2009) A novel biosynthetic pathway of archaetidyl-*myo*-inositol via archaetidyl-*myo*-inositol phosphate from CDP-archaeol

and d-glucose 6-phosphate in methanoarchaeon *Methanothermobacter thermautotrophicus* cells. *J Biol Chem* **284:** 30766–30774.
Morii, H., Ogawa, M., Fukuda, K., Taniguchi, H., and Koga, Y. (2010) A revised biosynthetic pathway for phosphatidylinositol in Mycobacteria. *J Biochem* **148:** 593–602.
Nanavati, D.M., Thirangoon, K., and Noll, K.M. (2006) Several archaeal homologs of putative oligopeptide-binding proteins encoded by *Thermotoga maritima* bind sugars. *Appl Environ Microbiol* **72:** 1336–1345.
Neelon, K., Wang, Y., Stec, B., and Roberts, M.F. (2005) Probing the mechanism of the *Archaeoglobus fulgidus* inositol-1-phosphate synthase. *J Biol Chem* **280:** 11475–11482.
Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., *et al*. (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399:** 323–329.
Nesbø, C.L., L'Haridon, S., Stetter, K.O., and Doolittle, W.F. (2001) Phylogenetic analyses of two 'archaeal' genes in *Thermotoga maritima* reveal multiple transfers between archaea and bacteria. *Mol Biol Evol* **18:** 362–375.
Niederberger, T.D., Gotz, D.K., McDonald, I.R., Ronimus, R.S., and Morgan, H.W. (2006) *Ignisphaera aggregans* gen. nov., sp. nov., a novel hyperthermophilic crenarchaeote isolated from hot springs in Rotorua and Tokaanu, New Zealand. *Int J Syst Evol Microbiol* **56:** 965–971.
Noll, K.M., Lapierre, P., Gogarten, J.P., and Nanavati, D.M. (2008) Evolution of mal ABC transporter operons in the *Thermococcales* and *Thermotogales*. *BMC Evol Biol* **8:** 7.
Nunes, O.C., Manaia, C.M., da Costa, M.S., and Santos, H. (1995) Compatible solutes in the thermophilic bacteria *Rhodothermus marinus* and '*Thermus thermophilus*'. *Appl Environ Microbiol* **61:** 2351–2357.
Pruesse, E., Quast, C., Knittel, K., Fuchs, B., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35:** 7188–7196.
Rodionov, D.A., Kurnasov, O.V., Stec, B., Wang, Y., Roberts, M.F., and Osterman, A.L. (2007) Genomic identification and in vitro reconstitution of a complete biosynthetic pathway for the osmolyte di-*myo*-inositol-phosphate. *Proc Natl Acad Sci USA* **104:** 4279–4284.
Rodrigues, M.V., Borges, N., Henriques, M., Lamosa, P., Ventura, R., Fernandes, C., *et al*. (2007) Bifunctional CTP : inositol-1-phosphate cytidylyltransferase/CDP inositol : inositol-1-phosphate transferase, the key enzyme for di-*myo*-inositol-phosphate synthesis in several (hyper)thermophiles. *J Bacteriol* **189:** 5405–5412.
Rodrigues, M.V., Borges, N., Almeida, C.P., Lamosa, P., and Santos, H. (2009) A unique β-1,2-mannosyltransferase of *Thermotoga maritima* that uses di-*myo*-inositol phosphate as the mannosyl acceptor. *J Bacteriol* **191:** 6105–6115.
Ronquist, F., and Huelsenbeck, J.P. (2003) MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:** 1572–1574.
Santos, H., Lamosa, P., Borges, N., Gonçalves, L.G., Pais, T.M., and Rodrigues, M.V. (2011) Organic compatible solutes of prokaryotes that thrive in hot environments: the importance of ionic compounds for thermostabilization. In *Extremophiles Handbook*. Horikoshi, K., Antranikian, G., Bull, A.T., Robb, F.T., and Stetter, K.O. (eds). Tokyo: Springer, pp. 497–520.
Scholz, S., Sonnenbichler, J., Schäfer, W., and Hensel, R. (1992) Di-*myo*-inositol-1,1′-phosphate: a new inositol phosphate isolated from *Pyrococcus woesei*. *FEBS Letters* **306:** 239–242.
Shibata, T., Nakasu, S., Yasui, K., and Kikuchi, A. (1987) Intrinsic DNA-dependent ATPase activity of reverse gyrase. *J Biol Chem* **262:** 10419–10421.
Silva, Z., Borges, N., Martins, L.O., Wait, R., da Costa, M.S., and Santos, H. (1999) Combined effect of the growth temperature and salinity of the medium on the accumulation of compatible solutes by *Rhodothermus marinus* and *Rhodothermus obamensis*. *Extremophiles* **3:** 163–172.
Swoboda, J.G., Campbell, J., Meredith, T.C., and Walker, S. (2010) Wall teichoic acid function, biosynthesis, and inhibition. *Chembiochem* **11:** 35–45.
Yanai, I., Wolf, Y.I., and Koonin, E.V. (2002) Evolution of gene fusions: horizontal transfer versus independent events. *Genome Biol* **3:** research0024.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. A multialignment of the IPCT/DIPPS and the concatenated IPCT and DIPPS sequences with Muscle programme and TrimAl was used to curate the multialignment. Colouring of the multialignment has been made with clustalx matrix.
Fig. S2. Bayesian phylogenetic unrooted tree of the IPCT/DIPPS proteins, constructed on the basis of amino acid sequences. The tree was calculated by MrBayes. Numbers at nodes are posterior probabilities computed by MrBayes. The scale bar represents the number of substitutions per site. Bacteria (red), Euryarchaeota (green), Crenarchaeota (blue), and Korarchaeota (purple).
Fig. S3. Comparison of IPCT/DIPPS (left) and 16S rRNA (right) phylogenetic trees of the organisms considered in this study. Disagreements between the two phylogenetic trees are highlighted with dotted lines. Bacteria (red), *Euryarchaeola* (green), *Crenarchaeota* (blue), and *Korarchaeota* (purple). IPCT and DIPPS GenPept accession numbers are indicated.
Fig. S4. Pathways for the synthesis of di-*myo*-inositol-1,3′-phosphate-1′-phosphate, glycero-phospho-*myo*-inositol-phosphate (A), and phosphatidyl-*myo*-inositol-phosphate (B). Enzymes: DIPPS, di-*myo*-inositol-1,3′-phosphate-1′-phosphate synthase and PIPS, phosphatidyl-*myo*-inositol-phosphate synthase.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.