Insight into the evolution of microbial metabolism from the deep-branching bacterium, *Thermovibrio ammonificans*

Donato Giovannelli[1,2,3,4*], Stefan M. Sievert[5], Michael Hügler[6], Stephanie Markert[7], Dörte Becher[8], Thomas Schweder [8], and Costantino Vetriani[1,9*]

[1]Institute of Earth, Ocean and Atmospheric Sciences, Rutgers University, New Brunswick, NJ 08901, USA

[2]Institute of Marine Science, National Research Council of Italy, ISMAR-CNR, 60100, Ancona, Italy

[3]Program in Interdisciplinary Studies, Institute for Advanced Studies, Princeton, NJ 08540, USA

[4]Earth-Life Science Institute, Tokyo Institute of Technology, Tokyo 152-8551, Japan

[5]Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA

[6]DVGW-Technologiezentrum Wasser (TZW), Karlsruhe, Germany

[7]Pharmaceutical Biotechnology, Institute of Pharmacy, Ernst-Moritz-Arndt-University Greifswald, 17487 Greifswald, Germany

[8]Institute for Microbiology, Ernst-Moritz-Arndt-University Greifswald, 17487 Greifswald, Germany

[9]Department of Biochemistry and Microbiology, Rutgers University, New Brunswick, NJ 08901, USA

*Correspondence to:

Costantino Vetriani
Department of Biochemistry and Microbiology
and Institute of Earth, Ocean and Atmospheric Sciences
Rutgers University
71 Dudley Rd
New Brunswick, NJ 08901, USA
+1 (848) 932-3379
vetriani@marine.rutgers.edu

Donato Giovannelli
Institute of Earth, Ocean and Atmospheric Sciences
Rutgers University
71 Dudley Rd
New Brunswick, NJ 08901, USA
+1 (848) 932-3378
giovannelli@marine.rutgers.edu

Abstract

Anaerobic thermophiles inhabit relic environments that resemble the early Earth. However, the lineage of these modern organisms co-evolved with our planet. Hence, these organisms carry both ancestral and acquired genes and serve as models to reconstruct early metabolism. Based on comparative genomic and proteomic analyses, we identified two distinct groups of genes in *Thermovibrio ammonificans*: the first codes for enzymes that do not require oxygen and use substrates of geothermal origin; the second appears to be a more recent acquisition, and may reflect adaptations to cope with the rise of oxygen on Earth. We propose that the ancestor of the *Aquificae* was originally a hydrogen oxidizing, sulfur reducing bacterium that used a hybrid carbon fixation pathway for CO_2 fixation. With the gradual rise of oxygen in the atmosphere, more efficient terminal electron acceptors became available and this lineage acquired genes that increased its metabolic flexibility while retaining ancestral metabolic traits.

Keywords: *Aquificae*, carbon fixation, rTCA, comparative genomic, proteomic

Introduction

Deep-branching, anaerobic, thermophilic *Bacteria* and *Archaea* inhabit relic environments that resemble the early Earth (Baross and Hoffman, 1985; Martin et al., 2008). Thermophily (Di Giulio, 2003, 2000), anaerobic metabolism (Baross and Hoffman, 1985; Martin et al., 2008; Schopf, 1983) and reliance on substrates of geothermal origin are among the proposed ancestral traits of these microorganisms (Baross and Hoffman, 1985; Di Giulio, 2003, 2000; Lane et al., 2010; Martin et al., 2008; Russell and Martin, 2004; Schopf, 1983). At the same time, their lineages have co-evolved with Earth and their genomes also carry more recently acquired traits. Therefore, these microorganisms can be used as models to reconstruct the evolution of metabolism.

Thermovibrio ammonificans is part of the phylum *Aquificae*, a deep-branching group of thermophilic bacteria found in geothermal environments (Lebedinsky et al., 2007; Sievert and Vetriani, 2012). Based on phylogenetic analyses of the 16S rRNA gene as well as whole genomes, *Aquificae* are believed to be the earliest bacterial lineages having emerged on Earth along with the phyla *Thermotogae* and *Thermodesulfobacteria* (Di Giulio, 2003; Pitulle et al., 1994; Battistuzzi et al, 2004) (Figure 1, Figure 1-figure supplement 1). All the cultured members of this phylum are chemolithoautotrophs that use hydrogen as an energy source, have optimum growth temperatures between 65˚C and 95˚C and rely on the reductive tricarboxylic acid cycle (rTCA) to convert carbon dioxide into biomass (Hügler et al., 2007; Hügler and Sievert, 2011; Sievert and Vetriani, 2012) (Table 1), making them ideal candidates to investigate the evolution of early metabolism.

The ability to synthesize new biomass from inorganic precursors, *i.e*. autotrophic carbon fixation, is a critical step in the global carbon cycle and is considered a key invention during the early evolution of metabolism (Braakman and Smith, 2012; Fuchs, 2011; Hügler and Sievert, 2011). Among the six known pathways of carbon fixation (reviewed in Fuchs, 2011; Hügler and Sievert, 2011), the rTCA cycle (present in the *Aquificae*, among others) and the reductive acetyl-CoA pathway (present in

acetogenic bacteria and methanogens), represent good candidates for the ancestral carbon fixation pathway (Martin and Russell, 2003; Russell and Hall, 2006). The emergence of a reductive acetyl-CoA pathway has been associated with the FeS-rich minerals at alkaline hydrothermal vents (Martin and Russell, 2007, 2003; Wächtershäuser, 1988) and the presence of homologous core enzymes in both *Bacteria* and *Archaea* potentially support its ancestral nature (Fuchs, 2011; Russell and Martin, 2004). The rTCA cycle is considered the modern version of a proposed prebiotic autocatalytic cycle fueled by the formation of the highly insoluble mineral pyrite in sulfur-rich hydrothermal environments (Wächtershäuser, 1988). The rTCA cycle is widespread among anaerobic and microaerophilic bacteria including all *Aquificae*, chemolithoautotrophic *Epsilonproteobacteria*, *Chlorobi,* and *Nitrospirae* (Berg, 2011; Fuchs, 2011; Hügler and Sievert, 2011) in addition to few other bacterial strains. A recent phylometabolic reconstruction hypothesized that all extant pathways can be derived from an ancestral carbon fixation network consisting of a hybrid rTCA cycle / reductive acetyl-CoA pathway (Braakman and Smith, 2012).

Here, we used a comparative genomic approach coupled to proteomic analyses to reconstruct the central metabolism of *Thermovibrio ammonificans* (Giovannelli et al., 2012; Vetriani et al., 2004), and provide evidence of ancestral and acquired metabolic traits in the genome of this bacterium. We suggest that the last common ancestor within the *Aquificae* possessed the reductive acetyl-CoA pathway in addition to the previously described rTCA cycle. Furthermore, we show that these two pathways may still coexist within the *Desulfurobacteraceae* lineage. The simultaneous presence of both pathways of carbon fixation may represent a modern analog of the early carbon fixation phenotype, and suggests that the redundancy of central metabolic pathways could have been common in ancestral microorganisms.

Results

General genome structure and central metabolism of *T. ammonificans*. The genome of *T.*

ammonificans HB-1 consists of one chromosome and one plasmid of 1,682,965 bp and 76,561 bp, respectively (Figure 2). The genome encodes for 1,890 genes, 1,831 of which are protein coding genes and 75.16% of the protein coding genes could be assigned a putative function (Giovannelli et al., 2012). The chromosome contains three ribosomal RNA operons, the first two with a 5S-23S-16S alignment (coordinates 149,693 – 152,859 bp and 1,168,776 – 1,172,141 bp antisense strand, respectively) and the other with a 16S-23S-5S (coordinates 1,604,527 – 1,609,585 bp). The second copy of the operon is flanked by the largest CRISPR found in the genome (Figure 2, circle 6). Several other repeats were identified in the chromosome (Figure 2, circle 5). The plasmid contained mainly ORFs with unknown hypothetical functions (Figure 2, circle 9) with the exception of a putative RNA polymerase sigma factor of probable archaeal origin, a DNA Topoisomerase I (43.2% similarity to *Hydrogenivirga* sp. 128-5-R1-6), a type II secretion protein E (identified also in the preliminary proteome, Figure 3, Figure 3-source data 1) and an ArsR-like regulatory protein, both of which had homologs among the *Firmicutes*. All other plasmid genes were ORFans with no significant similarities in the database and contained highly repeated DNA (Figure 2, circle 5).

By combining genome-scale analyses with known physiological traits, we were able to reconstruct the metabolic network of *T*. *ammonificans* (Figure 3). As part of the central metabolism of *T*. *ammonificans,* we identified the major pathways for carbon fixation, hydrogen oxidation and nitrate reduction to ammonia, and confirmed them by identifying the corresponding expressed enzymes in the proteome (Figure 3, Figure 3-source data 1). Further, we compared the central metabolism of *T*. *ammonificans* with closely related species within the *Aquificae,* and with ecologically similar *Epsilonproteobacteria*. Comparative genomic analyses with closely related strains provide information about the evolutionary history of the group, while comparison with phylogenetically distant, but ecologically similar species may reveal common adaptive strategies to similar habitats.

The proteome of *T. ammonificans*. 780 proteins, comprising approximately 43% of all

predicted ORFs in the database, were identified in cell samples of *T*. *ammonificans* grown under nitrate reducing conditions (Figure 3, Figure 3-source data 1), indicating a very high congruence between database entries and sample peptides. Among the most abundant proteins were – besides those involved in general housekeeping functions like the translation elongation factor Tu, ribosomal proteins and chaperones – many rTCA cycle enzymes. The four key enzymes of the rTCA cycle (see next paragraph for details) together yielded a relative abundance of 6% of all identified proteins in the samples. Most of the enzymes putatively involved in the reductive acetyl-CoA pathway (Table 2), including the carbon-monoxide dehydrogenase (CodH), were also detected, although at a lower abundance of 1.25% of all proteins.

Carbon Fixation. The genome and proteome of *T*. *ammonificans* (Figure 3) supports previous evidence, based on detection of key genes and measurements of enzyme activities, that carbon fixation occurs via the rTCA cycle (Hügler et al., 2007). The rTCA is widespread among anaerobic and microaerophilic bacteria including all *Aquificae*, chemolithoautotrophic *Epsilonproteobacteria*, *Chlorobi, Nitrospina* and *Nitrospirae,* in addition to few other bacterial strains (*e*.*g*. *Desulfobacter hydrogenophilus*, *Magnetococcus marinus* MC-1 and the endosymbiont of the vent tubeworm, *Riftia pachyptila*; Hügler and Sievert 2011). We identified all the genes responsible for the functioning of the rTCA in the genome and proteome of *T*. *ammonificans* (Figure 3), including the key enzymes fumarate reductase (*frdAB*, gene loci Theam_1270-1275), ATP-citrate (pro-S)-lyase (*aclAB*, 1021-1022), 2-oxoglutarate synthase (*oorABCD*, 1410-1413) and isocitrate dehydrogenase (*idh2*, 1023). *T*. *ammonificans* synthesizes 2-oxoglutarate from succinyl-CoA rather than via citrate, a feature in common with methanogenic *Archaea* and *Desulfurococcales*, and present in numerous strict anaerobes that use the rTCA cycle (Fuchs, 2011). Comparative analyses showed that several genes associated with the rTCA cycle and the gluconeogenesis pathway were highly conserved in all *Aquificae*, with the exception of the genes coding for the ATP-citrate lyase and the isocitrate-dehydrogenase, which are

absent in the *Aquificaceae* (see details below) (Hügler et al., 2007). In general, *T*. *ammonificans* genes had a high percent identity to the genes of *D*. *thermolithotrophum,* its closest relative whose genome sequence is available.

Two separate variants of the rTCA cycle are known (Braakman and Smith, 2012; Hügler et al., 2007). In the two-step variant – also known as symmetric variant, with respect to the enzymatic reaction catalyzed in the two arcs of the rTCA cycle – citrate cleavage is accomplished by the combined action of the enzymes citryl-CoA synthetase (CCS) and citryl-CoA lyase (CCL) and the carboxylation of 2-oxoglutarate is catalyzed by 2-oxoglutarate carboxylase and oxalosuccinate reductase (Aoshima, 2007; Aoshima and Igarashi, 2008; Hügler and Sievert, 2011). This variant of the rTCA cycle was previously proposed as ancestral (Braakman and Smith, 2012; Hügler et al., 2007). By contrast, in the one-step variant of the rTCA cycle, also known as asymmetric variant and more widely distributed, the cleavage of citrate and the carboxylation of 2-oxoglutarate are performed in a one-step fashion (Braakman and Smith, 2012; Hügler et al., 2007)(Figure 3, Figure 3-figure supplement 1). This variant is found in *T*. *ammonificans*.

The genome of *T*. *ammonificans* also codes for an incomplete reductive acetyl-CoA pathway (Figure 3 and Table 2). The key enzyme of this carbon fixation pathway, which is found in acetogens, methanogens, sulfate-reducers and anaerobic ammonium oxidizers (anammox; Berg et al., 2010), is the bifunctional heterotetramer enzyme carbon monoxide dehydrogenase/acetyl-CoA synthase (CODH/ACS). The reductive acetyl-CoA pathway is believed to be the most ancient carbon fixation pathway on Earth (Fuchs, 2011). In *T*. *ammonificans*, the genes coding for the CO-dehydrogenase (*codh* catalytic subunit Theam_1337; Table 2 and Figure 3) is present and also expressed. However, rather than coding for the classical bifunctional type II CodH, *T*. *ammonificans* codes for the type V, which is present in numerous deep-branching organisms (Figure 4) and whose function is unknown (Techtmann et al., 2012). Sequence analysis of the *T*. *ammonificans* CodH protein sequence revealed

the presence of conserved residues necessary for the interaction with Acs and for the coordination of the NiFeS cluster. We hypothesize that the type V CodH, which, based on structural and size considerations may represent the most ancestral version of this class of enzymes (Frank Robb, personal communication), may catalyze the reduction of CO_2 to CO. The *codh* gene of *T*. *ammonificans* shared high similarity with that of *D*. *thermolithotrophum* (87 similarity), while the similarity with the next closest homologous genes dropped below 54%. The genome of *D*. *thermolithotrophum* also codes for a classical bifunctional Type II CodH (Dester_0417 and Dester_0418; Figure 4) in addition to the type V CodH. The two genomes also share a CodH iron-sulfur accessory protein (Theam_0999 and Dester_0418, 81% similarity). Additional laboratory analyses will be required to confirm the role of the type V CodH in carbon fixation.

Gluconeogenesis. *T*. *ammonificans* uses the Embden-Meyerhof-Parnas pathway to synthesize pentose and hexose monosaccharides. We identified in the genome the key genes of this pathway, fructose-1,6-biphosphatase (fbp, Theam_1323; identified also in the proteome, Figure 3) and all other genes necessary for the functioning of the pathway (Figure 3). While the *fbp* gene is present in *Desulfurobacterium thermolithotrophum*, and in the genome of *Sulfurohydrogenibium* sp. YO3AOP1, *Persephonella marina* and *Hydrogenivirga* sp. 128-5-R1-6, it is absent from the genome of *A*. *aeolicus* (Deckert et al., 1998) and *H*. *thermophilus*, suggesting that in those species an alternative unidentified pathway may be active.

Hydrogen oxidation. Hydrogen is the only electron donor used by *T*. *ammonificans* and its importance in the metabolism of this bacterium is reflected by the diversity of hydrogenases found in the genome, whose encoding genes are organized in two large clusters (Figure 2 circle 9). The first cluster is composed of eight genes encoding for the complete Group 4 Ech hydrogenases (*echABCDEFG,* Theam_0476-0482) and the alpha, beta, gamma and delta subunits of the Group 3b multimeric cytoplasmic Hyd sulfhydrogenase (*hydABCD*, Theam_0484-0487). The second

hydrogenase cluster in the genome codes for the maturation protein HypF (*hypF*, Theam_1116), the three subunit of the Group 3a cytoplasmic F420-reducing hydrogenase Frh (*frhACB*, Theam_1117-1119) and the Group 1 hydrogenases Hyn (*hynABC*, Theam_1121-1123) with associated maturation factors/chaperones Hyp (Theam_1124-1128). We also found other hydrogenase maturation factors (Theam_0922-0924) and the formate dehydrogenase Fdh (*fdhABC*, Theam_1020 and Theam_0966-0967 respectively) scattered in the genome.

We reconstructed the phylogeny of *T*. *ammonificans* [NiFe]-hydrogenases (Figure 5), and we assigned them to known classes of hydrogenases (Groups 1 through 4) according to their phylogenetic position in the tree and the presence of known conserved amino acid motifs typical of each group (Vignais and Billoud, 2007). Group 1 [NiFe]-hydrogenases are normally coupled to anaerobic respiration through electron transfer to the membrane quinone and menaquinone pool. Comparative analyses of the hydrogenases revealed that the Group 1 cluster in *T*. *ammonificans* (Hyn) is more similar to the hydrogenases found in *Deltaproteobacteria* than to homologs in other *Aquificae* (Figure 5). This finding raises interesting questions about the origin of the hydrogen oxidation Group 1 enzymes in *T*. *ammonificans* and suggests that an event of lateral gene transfer occurred between *T*. *ammonificans* and the *Deltaproteobacteria*. Further, the periplasmic nitrate reductase (*napA*) and the *nap* operon structure in *T*. *ammonificans* also appears to be closely related to that of *Deltaproteobacteria*, suggesting that a horizontal gene transfer event involved the entire hydrogen oxidation (Hyn)/nitrate reduction respiratory pathway (see discussion).

Contrarily to members of the *Aquificaceae* and *Hydrogenothermaceae*, *T*. *ammonificans* and *D*. *thermolithotrophum* genomes code for both Group 3 cytoplasmic hydrogenases, Group 3a F420-reducing hydrogenases (*frhACB*) and Group 3b multimeric cytoplasmic sulfhydrogenases (*hydABCD*; Figure 5; Jeon et al., 2014). Comparative analyses of these two clusters revealed that other members of the *Desulfurobacteriaceae* family were the closest relatives, while the next closest enzymes were those

found in hyperthermophilic *Euryarchaeota* (*i.e*. *Pyrococcus abyssi*) and in methanogens (Figure 5). Overall, these results imply a polyphyletic origin for the group 3 cytoplasmic hydrogenases. Comparative analyses revealed similarities with hyperthermophilic *Euryarchaeota* and methanogens (Figure 5), suggesting a possible common ancestry for the group 3 cytoplasmic hydrogenases of *Archaea* and *Desulfurobacteriaceae*. Together with Group 3b sulfhydrogenases and Group 4 Ech, F420-reducing hydrogenases are in fact typically found in methanogens in which the F420-cofactor is used as an electron carrier in the reduction of CO_2 or other C1 compounds to methane (Vignais and Billoud, 2007). Cultures of *T*. *ammonificans* grown with nitrate as the terminal electron acceptor did not present the typical autofluorescence of methanogens due to F420-cofactor fluorescence when observed under UV in epifluorescence microscopy and the genome lacks the gene involved in the biosynthetic pathway of the F420-coenzyme. Due to these observations, we suggest that these hydrogenases may work in conjunction with ferredoxin in *T*. *ammonificans*. Group 3b Hyd sulfhydrogenases appear to be of mixed origins despite being present in a single operon. Three of the subunits share similarities with the *Thermococcales* (*Euryarchaeota; hydBCD*) while the catalytic subunit (HydA) appears to be unique to the *Desulfurobacteriacea*, not having any other significant homolog in the database (Figure 6). Comparative analyses indicated that Ech shared similarity with *Epsilonproteobacteria* Group 4 hydrogenases (Figure 5). Interestingly, Group 2 hydrogenases are found in microaerobic members of the *Aquificaceae* and *Hydrogenothermaceae* and in microaerobic *Epsiloproteobacteria* but are not found in anaerobes, suggesting a possible link with oxygen adaptation. In contrast, Group 3a is found in strict anaerobes, including *T*. *ammonificans* and methanogens (Figure 5).

Finally, Ech hydrogenases are the only Group 4 hydrogenases found in *T*. *ammonificans*. Both the reductive TCA cycle and the reductive acetyl-CoA pathway involve reaction steps that are driven by reduced ferredoxin (Fuchs, 2011). In the case of *T*. *ammonificans,* the enzymes formate

dehydrogenase, CO-dehydrogenase, pyruvate synthase and 2-oxoglutarate synthase require reduced ferredoxin. We hypothesize that *T*. *ammonificans* can accomplish ferredoxin reduction in two different ways: the first involves cytosolic hydrogenases (Group 3, Frh and Hyd; Figure 3) commonly found in methanogens, where they reduce ferredoxin via electron bifurcation (Fuchs, 2011); the second involves membrane bound hydrogenases (Group 4, Ech; Figure 3). With the exception of the F420-reducing (Frh) hydrogenases, representatives of the remaining three groups were identified in the proteome of *T*. *ammonificans* (Figure 3).

Nitrate ammonification. *T*. *ammonificans* HB-1 can use nitrate as a terminal electron acceptor and reduces it to ammonium (Vetriani et al., 2004). Periplasmic nitrate reductases (Nap) have been studied extensively in some *Proteobacteria*, and are comprised of the catalytic enzymes – typically a heterodimer (NapAB) – associated with various other subunits involved in channeling electrons to the NapA reactive center (Moreno-Vivián et al., 1999; Vetriani et al., 2014). In *T*. *ammonificans,* the genes encoding for the Nap complex are organized in a single operon, *napCMADGH*. Comparative genomic analyses with genomes from *Aquificae* and *Proteobacteria* indicate that the *nap* operon structure of *T*. *ammonificans* is unique to this organism, and shares similarity with *Desulfovibrio desulfuricans* and other *Deltaproteobacteria*. All other *Aquificae* whose genomes are available have a Nap operon similar to that of *Epsilon-* and *Gammaproteobacteria* (Vetriani et al., 2014). Moreover, unlike other members of the *Aquificae*, the nitrate reductase of *T*. *ammonificans* (encoded by gene Theam_0423, expressed in the proteome; Figure 3) appears to be of the monomeric type (*sensu*; Jepson et al., 2006), and the NapB subunit is missing (Figure 3). We also identified a putative nitrite reductase-encoding gene possibly involved in the reduction of nitrite to ammonium (NirA; Theam_1000, Figure 3), which was also expressed in the proteome (Figure 3). However, the exact mechanism through which *T*. *ammonific*ans reduces nitrite to ammonium remains to be experimentally elucidated.

Sulfur reduction. While *T*. *ammonificans* is able to reduce elemental sulfur to hydrogen sulfide

(Vetriani et al., 2004), the sulfur reduction pathway remains unclear. We identified in the genome a putative polysulfide reductase of the sulfide-quinone reductase family (*sqr*, Theam_0841, Figure 3). We propose that this membrane-bound complex can utilize polysulfide formed from the reaction of S^0 with sulfide (naturally enriched in hydrothermal systems) via quinone oxidation (Figure 3). The putative *sqr* gene of *T*. *ammonificans* shares similarities with other *Aquificae* and *Epsilonproteobacteria* (average 56% similarity to both groups). In particular, the *sqr* gene appears to be conserved also in *Caminibacter mediatlanticus* (Figure 7; Giovannelli et al., 2011), whose sulfur reduction pathway is yet to be elucidated. Another possibility is that *T*. *ammonificans* uses a NAD or FAD-dependent reductase (*nsr*, Theam_1321) to reduce sulfur, a mechanism previously described in the sulfur-reducing archaeon *Pyrococcus furiosus* (Schut et al., 2013, 2007). However, the putative polysulfide reductase of *T*. *ammonificans* has only about 33% identity to homologous enzymes of sulfur reducing bacteria and archaea identified in the database.

Resistance to oxidative stress. *T*. *ammonificans* is a strict anaerobe. However, its genome codes for genes involved in the detoxification of oxygen radicals, including a catalase/peroxidase (Theam_0186), whose activity was previously detected (Vetriani et al., 2004), a putative superoxide reductase (Theam_0447), a cytochrome-C peroxidase (Theam_1156) and a cytochrome bd complex (Theam_0494-0496, Figure 3). The latter three enzymes were identified in the proteome (Figure 3). The cythochrome bd complex has been shown to contribute to oxygen tolerance in other anaerobic bacteria (Das et al., 2005), and to contribute to the detoxification of nitrous oxide radicals (Mason et al., 2009).

Energy conversion. The genome of *T*. *ammonificans* contains two gene clusters coding for the NADH dehydrogenase (Theam_0731-0745 and Theam_1488-1498). The genes appear to have undergone an inversion during the duplication process and the second copy is missing 4 genes in the middle of the operon (*nuoEFG* and a transcriptional regulator). The two clusters share on average 60%

gene similarity. Only one copy of the gene for the synthesis of ATP synthetase complex was present (Theam_1605-1606 and Theam_1656-1662). Three different hydrogen-sodium symporters (*nhe*, Theam_0443, 0413 and 0647) were present with low pairwise similarity. NADH dehydrogenase second cluster and ATP synthetase complex appear to be conserved in the close relative *D. thermolithotrophum* and in the *Epsilonproteobacterium, C. mediatlanticus* TB-2 (Figure 3).

Motility, cell sensing and biofilm formation. *T. ammonificans* posses one to two terminal flagella (Vetriani et al., 2004). The genes involved in flagella formation and assembly are prevalently organized in a single large cluster in the genome (from Theam_1440 to Theam_1468, Figure 7), putatively organized in three distinct operons. The hook-filament junction proteins genes *flgK* and *flgL* (Theam_1384 and Theam_1385), the filament and filament cap proteins *fliC* and *fliD* (Theam_1160 and Theam_1162) together with few flagellar maturation factors and motor switch genes (*motA* and *motB*, Theam_1087 and Theam_1085-1086, respectively) are instead scattered in the genome. The entire cluster of genes shares similarities with *D. thermolitotrophum,* although the similarity for some of the proteins is as low as 50%. Despite this, comparative analyses revealed a similar organization in *D. thermolitotrophum* where the cluster appears inverted in the same relative genomic position and constituted one of the main region of synteny (Figure 7b). By contrast, the flagellar genes in the *Epsilonproteobacterium, Caminibacter mediatlanticus* TB-2, are scattered throughout the genome (Figure 7b). Numerous genes for adhesion and pili were present, generally organized in small clusters (Figure 7b).

Four different putative methyl-accepting chemotaxis sensory transducer proteins were found in the genome (*mcp*, Theam_0157, 0165, 0845 and 1027). One of those, Theam_0845, is surrounded by receptors *cheW* and *cheA*, and by the response regulator *cheY*. The entire group is organized in a single operon, *cheYVBWAZ* conserved in *D. thermolithotrophum* and other *Aquificales* with the exception of *cheW,* which shares higher similarities with the *Epsilonproteobacterium, C. mediatlanticus* TB-2

(Figure 6). We failed to find in the genome of *T. ammonificans* a homolog of the primary receptor encoding gene *cheR* (Wadhams and Armitage, 2004). We identified in the genome numerous membrane transporters for molybdenum (*modCBA*, Theam_0787-0789), iron (*fhuDBC*, Theam_0192-0194), zinc (*znuAB*, Theam_0238 and 0689) and cobalt/nickel (*cbiOQMN*, Theam_1522, 1523, 1525 and 1526) all involved in the uptake of important micro-elements for enzyme and cofactor synthesis. We also identified the complete type II secretion pathway, which is conserved in gram-negative bacteria and responsible for protein and toxin translocation to the extracellular milieu. Proteins secreted by the type II pathway include proteases, cellulases, pectinases, phospholipases, lipases, toxins and in some cases type III and IV pili (Douzi et al., 2012). In some bacteria, the expression of secretion type II pathway genes are regulated either by quorum-sensing or by the environmental factors at the site of colonization (Douzi et al., 2012; Sandkvist, 2001). Biofilm formation in bacteria is often linked to quorum-sensing mechanisms, regulating the settlement of planktonic cells in relationship to environmental sensing, substrate suitability and cell densities. It is likely that *T. ammonificans* maintains a mostly attached lifestyle in hydrothermal vent environments due to the high turbulence associated with fluid flux and mixing with seawater. We speculate that the type II pathway genes, together with chemotaxis and flagellar/adhesion genes may play a key role in the formation of biofilm. Extracellular secreted proteins may in fact be used to "condition" the colonized substrate and to interact/remodel existing and newly formed biofilms.

Activated methyl cycle. The activated methyl cycle is a conserved pathway in which S-methyl groups from L-methionine are converted into a biochemically reactive form through insertion into an adenosyl group (Vendeville et al., 2005). S-adenosyl-L-methionine provides activated methyl groups for use in the methylation of proteins, RNA, DNA and certain metabolites central to the core cell machinery. The activated methyl cycle exists in two recognized forms, involving respectively *luxS* or *sahH* genes (Vendeville et al., 2005). Genome-based reconstruction of the activated methyl cycle

indicated that, in *T*. *ammonificans,* involves the *sahH* gene, while a *luxS* homolog is absent (Figure 3-figure supplement 2). Methionine methyl groups are adenosylated by a methionine adenosyltransferase (*samS*, Theam_1075). The methyl group is thus activated and readily available for the S-adenosylmethionine-dependent methyltransferase (*samT*) for methylation of substrate. In *T*. *ammonificans, samT* is a pseudogene, as it contains two stop codons (Theam_0630). Homologs of *samT* are absent in the genome of the other *Aquificae* and the pathway appears to be incomplete. Only in *P*. *marina* and *D*. *themolithotrophum* the functionally equivalent DNA-cytosine methyltransferase (*dcm*) is present. The presence of a gap in such a central pathway in the *Aquificae* and the retrieval of a *samT*-like pseudogene in *T*. *ammonificans* raises interesting question on the functioning of the activated methyl cycle in those organisms. An interesting working hypothesis is that enzyme thermal stability at the physiological temperature of these bacteria might have driven the loss of *samT* and the replacement by a not yet identified alternative enzyme in most *Aquificae*.

Discussion

Evidence for a reductive acetyl-CoA pathway in t*he Desulfurobacteraceae* lineage. The reductive acetyl-CoA pathway is present in acetogenic bacteria, methanogens, sulfate-reducers and anammox bacteria (Berg, 2011; Fuchs, 2011; Hügler and Sievert, 2011). It is believed to be the oldest carbon fixation pathway on Earth, due to the centrality of acetyl-CoA in metabolism, the low energy requirement (~1 ATP required *vs* 2-3 ATP for the rTCA cycle) and the limited necessity of *de novo* protein assembly (Berg et al., 2010; Fuchs, 2011). We hypothesize that the reductive acetyl-CoA pathway is used as an additional or alternative pathway to fix CO_2 in *T*. *ammonificans*.

The finding of a potentially active reductive acetyl-CoA pathway (reactions 1 to 6; Figure 3) in *T*. *ammonificans* and other members of the *Desulfurobacteraceae* is intriguing. The pathway is missing reaction 2, the formyl-THF synthesis for which no obvious 10-formyl-THF synthetase was found (Figure 3). However, the 10-formyl-THF synthetase of the reductive acetyl-CoA pathway may have

been replaced by alternative enzymes, such as 5-formyl-THF cycloligase (Theam_1206) or 10-formyl-THF deformylase (Theam_0826) working in reverse (Table 2, Figure 3). Promiscuous enzymes that function in more than one pathway have been long hypothesized and recently described in *Escherichia coli* (Jensen, 1976; Kim et al., 2010). Assuming low substrate specificity, other candidate enzymes with similar substrate requirements are the phosphoribosylglycinamide formyltransferase (Theam_1211) or phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase (Theam_0328), which could catalyze the synthesis of formyl-THF, albeit with sub-optimal kinetics. While all four enzymes were detected in the proteomic analysis (Figure 3), we think that the possible involvement of a 5-formyl-THF cycloligase in the synthesis of formyl-THF in *T*. *ammonificans* is particularly appealing, as this would provide an evolutionary link between the N^5-formate uptake of methanogens and N^{10}-formate uptake in acetogens (Braakman and Smith, 2012). Despite the absence of the 10-formyl-THF synthetase, the genome of *T*. *ammonificans* encodes and expresses a type V CodH (Figure 3).

We investigated the available genomes of other members of the *Aquificae* for the presence of genes homologous to the putative reductive acetyl-CoA pathway of *T*. *ammonificans* (Table 2). The similarity of the reductive acetyl-CoA pathway-related genes of *T*. *ammonificans* to those from organisms outside the *Desulfurobacteriaceae* is low. The type V *codH* gene identified in *T*. *ammonificans* is found exclusively in members of the family *Desulfurobacteriaceae*, with exception of a low similarity homolog present in *Persephonella marina* (41% amino acid identity) (Table 2). Furthermore, most of the genes involved in reaction 3, 4 and 5 (Figure 3) have homologs in other *Aquificae* or are substituted by functional equivalents with low amino acid identity (< 40%). To understand the relationship of the *T*. *ammonificans* CodH with the homologous enzymes of methanogens, homoacetogens and sulfate-reducers, we reconstructed the phylogenetic history of these enzymes with a particular focus on type V CodH (Figure 4). Sequences from the

393 *Desulfurobacteriaceae* appear to be related to those present in the *Archaea, Ferroglobus placidus* and
394 *Archaeoglobus* spp., and similar to *Methanosarcina* spp. and homoacetogenic bacteria. The CodH
395 found in CO-utilizing bacteria is only distantly related to the one found in *T*. *ammonificans* (Techtmann
396 et al., 2012), while the CodH sequence of *Persephonella marina* appears to be one of a kind, with the
397 enzyme of *Ferroglobus placidus* as its closest relative (44% similarity). Taken together, these findings
398 suggest that, within the *Aquificae*, the type V CodH is exclusive to the *Desulfurobacteriaceae*.

399 The retrieval of the CodH enzyme from the proteome of *T*. *ammonificans* (Figure 3, Figure 3-
400 source data 1), along with the catalytic subunit of the formate dehydrogenase (FdhA), suggests that the
401 reductive acetyl-CoA pathway could be operational in *T*. *ammonificans* in addition to the rTCA cycle.
402 The simultaneous presence of two distinct carbon fixation pathways has been confirmed so far only in
403 the endosymbiont of the deep-sea vent tubeworm *Riftia pachyptila* (Markert et al., 2007). Recent
404 findings suggest that the *Riftia* symbiont can switch from the Calvin-Benson-Bassham (CBB) to the
405 rTCA cycles in response to low-energy supply (*e*.*g*., low concentrations of hydrogen sulfide in the vent
406 fluids, being the rTCA cycle more energetically favorable than the CBB cycle; Markert et al., 2007).
407 Genomic evidence of the simultaneous presence of different carbon fixation pathways came also from
408 genome analysis of *Ammonifex degensii* and *Ferroglobus placidus* (Berg, 2011; Hügler and Sievert,
409 2011), where the reductive acetyl-CoA pathway seems to be coupled to an incomplete CBB cycle. The
410 use of different carbon fixation strategies could be advantageous under energy limiting conditions (*e*.*g*.,
411 deep biosphere), could optimize overall carbon fixation or provide different precursors for biosynthetic
412 pathways, and could be more widespread than originally thought.

413 Further laboratory analyses will identify the exact role of the type V CodH in *T*. *ammonificans*
414 and the potential contribution of the reductive acetyl-CoA to overall carbon fixation.

415 **Comparative genomic of *T*. *ammonificans*: ancestral and acquired metabolic traits.** The
416 genome of *T*. *ammonificans* HB-1 displays a large degree of mosaicism (Figure 6). When comparing

417 the coding sequences (CDS) of *T*. *ammonificans* with available genomes, 34% of the CDS shared
418 higher similarity to genes outside the *Aquificae*, suggesting that these genes were acquired from more
419 distantly related taxa.

420 We carried out comparative genomic analyses among *T*. *ammonificans* and all available
421 *Aquificae* genomes (Table 3). *Desulfurobacterium thermolithotrophum* DSM 11699 (its closest
422 relative whose genome was sequenced) and *Caminibacter mediatlanticus* TB-2 (an
423 *Epsilonproteobacterium* which, albeit phylogenetically distant, shares the same physiology and
424 occupies a similar ecological niche as *T*. *ammonificans*, only at lower temperature; Figure 7) were
425 further used as a direct comparison. We identified areas of synteny between *T*. *ammonificans* and the
426 closely related *D*. *thermolithotrophum* in the region surrounding the genes encoding for key enzymes
427 responsible for the citrate cleavage in the rTCA cycle (Figure 7b and Figure 8). This region is
428 conserved also within the *Hydrogenothermaceae* (in particular *P*. *marina*) and members of the
429 *Epsilonproteobacteria*.

430 The conserved regions between the genomes of *T*. *ammonificans* and *D*. *thermolithotrophum*
431 included the two hydrogenase clusters, the flagellum and the NADP dehydrogenase complex. The
432 order and position of the hydrogenase and NADP dehydrogenase clusters were highly conserved also in
433 *C*. *mediatlanticus*, while flagellar genes were scattered throughout the genome (Figure 7). These results
434 suggest that the genes encoding hydrogen utilization and functions related to energy conversion are
435 among the oldest genomic regions present in those organisms, and overall are conserved across phyla.
436 When we analyzed the central metabolism of *T*. *ammonificans* using comparative genomic approaches,
437 the presence of two distinct groups of genes became evident: the first group was related to early-
438 branching bacterial or archaeal lineages and coded for enzymes involved in several central metabolic
439 pathways, while the second group of genes appeared to have been acquired later. Among the first group
440 of genes, we identified: (I) the cytoplasmic [Ni-Fe]-hydrogenases Group 3 that were related to the

441 enzymes found in methanogens and thermophilic *Euryarchaeota,* in which they are involved in
442 ferredoxin reduction (Figure 3 and Figure 8); (II) sulfur-reduction related genes (Figure 3); and (III) the
443 genes coding for the enzymes of the reductive acetyl-CoA pathway (Figure 3). These pathways are
444 either present in early branching *Archaea*, or are directly involved in metabolic reactions that do not
445 require oxygen (or oxygen by-products) and use substrates of geothermal origin that are likely to have
446 been abundant on Early Earth. Moreover, most key reactions in these pathways are catalyzed by
447 enzymes that are extremely sensitive to oxygen. Our findings are consistent with a recent
448 reconstruction of the ancestral genome of the last universal common ancestor (LUCA), which suggest
449 that LUCA was a hydrogen-dependent autotroph capable of S utilization that could fix CO2 via the
450 reductive acetyl-CoA pathway (Weiss et al., 2016). Taken together, these observations suggest that part
451 of the central metabolism of *T*. *ammonificans* is ancestral and emerged prior to oxygenic
452 photosynthesis.

453 By contrast, some genes and associated metabolic pathways appear to have been acquired by
454 the *T*. *ammonificans* lineage at a later time. One example is the ability to conserve energy by nitrate
455 reduction. The hypothesis that nitrate respiration is an acquired trait in *T*. *ammonificans* is supported by
456 the structure of the *nap* operon. The monomeric *napA* gene of *T*. *ammonificans* is homologous to that
457 of *Desulfovibrio desulfuricans* and other *Deltaproteobacteria* (Figure 6; online interactive version) and
458 apparently has been acquired laterally. Furthermore, it shares high similarity with assimilatory nitrate
459 reductases (Nas) and it could be an evolutionary intermediate form between assimilatory Nas and the
460 dimeric respiratory Nap. In line with this observation, phylogenetic analyses suggest that the membrane
461 bound [Ni-Fe]-hydrogenases of Group 1 and 4 are of delta/epsilonproteobacterial origin (Figure 5).
462 Based on the finding that Group 1 Hyn hydrogenases were expressed when *T*. *ammonificans* was
463 grown under nitrate-reducing conditions (Figure 3), we suggest that these enzymes are linked through
464 the membrane quinone pool to nitrate reduction and may have been acquired simultaneously with the

nap genes.

A second example is represented by the oxygen radical detoxification enzymes encoded by the genome of *T*. *ammonificans*. We propose that such genes are not part of the core, or ancestral genome of *T*. *ammonificans*, but that they evolved as a response to exposure to toxic oxygen radicals. In the modern ocean, catalase and peroxidase may provide protection to oxygen-sensitive enzymes during transient exposure to oxygen associated with entrainment of deep seawater in hydrothermal fluids, or when the organism is displaced into the water column. This second scenario may have important implication for the dispersal of vent microorganisms to other vent sites. We hypothesize that *T*. *ammonificans* is able to deal with oxygen exposures using two different mechanisms: the first response mechanism would be activated following exposure to oxygen at physiological temperatures (60 – 80 °C). In that case, the oxygen-stress related genes are induced, protecting the cell machinery against damages. The second mechanism would be a response to prolonged exposure to oxygen at temperatures below the physiological threshold. Such prolonged exposure would trigger a metabolic shut-down, enabling survival of the organism in cold seawater. The latter hypothesis is supported by experiments where batch cultures of *T*. *ammonificans* were exposed for a prolonged time at 4°C, which revealed the survival of the bacterium despite being exposed to oxygen (data not shown).

In conclusion, the nitrate reduction pathway, as well as oxygen stress-related genes, may represent adaptations acquired by the *T*. *ammonificans* lineage to cope with the rise of oxygen on Earth.

Insight into the evolution of carbon fixation. Six different pathways of carbon fixation are known to date (Fuchs, 2011). In the modern biosphere, the Calvin-Benson-Bassham cycle is the dominant mechanism of CO_2 fixation. Yet, other anaerobic pathways have been investigated and are proposed to represent the ancestral autotrophic carbon fixation pathway. In particular, the reductive acetyl-CoA pathway is thought to be among the earliest carbon fixation pathways that have emerged on

Earth (Fuchs, 2011). This hypothesis is supported by numerous observations: (I) the presence of this pathway in the early branching methanogenic archaea; (II) its low energy requirements and low need of *de novo* protein synthesis; (III) its capacity to incorporate different one-carbon compounds and carbon monoxide of geothermal origin, and (IV) the extreme oxygen sensitivity of its key enzymes which have common roots in *Bacteria* and *Archaea* (reviewed in Berg et al., 2010; Fuchs, 2011; Hügler and Sievert, 2011). Despite differences in their structure, all extant carbon fixation pathways can be theoretically derived from a putative rTCA cycle/reductive acetyl-CoA hybrid pathway, as proposed by a recent phylometabolic reconstruction of the evolution of carbon fixation (Braakman and Smith, 2012). According to this study, both *Aquificae* and acetogenic bacteria represent the closest living example of the archetypal network, and both diverged from the pre-LUCA pathway under the selective pressure of energy-efficiency (acetogens) and oxygen sensitivity (*Aquificae*). Further, a recent reconstruction of the genome of LUCA based on gene phyletic pattern reconstruction is consistent with some of these findings and suggests that LUCA's genomic makeup point to autotrophic acetogenic and methanogenic roots and to the ancestry of the reductive acetyl-CoA pathway (Weiss et al., 2016).

The presence of a CodH type V enzyme in the genome of *T*. *ammonificans* and other members of the *Desulfurobacteriaceae*, and the presence of a type II CodH in *Desulfurobacterium thermolithotrophum* suggest that the reductive acetyl-CoA pathway could be operational in extant members of the *Desulfurobacteriaceae*. Furthermore, this finding supports the scenario proposed by Braakman and Smith that a complete and operational reductive acetyl-CoA pathway was present in the ancestor of the phylum *Aquificae* (Braakman and Smith, 2012). Comparative analyses revealed that the CodH enzyme is conserved only within the *Desulfurobacteriaceae*, consistent with the strict anaerobic nature of the members of this family and the extreme oxygen sensitivity of CodH. The rise of oxygen has been interpreted as one of the factors responsible for the diversification of carbon fixation from the ancestral pathway, and could explain the subsequent loss of the CodH in the generally facultative

513 microaerobic *Hydrogenothermaceae* and *Aquificaceae* within the *Aquificae* (Table 2).

514 Comparative genomic analyses of rTCA cycle genes allow the description of a possible
515 evolutionary scenario for the rTCA cycle in *Aquificae*. Members of the *Aquificaceae* (*e.g., Aquifex*
516 *aeolicus*; Figure 9 and Figure 3-figure supplement 1) have the two-step version of the rTCA cycle
517 (Figure 3-figure supplement 1A), where citrate cleavage is accomplished by the combined action of the
518 enzymes citryl-CoA synthetase and citryl-CoA lyase (encoded by the *ccl* gene), and the carboxylation
519 of 2-oxoglutarate is catalyzed by the two enzymes 2-oxoglutarate carboxylase and oxalosuccinate
520 reductase (Aoshima et al., 2004; Braakman and Smith, 2012). Since citryl-CoA synthetase and 2-
521 oxoglutarate carboxylase likely evolved by duplication of the genes for the succinyl-CoA synthetase
522 and pyruvate carboxylase (Aoshima, 2007; Braakman and Smith, 2012), a complete rTCA cycle could
523 have evolved in the *Aquificaceae* from an ancestral, incomplete version of the cycle. In contrast, the
524 two other groups of *Aquificae*, the *Hydrogenothermaceae* and the *Desulfurobacteriaceae* (Figure 1-
525 figure supplement 1), use the one-step – and more recent – version of the rTCA cycle (Figure 10,
526 Figure 9 and Figure 3-figure supplement 1B), involving ATP citrate lyase (ACL, encoded by the *acl*
527 gene) and isocitrate dehydrogenase, that carry out citrate cleavage and 2-oxoglutarate carboxylation in
528 two single enzyme reactions, respectively (Braakman and Smith, 2012).

529 The enzyme responsible for citrate cleavage, the ATP citrate lyase, likely evolved through gene
530 fusion of the genes of CCS and CCL (Aoshima et al., 2004). Phylogenetic analyses of ACL suggests
531 that this gene fusion event did not happen within the *Aquificae*, as *Nitrospira* and *Chlorobia* have an
532 evolutionary older version of ACL than *Hydrogenothermaceae* and *Desulfurobacteriaceae* (Figure 9;
533 Hügler et al., 2007). Hence, it is likely that these two groups acquired ACL through HGT (Figure 9 and
534 Figure 8; Hügler et al., 2007). Furthermore, our comparative analysis showed that: (I) in the
535 *Hydrogenothermaceae* and in the *Desulfurobacteriaceae*, the enzymes of the first half of the rTCA
536 cycle, as well as aconitase (*acnA*), share a common ancestor with the *Aquificaceae* and can be

considered part of the core genome of the phylum *Aquificae*, while the remaining enzymes are either the result of gene duplication or have been acquired by horizontal gene transfer (Hügler et al., 2007); (II) the two-step citrate cleavage is exclusive to the *Aquificacea* (Figure 9); (III) the *ccl* gene is still present in the *Hydrogenothermacea* (in addition to *acl*), and one sequenced strain of the *Hydrogenothermaceae* (*S. azorense*) still uses the two-step carboxylation of 2-oxoglutarate, suggesting that the two-step version of the rTCA cycle was present in the ancestor of both the *Aquificaceae* and the *Hydrogenothermaceae* (Figure 9 and Figure 10). However, neither of the genes for the two-step citrate cleavage or two-step 2-oxoglutrate carboxylation is present in the *Desulfurobacteriaceae* (Figure 9). Finally, synteny analyses show that the four enzymes necessary to complement the one-step rTCA cycle variant in the *Desulfurobacteriaceae* from a theoretical ancestral linear rTCA are organized in a single operon (Figure 8).

Taking into consideration evidence from comparative genomics and phylogenetic analyses, the most parsimonious interpretation of the data suggests that the last common ancestor of the *Aquifica*e had an incomplete form of the rTCA that did not proceed past the synthesis of 2-oxoglutarate, and that later on the cycle was closed following two independent evolutionary trajectories (Figure 10): (I) Gene duplication in the lineage that lead to the *Aquificaceae* and *Hydrogenothermaceae*; and (II) gene acquisition by horizontal gene transfer in the *Desulfurobacteriaceae* and *Hydrogenothermaceae* (the latter replaced the two-step version of the rTCA cycle with the one-step version) (Figure 9 and figure 10). The reasons behind the presence of two distinct rTCA cycle variants within the extant *Aquificae* are not known. Braackman and Smith hypothesized that the one-step reactions might have evolved as a way to improve the thermodynamic efficiency of the rTCA cycle (Braakman and Smith, 2012). We hypothesize that temperature may also have played a role in preserving the ancient, more symmetric, two-step citrate cleavage rTCA cycle variant in *Aquificaceae* (Figure 10). Members of this group have optimum growth temperatures (75 – 95°C) higher than those of the two other groups (60 – 75°C; Table

561 1), and the "ancient" enzymes might be more stable at these high temperatures. In contrast, the "newer"
562 enzymes that catalyze the one-step citrate cleavage might have evolved to function optimally at lower
563 temperatures.

564 Different members of the *Aquificae* use either one of the two versions of the rTCA cycle, and
565 all the extant members of this phylum possess the genes encoding for the enzymes of the reductive
566 acetyl-CoA pathway, with the exception of *codH* (encoding for the CO-dehydrogenase), which is only
567 found in the *Desulfurobacteriaceae*. Therefore, we hypothesize that the last common ancestor of the
568 *Aquificae* possessed the complete reductive acetyl-CoA pathway (Figure 10). While members of the
569 *Desulfurobacteriaceae* kept the CodH due to their obligate anaerobic lifestyle, microaerophilic
570 *Aquificae* (*Hydrogenothermaceae* and *Aquificaceae*) lost this extremely oxygen-sensitive enzyme
571 (Figures 4 and 10). The presence of the gene encoding CodH in *P. marina* (*Hydrogenothermaceae*) is
572 the only exception, and suggests that *codH* was lost independently in the two lineages (Figure 10).

573 Alltogether, our results suggest that the last common ancestor of the *Aquificae* combined the
574 reductive acetyl-CoA pathway with an incomplete form of the rTCA that did not proceed past the
575 synthesis of 2-oxoglutarate (Figure 10). A similar incomplete version of the rTCA pathway, consisting
576 only of the reactions from acetyl-CoA to 2-oxoglutarate, is present in extant methanogens (Berg, 2011).
577 Thus, our phylometabolic reconstruction of the ancestral state of carbon fixation in the *Aquificae*
578 (Figure 10) is conceptually consistent with chemoautotrophic processes of extant *Bacteria* and
579 *Archaea*, and may represent the earliest carbon fixation pathway.

580 **Conclusions**

581 We propose that the ancestor of *Thermovibrio ammonificans* was originally a hydrogen
582 oxidizing, sulfur reducing bacterium that used a hybrid carbon fixation pathway for CO_2 fixation. The
583 simultaneous presence of the rTCA cycle and of the reductive acetyl-CoA pathway of carbon fixation
584 in *T. ammonificans* may represent a modern analog of the early carbon fixation phenotype, and

suggests that the redundancy of central metabolic pathways was common in ancestral microorganisms. With the gradual rise of oxygen in the atmosphere, more efficient terminal electron acceptors became available and this lineage acquired genes that increased its metabolic flexibility - *e.g.*, the capacity to respire nitrate - along with enzymes involved in the detoxification of oxygen radicals. However, this lineage also retained its core, or ancestral, metabolic traits. Given the early branching nature of the phylum *Aquificae* and the ability of *T. ammonificans* and the *Desulfurobacteriaceae* to thrive in hydrothermal environments relying on energy sources of geothermal origins (namely carbon dioxide, hydrogen and elemental sulfur), we argue that these microorganisms represent excellent models to investigate how metabolism co-evolved with Earth's changing environmental conditions.

Methods

Strain isolation was described in (Vetriani et al., 2004). Growth condition, DNA extraction, sequencing strategy and automatic annotation were published in (Giovannelli et al., 2012).

Manual curation of the genome. Manual curation of the genome was performed using blastn and blastp (McGinnis and Madden, 2004) against the non-redundant database (Pruitt et al., 2007), the conserved domain database (Marchler-Bauer et al., 2005), the Kyoto Encyclopedia of Genes and Genomes (Kanehisa and Goto, 2000) and the PFAM database (Sonnhammer et al., 1998). Coding sequence similarities were compared using translated protein sequence. Metabolic pathways were reconstructed on the basis of available genomic, physiologic and biochemical information and using Kyoto Encyclopedia of Genes and Genomes (Kanehisa and Goto, 2000) and SEED (Overbeek et al., 2014) pathways as template.

Comparative genomics. Genome maps were drawn using Circos (Krzywinski et al., 2009), parsing blast results with *ad hoc* bash scripting. Comparative analyses between the genomes of *T. ammonificans*, those of representative members of the *Aquificae* (reported in Table 3),

608 *Desulfurobacterium thermolithotrophum* (Goker et al., 2011) and *Caminibacter mediatlanticus*
609 (Giovannelli et al., 2011) were performed using the GenomeEvolution pipeline CoGe (Lyons et al.,
610 2008). *Aquificae* genomes were selected among all available genome within this phylum to maximize
611 diversity while minimizing redundancy. All the available genomes of validly published *Aquificae*
612 species were selected for analysis. To this set we added the reference genomes for the genera
613 *Hydrogenivirga* and *Hydrogenobaculum*, as no genome sequence is available for validly published
614 species of these genera. We also selected two additional genomes belonging to the genera
615 *Desulfurobacterium* (the closest relative to the genus *Thermovibrio*) and *Persephonella,* respectively.
616 Excluded genomes include either alternative assemblies of selected genomes or closely related
617 genomes with a gapped genome similarity above 90%. LastZ pairwise alignments of selected genomes
618 were used to draw three-way plots using the Hive Plot software (Krzywinski et al., 2012). Gene context
619 and operonic structures were reconstructed using BioCyc (Karp et al., 2005) and FgenesB. Operons
620 were manually screened and their structure selected based on gene context and available literature on
621 the specific gene. When it was not possible to discriminate between the two alternative operonic
622 predictions, both were reported in the text. Similarities between *T*. *ammonificans* genes and other
623 prokaryotic genomes were found performing blastp analyses against the nr database. The top three blast
624 results were retained and further analyzed, ranking the genes for their best hits. The procedure was
625 repeated removing from the database the genome of *D*. *thermolithotrophum* and *Desulfurobacterium*
626 sp. TC5-1, thus searching for the best hit outside of the *Desulfurobacteriaceae* family. The results were
627 analyzed and interactive Krona plots (Ondov et al., 2011) drawn linking the *T*. *ammonificans* genes
628 with its closest match in the database. The interactive plots are accessible at DOI:
629 10.6084/m9.figshare.3178528. Images were drawn or edited using the open source vector drawing
630 program Inkscape (http://inkscape.org/).

631 **Phylogenetic analyses.** Phylogenetic analyses were performed using the approaches defined

below for each tree. The phylogenetic tree presented in Figure 1 was computed from the current 16S rRNA database alignment available from the ARB-SILVA project (http://www.arb-silva.de). The maximum likelihood three was computed from the ARB-SILVA alignment using PHYLML (Guindon and Gascuel, 2003) and the GTR model. The phylogenetic tree in Figure 1-figure supplement 1 was constructed by aligning complete or near complete 16S rRNA sequences obtained from NCBI and representing the phylum *Aquificae*. The 16S rRNA sequences were aligned with ClustalO (Thompson et al., 1997) and Gblocks (Castresana, 2000) and the alignment was manually refined using SEAVIEW (Galtier et al., 1996). The maximum likelihood phylogeny was inferred from the alignment of 1455 sites using PHYML (Guindon and Gascuel, 2003), the GRT model and 1000 bootstrap replications. The CODH tree presented in Figure 4 was computed using a selected set of amino acid sequences and the neighbor-joining method. Alignments were obtained using Muscle (Edgar, 2004) and Gblocks (Castresana, 2000), manually refined using SEAVIEW, and phylogenetic distances calculated using 255 sites and the Observed Divergence matrix. The neighbor-joining method was used to evaluate tree topologies using Phylo_win (Galtier et al., 1996) and their robustness was tested by bootstrap analysis with 1000 resamplings. The tree for the catalytic subunit of the [NiFe]-hydrogenases presented in Figure 5 was computed using a selected set of amino acid sequences and the same approach described above for the CODH tree and was based on 539 sites. The accession number for the sequences used in trees presented in Figure 1-figure supplement 1, Figure 4 and 5 are reported within each tree. The ATP citrate lyase phylogenetic tree was reconstructed with Bayesian inference and maximum likelihood methods. Both subunit of the ATP citrate lyase (AclB and AclA) were aligned individually to retrieved homologs using Muscle (Edgar, 2004) and SEAVIEW (Galtier et al., 1996). The alignments were concatenated and refined using Gblock (Castresana, 2000). A hypothetical ancestral ATP citrate lyase enzyme was manually constructed by concatenating the citryl-CoA synthetase and citryl-CoA lyase of *H. thermophilus* and *A. aeolicus*, respectively, and used as the outgroup (Hügler et al., 2007). The maximum likelihood phylogeny was inferred from the alignment using PHYML with the Wag

substitution model (Whelan and Goldman, 2001) and 1000 bootstrap resamplings. The substitution model was selected based on AIC values using ProTest3 (Darriba et al., 2011). Bayesian phylogeny was inferred using MrBayes (Ronquist and Huelsenbeck, 2003) performing 500,000 generations with two parallel searches with the Wag amino acid matrix model (selected using forward selection among all possible substitution models) and a burn-in of 125,000 generations. Both tree were computed on 1035 identified sites. Accession numbers for the tree presented are reported in Table S4. A combination of phylogenetic analysis and comparative genomic analyses were used to identify lateral gene transfer events.

Phylometabolic reconstruction of the carbon fixation pathway within the *Aquificae*.

Phylometabolic analysis (Braakman and Smith, 2012) was used to investigate the carbon fixation pathway in *T*. *ammonificans* and its evolutionary relationship to the carbon fixation pathways present in other members of the *Aquificae* phylum. In phylometabolic analyses, the metabolic pathways of the organism under investigation are compared to those found in related organisms both within and across neighboring clades. By focusing on the pathways, the comparison may reveal variations in multi-enzyme functional units, providing context for the completion of the pathway within the networks of individual organism, while also allowing for the identification of ancestral states and horizontal gene transfer events. The resulting phylometabolic tree includes multiple complete pathways to common essential metabolites, and suggests which evolutionary substitutions are allowed (at either organism or ecosystem levels) among these pathways (see Braakman and Smith, 2012 and reference therein for a more extensive description of the principles underlying this approach). We reconstructed the carbon fixation pathways in representative genomes of the *Aquificae*, and compared them. Information regarding the carbon fixation metabolic network was implemented using phylogenetic and comparative genomic information to help reconstruct possible ancestral states of the carbon fixation network based on maximum parsimony principles.

Protein extraction, digestion and identification. *T. ammonificans* cell pellets were washed in TE buffer (10 mM Tris-HCl pH 7.5, 10 mM EDTA pH 8.0, containing Roche cOmplete protease inhibitor) and soluble proteins were extracted as described by (Heinz et al., 2012). Briefly, cells were disrupted by sonication (2x25 s), cell debris was pelleted and protein concentrations in the supernatant were determined according to (Bradford, 1976). 20 µg of protein extract were loaded onto a precast 10% polyacrylamide mini gel in technical triplicates for 1D PAGE (150 V, 45 min). After staining with Coomassie Brilliant Blue, protein-containing gel lanes were excised and cut into 10 equal subsamples each, which were destained (200 mM NH_4HCO_3, 30% acetonitrile) and digested with trypsin solution (1 µg/ml, Promega, Madison WI, USA) at 37°C over night, before peptides were eluted from the gel pieces in an ultrasonic bath (15 min). As described by (Xing et al. 2014), peptide mixes were subjected to reversed phase C18 column chromatography on a nano-ACQUITY-UPLC (Waters Corporation, Milford, MA, USA). Mass spectrometry (MS) and MS/MS data were recorded with an online-coupled LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). MS data were searched against the forward-decoy *T. ammonificans* protein database using Sequest (Thermo Fisher Scientific, San Jose, CA, USA; version 27, revision 11) and identifications were filtered and validated in Scaffold (http://www.proteomesoftware.com), as described previously (Heinz et al., 2012). Data for all three technical replicates were merged and exclusive unique peptide count values of all proteins were exported for calculation of normalized spectral abundance factors (NSAF), which are given in Table S3 in %, *i.e.,* as relative abundance of each protein in % of all identified proteins.

Funding

The genome of *Thermovibrio ammonificans* was sequenced under the auspices of the US Department of Energy. Work on *T. ammonificans* was supported, in part, by NSF grants MCB 04-56676, OCE 03-27353, MCB 08-43678, OCE 09-37371, OCE 11-24141, OCE11-36451, and NASA

grant NNX15AM18G to CV, NSF grant MCB 15-17567 to CV and DG, and by the New Jersey Agricultural Experiment Station. DG was supported by a Postdoctoral Fellowship from the Institute of Marine and Coastal Sciences and a Postdoctoral Fellowship from the Center for Dark Energy Biosphere Investigations (C-DEBI). Funding to SMS was provided by NSF grant OCE-1136727 and a senior fellowship by the Alfried Krupp Wissenschaftskolleg Greifswald, Germany. This publication was in part supported by the ELSI Origins Network (EON), which is supported by a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

Acknowledgements

The authors gratefully acknowledge the support of the Deep Carbon Observatory. Thanks to Sandra Beyer for assistance in the lab and to Sebastian Grund for MS measurements.

Competing Interests Statement: None of the authors have any competing interests

719 **References**

Aoshima, M., 2007. Novel enzyme reactions related to the tricarboxylic acid cycle: phylogenetic/functional implications and biotechnological applications. Appl. Microbiol. Biotechnol. 75, 249–255.

Aoshima, M., Igarashi, Y., 2008. Nondecarboxylating and Decarboxylating Isocitrate Dehydrogenases: Oxalosuccinate Reductase as an Ancestral Form of Isocitrate Dehydrogenase. J. Bacteriol. 190, 2050–2055. doi:10.1128/JB.01799-07

Aoshima, M., Ishii, M., Igarashi, Y., 2004. A novel enzyme, citryl-CoA lyase, catalysing the second step of the citrate cleavage reaction in Hydrogenobacter thermophilus TK-6. Mol. Microbiol. 52, 763–770. doi:10.1111/j.1365-2958.2004.04010.x

Arai, H., Kanbe, H., Ishii, M., Igarashi, Y., 2010. Complete Genome Sequence of the Thermophilic, Obligately Chemolithoautotrophic Hydrogen-Oxidizing Bacterium Hydrogenobacter thermophilus TK-6. J. Bacteriol. 192, 2651–2652. doi:10.1128/JB.00158-10

Baross, J.A., Hoffman, S.E., 1985. Submarine hydrothermal vents and associated gradient environments as sites for the origin and evolution of life. Orig. Life Evol. Biospheres 15, 327–345.

Berg, I.A., 2011. Ecological Aspects of the Distribution of Different Autotrophic CO2 Fixation Pathways. Appl. Environ. Microbiol. 77, 1925–1936. doi:10.1128/AEM.02473-10

Berg, I.A., Kockelkorn, D., Ramos-Vera, W.H., Say, R.F., Zarzycki, J., Hügler, M., Alber, B.E., Fuchs, G., 2010. Autotrophic carbon fixation in archaea. Nat. Rev. Microbiol. 8, 447–460. doi:10.1038/nrmicro2365

Battistuzzi, F.U., Feijao, A., Balir Hedger, S. 2004. A genomic timescale of prokaryotic evolution: insights into the origin of methanogenesis, phototrophy and colonization of land. BMC Evol. Biol. 4:44. doi:10.1186/1471-2148-4-44

Braakman, R., Smith, E., 2012. The Emergence and Early Evolution of Biological Carbon-Fixation. PLoS Comput Biol 8, e1002455. doi:10.1371/journal.pcbi.1002455

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248–254.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27, 1164–1165. doi:10.1093/bioinformatics/btr088

Das, A., Silaghi-Dumitrescu, R., Ljungdahl, L.G., Kurtz, D.M., 2005. Cytochrome bd Oxidase, Oxidative Stress, and Dioxygen Tolerance of the Strictly Anaerobic Bacterium Moorella thermoacetica. J. Bacteriol. 187, 2020–2029. doi:10.1128/JB.187.6.2020-2029.2005

Deckert, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D.E., Overbeek, R., Snead, M.A., Keller, M., Aujay, M., Huber, R., Feldman, R.A., Short, J.M., Olsen, G.J., Swanson, R.V., 1998. The complete genome of the hyperthermophilic bacterium Aquifex aeolicus. Nature 392, 353–358. doi:10.1038/32831

Di Giulio, M., 2003. The ancestor of the Bacteria domain was a hyperthermophile. J. Theor. Biol. 224, 277–283. doi:10.1016/S0022-5193(03)00164-4

Di Giulio, M., 2000. The universal ancestor lived in a thermophilic or hyperthermophilic environment. J. Theor. Biol. 203, 203–213. doi:10.1006/jtbi.2000.1086

Douzi, B., Filloux, A., Voulhoux, R., 2012. On the path to uncover the bacterial type II secretion system. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 367, 1059–1072. doi:10.1098/rstb.2011.0204

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high

throughput. Nucleic Acids Res. 32, 1792–1797.

Fuchs, G., 2011. Alternative Pathways of Carbon Dioxide Fixation: Insights into the Early Evolution of Life? Annu. Rev. Microbiol. 65, 631–658. doi:10.1146/annurev-micro-090110-102801

Galtier, N., Gouy, M., Gautier, C., 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. CABIOS 12, 543–548. doi:10.1093/bioinformatics/12.6.543

Giovannelli, D., Ferriera, S., Johnson, J., Kravitz, S., Perez-Rodriguez, I., Ricci, J., O'Brien, C., Voordeckers, J.W., Bini, E., Vetriani, C., 2011. Draft genome sequence of Caminibacter mediatlanticus strain TB-2T, an epsilonproteobacterium isolated from a deep-sea hydrothermal vent. Stand. Genomic Sci. 5, 135–143. doi:10.4056/sigs.2094859

Giovannelli, D., Ricci, J., Perez-Rodriguez, I., Hugler, M., O'Brien, C., Keddis, R., Grosche, A., Goodwin, L., Bruce, D., Davenport, K.W., Detter, C., Han, J., Han, S., Ivanova, N., Land, M.L., Mikhailova, N., Nolan, M., Pitluck, S., Tapia, R., Woyke, T., Vetriani, C., 2012. Complete genome sequence of Thermovibrio ammonificans HB-1T, a thermophilic, chemolithoautotrophic bacterium isolated from a deep-sea hydrothermal vent. Stand. Genomic Sci. 7, 82–90. doi:10.4056/sigs.2856770

Goker, M., Daligault, H., Mwirichia, R., Lapidus, A., Lucas, S., Deshpande, S., Pagani, I., Tapia, R., Cheng, J.-F., Goodwin, L., Pitluck, S., Liolios, K., Ivanova, N., Mavromatis, K., Mikhailova, N., Pati, A., Chen, A., Palaniappan, K., Han, C., Land, M., Hauser, L., Pan, C., Brambilla, E.-M., Rohde, M., Spring, S., Sikorski, J., Wirth, R., Detter, J.C., Woyke, T., Bristow, J., Eisen, J.A., Markowitz, V., Hugenholtz, P., Kyrpides, N.C., Klenk, H.-P., 2011. Complete genome sequence of the thermophilic sulfur-reducer Desulfurobacterium thermolithotrophum type strain (BSAT) from a deep-sea hydrothermal vent. Stand. Genomic Sci. 5, 407–415. doi:10.4056/sigs.2465574

Götz, D., Banta, A., Beveridge, T.J., Rushdi, A.I., Simoneit, B.R.T., Reysenbach, A.L., 2002. Persephonella marina gen. nov., sp. nov. and Persephonella guaymasensis sp. nov., two novel,

thermophilic, hydrogen-oxidizing microaerophiles from deep-sea hydrothermal vents. Int. J. Syst. Evol. Microbiol. 52, 1349–1359. doi:10.1099/ijs.0.02126-0

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

Heinz, E., Williams, T.A., Nakjang, S., Noël, C.J., Swan, D.C., Goldberg, A.V., Harris, S.R., Weinmaier, T., Markert, S., Becher, D., 2012. The genome of the obligate intracellular parasite Trachipleistophora hominis: new insights into microsporidian genome dynamics and reductive evolution. PLoS Pathog. 8, e1002979.

Hetzer, A., McDonald, I.R., Morgan, H.W., 2008. Venenivibrio stagnispumantis gen. nov., sp. nov., a thermophilic hydrogen-oxidizing bacterium isolated from Champagne Pool, Waiotapu, New Zealand. Int. J. Syst. Evol. Microbiol. 58, 398–403. doi:10.1099/ijs.0.64842-0

Huber, R., Eder, W., Heldwein, S., Wanner, G., Huber, H., Rachel, R., Stetter, K.O., 1998. Thermocrinis ruber gen. nov., sp. nov., a Pink-Filament-Forming Hyperthermophilic Bacterium Isolated from Yellowstone National Park. Appl. Environ. Microbiol. 64, 3576–3583.

Huber, R., Wilharm, T., Huber, D., Trincone, A., Burggraf, S., König, H., Reinhard, R., Rockinger, I., Fricke, H., Stetter, K.O., 1992. Aquifex pyrophilus gen. nov. sp. nov., Represents a Novel Group of Marine Hyperthermophilic Hydrogen-Oxidizing Bacteria. Syst. Appl. Microbiol. 15, 340–351. doi:10.1016/S0723-2020(11)80206-7

Hügler, M., Huber, H., Molyneaux, S.J., Vetriani, C., Sievert, S.M., 2007. Autotrophic CO2 fixation via the reductive tricarboxylic acid cycle in different lineages within the phylum Aquificae: evidence for two ways of citrate cleavage. Environ. Microbiol. 9, 81–92. doi:10.1111/j.1462-2920.2006.01118.x

Hügler, M., Sievert, S.M., 2011. Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the

Ocean. Annu. Rev. Mar. Sci. 3, 261–289. doi:10.1146/annurev-marine-120709-142712

Jensen, R.A., 1976. Enzyme Recruitment in Evolution of New Function. Annu. Rev. Microbiol. 30, 409–425. doi:10.1146/annurev.mi.30.100176.002205

Jeon, J.H., Lim, J.K., Kim, M.-S., Yang, T.-J., Lee, S.-H., Bae, S.S., Kim, Y.J., Lee, S.H., Lee, J.-H., Kang, S.G., Lee, H.S., 2014. Characterization of the frhAGB-encoding hydrogenase from a non-methanogenic hyperthermophilic archaeon. Extremophiles 1–10. doi:10.1007/s00792-014-0689-y

Jepson, B.J.N., Marietou, A., Mohan, S., Cole, J.A., Butler, C.S., Richardson, D.J., 2006. Evolution of the soluble nitrate reductase: defining the monomeric periplasmic nitrate reductase subgroup. Biochem. Soc. Trans. 34, 122–126. doi:10.1042/BST0340122

Kanehisa, M., Goto, S., 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27–30. doi:10.1093/nar/28.1.27

Karp, P.D., Ouzounis, C.A., Moore-Kochlacs, C., Goldovsky, L., Kaipa, P., Ahrén, D., Tsoka, S., Darzentas, N., Kunin, V., López-Bigas, N., 2005. Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. Nucleic Acids Res. 33, 6083–6089. doi:10.1093/nar/gki892

Kawasumi, T., Igarashi, Y., Kodama, T., Minoda, Y., 1984. Hydrogenobacter thermophilus gen. nov., sp. nov., an Extremely Thermophilic, Aerobic, Hydrogen-Oxidizing Bacterium. Int. J. Syst. Bacteriol. 34, 5–10. doi:10.1099/00207713-34-1-5

Kim, J., Kershner, J.P., Novikov, Y., Shoemaker, R.K., Copley, S.D., 2010. Three serendipitous pathways in E. coli can bypass a block in pyridoxal-5′-phosphate synthesis. Mol. Syst. Biol. 6. doi:10.1038/msb.2010.88

Krzywinski, M., Birol, I., Jones, S.J., Marra, M.A., 2012. Hive plots—rational approach to visualizing networks. Brief. Bioinform. 13, 627–644. doi:10.1093/bib/bbr069

Krzywinski, M., Schein, J., Birol, İ., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A., 2009. Circos: An information aesthetic for comparative genomics. Genome Res. 19, 1639–1645. doi:10.1101/gr.092759.109

Lane, N., Allen, J.F., Martin, W., 2010. How did LUCA make a living? Chemiosmosis in the origin of life. BioEssays 32, 271–280. doi:10.1002/bies.200900131

Lebedinsky, A.V., Chernyh, N.A., Bonch-Osmolovskaya, E.A., 2007. Phylogenetic systematics of microorganisms inhabiting thermal environments. Biochem. Mosc. 72, 1299–1312. doi:10.1134/S0006297907120048

L'Haridon, S., Cilia, V., Messner, P., Raguénès, G., Gambacorta, A., Sleytr, U.B., Prieur, D., Jeanthon, C., 1998. Desulfurobacterium thermolithotrophum gen. nov., sp. nov., a novel autotrophic, sulphur-reducing bacterium isolated from a deep-sea hydrothermal vent. Int. J. Syst. Bacteriol. 48, 701–711. doi:10.1099/00207713-48-3-701

Lyons, E., Pedersen, B., Kane, J., Alam, M., Ming, R., Tang, H., Wang, X., Bowers, J., Paterson, A., Lisch, D., Freeling, M., 2008. Finding and Comparing Syntenic Regions among Arabidopsis and the Outgroups Papaya, Poplar, and Grape: CoGe with Rosids. Plant Physiol. 148, 1772–1781. doi:10.1104/pp.108.124867

Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Marchler, G.H., Mullokandov, M., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Yamashita, R.A., Yin, J.J., Zhang, D., Bryant, S.H., 2005. CDD: a Conserved Domain Database for protein classification. Nucleic Acids Res. 33, D192–D196. doi:10.1093/nar/gki069

Markert, S., Arndt, C., Felbeck, H., Becher, D., Sievert, S.M., Hügler, M., Albrecht, D., Robidart, J., Bench, S., Feldman, R.A., Hecker, M., Schweder, T., 2007. Physiological Proteomics of the Uncultured Endosymbiont of Riftia pachyptila. Science 315, 247–250.

doi:10.1126/science.1132913

Martin, W., Baross, J., Kelley, D., Russell, M.J., 2008. Hydrothermal vents and the origin of life. Nat. Rev. Microbiol. 6, 805–814. doi:10.1038/nrmicro1991

Martin, W., Russell, M.J., 2007. On the origin of biochemistry at an alkaline hydrothermal vent. Philos. Trans. R. Soc. B Biol. Sci. 362, 1887–1926. doi:10.1098/rstb.2006.1881

Martin, W., Russell, M.J., 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, 59–85. doi:10.1098/rstb.2002.1183

Mason, M.G., Shepherd, M., Nicholls, P., Dobbin, P.S., Dodsworth, K.S., Poole, R.K., Cooper, C.E., 2009. Cytochrome bd confers nitric oxide resistance to Escherichia coli. Nat. Chem. Biol. 5, 94–96. doi:10.1038/nchembio.135

McGinnis, S., Madden, T.L., 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 32, W20–W25.

Moreno-Vivián, C., Cabello, P., Martínez-Luque, M., Blasco, R., Castillo, F., 1999. Prokaryotic Nitrate Reduction: Molecular Properties and Functional Distinction among Bacterial Nitrate Reductases. J. Bacteriol. 181, 6573–6584.

Nakagawa, S., Nakamura, S., Inagaki, F., Takai, K., Shirai, N., Sako, Y., 2004. Hydrogenivirga caldilitoris gen. nov., sp. nov., a novel extremely thermophilic, hydrogen- and sulfur-oxidizing bacterium from a coastal hydrothermal field. Int. J. Syst. Evol. Microbiol. 54, 2079–2084. doi:10.1099/ijs.0.03031-0

Nunoura, T., Oida, H., Miyazaki, M., Suzuki, Y., 2008. Thermosulfidibacter takaii gen. nov., sp. nov., a thermophilic, hydrogen-oxidizing, sulfur-reducing chemolithoautotroph isolated from a deep-sea hydrothermal field in the Southern Okinawa Trough. Int. J. Syst. Evol. Microbiol. 58, 659–

665. doi:10.1099/ijs.0.65349-0

Ondov, B.D., Bergman, N.H., Phillippy, A.M., 2011. Interactive metagenomic visualization in a Web browser. BMC Bioinformatics 12, 385. doi:10.1186/1471-2105-12-385

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 42, D206–D214.

Pérez-Rodríguez, I., Grosche, A., Massenburg, L., Starovoytov, V., Lutz, R.A., Vetriani, C., 2012. Phorcysia thermohydrogeniphila gen. nov., sp. nov., a thermophilic, chemolithoautotrophic, nitrate-ammonifying bacterium from a deep-sea hydrothermal vent. Int. J. Syst. Evol. Microbiol. 62, 2388–2394. doi:10.1099/ijs.0.035642-0

Pitulle, C., Yang, Y., Marchiani, M., Moore, E.R., Siefert, J.L., Aragno, M., Jurtshuk, P.J., Fox, G.E., 1994. Phylogenetic position of the genus Hydrogenobacter. Int. J. Syst. Bacteriol. 44, 620–626.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Pruitt, K.D., Tatusova, T., Maglott, D.R., 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 35, D61–D65. doi:10.1093/nar/gkl842

Reysenbach, A.-L., Hamamura, N., Podar, M., Griffiths, E., Ferreira, S., Hochstein, R., Heidelberg, J., Johnson, J., Mead, D., Pohorille, A., Sarmiento, M., Schweighofer, K., Seshadri, R., Voytek, M.A., 2009. Complete and Draft Genome Sequences of Six Members of the Aquificales. J. Bacteriol. 191, 1992–1993. doi:10.1128/JB.01645-08

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574. doi:10.1093/bioinformatics/btg180

Russell, M.J., Hall, A.J., 2006. The onset and early evolution of life. Geol. Soc. Am. Mem. 198, 1–32. doi:10.1130/2006.1198(01)

Russell, M.J., Martin, W., 2004. The rocky roots of the acetyl-CoA pathway. Trends Biochem. Sci. 29, 358–363. doi:10.1016/j.tibs.2004.05.007

Sandkvist, M., 2001. Type II Secretion and Pathogenesis. Infect. Immun. 69, 3523–3535. doi:10.1128/IAI.69.6.3523-3535.2001

Schopf, J.W., 1983. Earth's earliest biosphere: its origin and evolution.

Schut, G.J., Boyd, E.S., Peters, J.W., Adams, M.W.W., 2013. The modular respiratory complexes involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and their evolutionary implications. FEMS Microbiol. Rev. 37, 182–203. doi:10.1111/j.1574-6976.2012.00346.x

Schut, G.J., Bridger, S.L., Adams, M.W.W., 2007. Insights into the Metabolism of Elemental Sulfur by the Hyperthermophilic Archaeon Pyrococcus furiosus: Characterization of a Coenzyme A-Dependent NAD(P)H Sulfur Oxidoreductase. J. Bacteriol. 189, 4431–4441. doi:10.1128/JB.00031-07

Sievert, S., Vetriani, C., 2012. Chemoautotrophy at Deep-Sea Vents: Past, Present, and Future. Oceanography 25, 218–233. doi:10.5670/oceanog.2012.21

Sonnhammer, E.L.L., Eddy, S.R., Birney, E., Bateman, A., Durbin, R., 1998. Pfam: Multiple sequence alignments and HMM-profiles of protein domains. Nucleic Acids Res. 26, 320–322. doi:10.1093/nar/26.1.320

Stohr, R., Waberski, A., Völker, H., Tindall, B.J., Thomm, M., 2001. Hydrogenothermus marinus gen. nov., sp. nov., a novel thermophilic hydrogen-oxidizing bacterium, recognition of Calderobacterium hydrogenophilum as a member of the genus Hydrogenobacter and proposal of the reclassification of Hydrogenobacter acidophilus as Hydrogenobaculum acidophilum gen. nov., comb.

nov., in the phylum “Hydrogenobacter/Aquifex”. Int. J. Syst. Evol. Microbiol. 51, 1853–1862. doi:10.1099/00207713-51-5-1853

Takai, K., Kobayashi, H., Nealson, K.H., Horikoshi, K., 2003a. Sulfurihydrogenibium subterraneum gen. nov., sp. nov., from a subsurface hot aquifer. Int. J. Syst. Evol. Microbiol. 53, 823–827. doi:10.1099/ijs.0.02506-0

Takai, K., Nakagawa, S., Sako, Y., Horikoshi, K., 2003b. Balnearium lithotrophicum gen. nov., sp. nov., a novel thermophilic, strictly anaerobic, hydrogen-oxidizing chemolithoautotroph isolated from a black smoker chimney in the Suiyo Seamount hydrothermal system. Int. J. Syst. Evol. Microbiol. 53, 1947–1954. doi:10.1099/ijs.0.02773-0

Techtmann, S.M., Lebedinsky, A.V., Colman, A.S., Sokolova, T.G., Woyke, T., Goodwin, L., Robb, F.T., 2012. Evidence for Horizontal Gene Transfer of Anaerobic Carbon Monoxide Dehydrogenases. Front. Microbiol. 3. doi:10.3389/fmicb.2012.00132

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Vendeville, A., Winzer, K., Heurlier, K., Tang, C.M., Hardie, K.R., 2005. Making “sense” of metabolism: autoinducer-2, LUXS and pathogenic bacteria. Nat. Rev. Microbiol. 3, 383–396. doi:10.1038/nrmicro1146

Vetriani, C., Speck, M.D., Ellor, S.V., Lutz, R.A., Starovoytov, V., 2004. Thermovibrio ammonificans sp. nov., a thermophilic, chemolithotrophic, nitrate-ammonifying bacterium from deep-sea hydrothermal vents. Int. J. Syst. Evol. Microbiol. 54, 175–181.

Vetriani, C., Voordeckers, J.W., Crespo-Medina, M., O’Brien, C.E., Giovannelli, D., Lutz, R.A., 2014. Deep-sea hydrothermal vent Epsilonproteobacteria encode a conserved and widespread

nitrate reduction pathway (Nap). ISME J. 8, 1510–1521. doi:10.1038/ismej.2013.246

Vignais, P.M., Billoud, B., 2007. Occurrence, Classification, and Biological Function of Hydrogenases: An Overview. Chem. Rev. 107, 4206–4272.

Wächtershäuser, G., 1990. Evolution of the first metabolic cycles. Proc. Natl. Acad. Sci. 87, 200–204.

Wächtershäuser, G., 1988. Before enzymes and templates: theory of surface metabolism. Microbiol. Rev. 52, 452–484.

Wächtershäuser, G., 1988. Pyrite Formation, the First Energy Source for Life: a Hypothesis. Syst. Appl. Microbiol. 10, 207–210. doi:10.1016/S0723-2020(88)80001-8

Wadhams, G.H., Armitage, J.P., 2004. Making sense of it all: bacterial chemotaxis. Nat. Rev. Mol. Cell Biol. 5, 1024–1037. doi:10.1038/nrm1524

Weiss, M.C., Sousa, F.L., Mrnjavac, N., Neukirchen, S. Roettger, M., Nelson-Sathi, S. and Martin, W.F. 2016. The physiology and habitat of the last universal common ancestor. Nat. Microbiol. 1, 16116-16123. doi: 10.1038/nmicrobiol.2016.116.

Whelan, S., Goldman, N., 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. 18, 691–699.

720

Figures

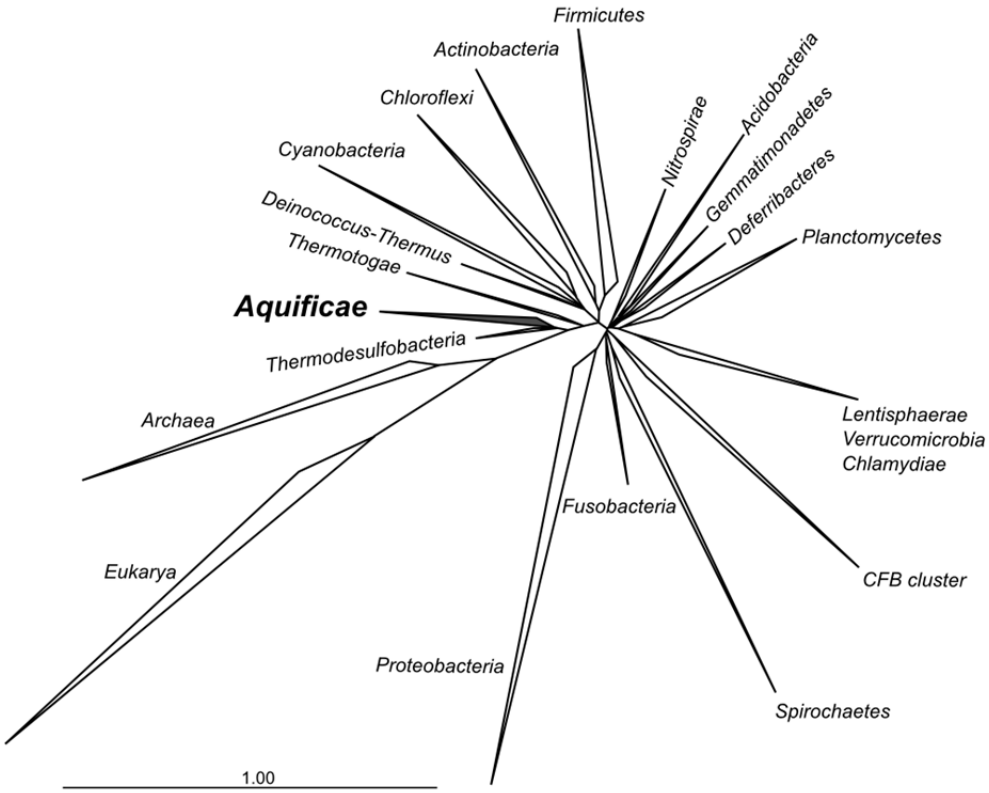
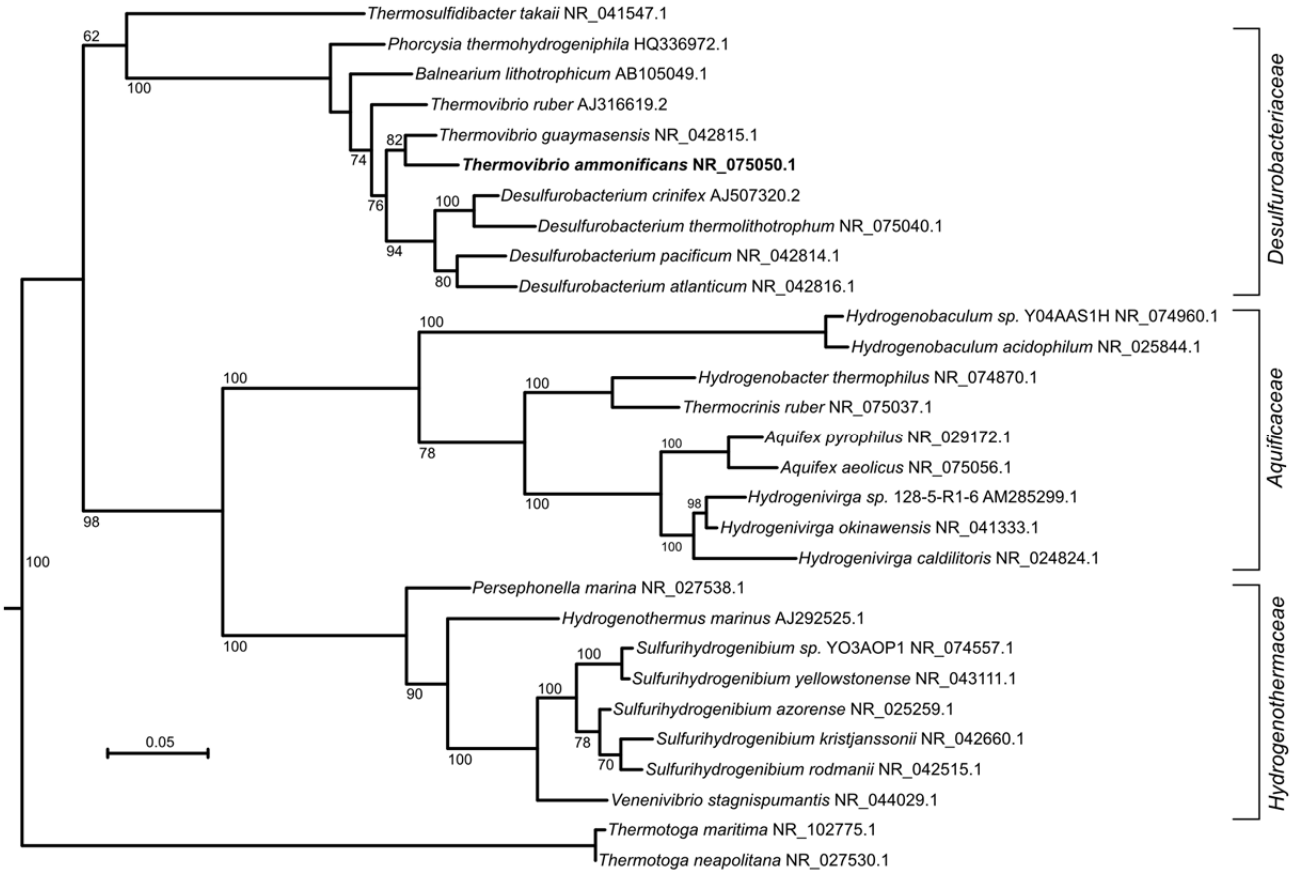


Figure 1. Phylogenetic tree of 16S rRNA sequences computed from the current version of the aligned 16S rRNA database obtained from the arb-SILVA project (http://www.arb-silva.de/).

727 .



728

729 **Figure 1-figure supplement 1**. Maximum likelihood tree showing the 16S rRNA relationship

730 of the *Aquificae* phylum. *Thermotoga* spp. were used as outgroup. Bootstrap values based on 1000

731 replication are shown at branch node. Bar, 5% estimated substitution.

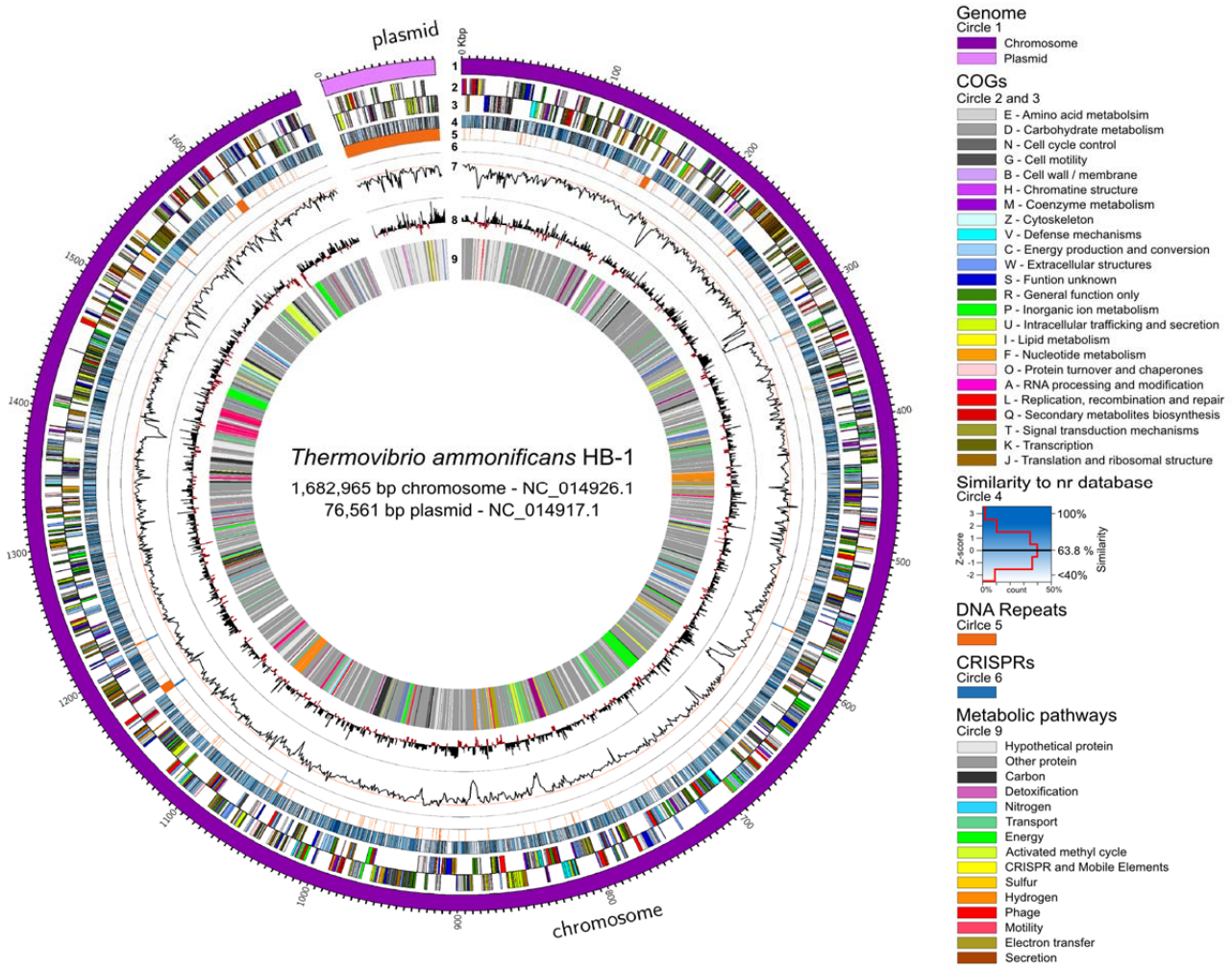
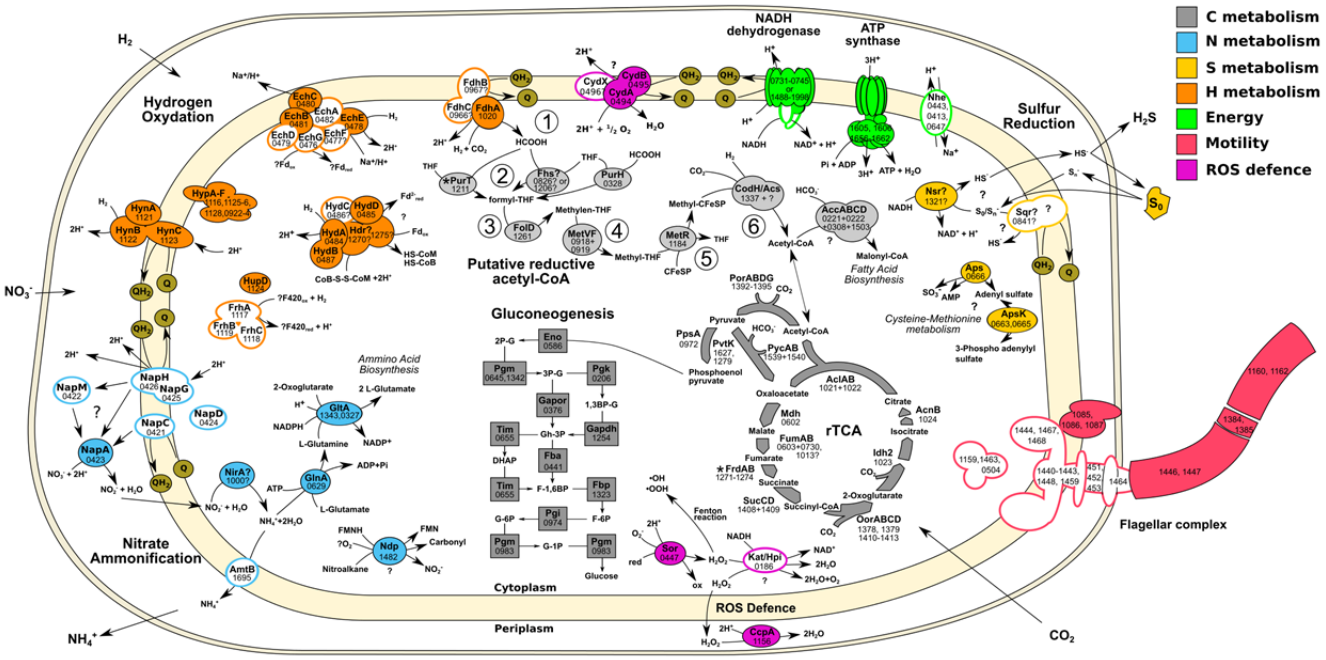
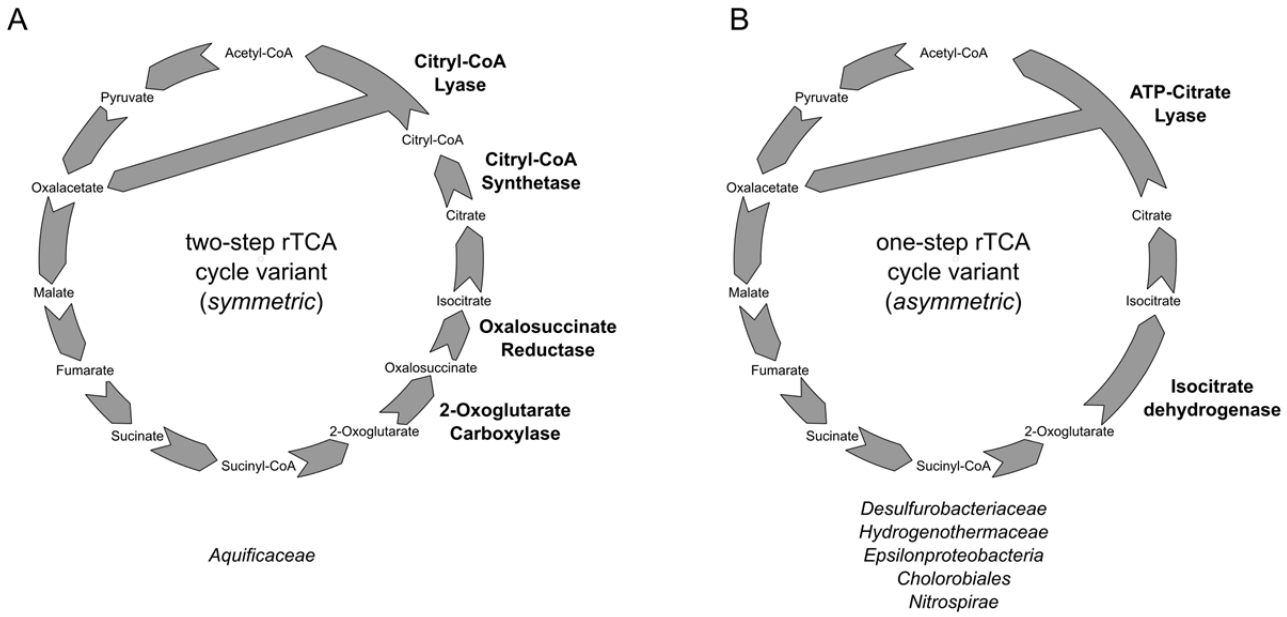


Figure 2. *Thermovibrio ammonificans* HB-1 genomic map highlighting main genomic features. The circular chromosome and plasmid are drawn together. Several other repeats were identified in the chromosome (circle 5). From the outer circle: 1 – Dimension of the genome in base pairs (chromosome in violet, plasmid in pink); 2 and 3 – sense and antisense coding sequences colored according to their COGs classification (see legend in the figure); 4 – similarity of each coding sequence with sequences in the non-redundant database; 5 – position of tandem repeat in the genome. A total of 46 tRNA were identified, comprising all of the basic 20 amino acid plus selenocysteine; 6 – position of CRISPRs; 7 – GC mol% content, the red line is the GC mean of the genome (52.1 mol%); 8 – GC skew calculated as G+C/A+T; 9 – localization of coding genes arbitrarily colored according to the metabolic pathways reconstructed (see Figure 1). Colors are consistent trough the entire paper. Total dimension and accession number for the chromosome and plasmid are given.



744 **Figure 3. Central metabolism of *T. ammonificans* HB-1.** Enzyme names are reported
745 together with the gene locus number (Theam_*number*). Primary compounds involved in reactions were
746 also reported, however visualized reactions are not complete. Pathways were arbitrarily color coded
747 according to their reconstructed function and are consistent throughout the paper. Solid shapes
748 represent genes for which the enzyme was found in the proteome, while outlined shapes were only
749 identified in the genome. Circled numbers 1 to 6 represent reaction numbers of the putative reductive
750 acetyl-CoA pathway as described in the text. Abbreviations: *NITRATE AMMONIFICATION*:
751 NapCMADGH – Periplasmic nitrate reductase complex; NirA – Putative nitrite reductase; AmtB –
752 ammonia transporter; GlnA – L-Glutamine synthetase; GltA – Glutamate synthetase. *HYDROGEN*
753 *OXIDATION*: HynABC – Ni-Fe Membrane bound hydrogenase; FrhACB – Cytosolic Ni-Fe
754 hydrogenase/putative coenzyme F420 hydrogenase; HupD – Cyrosolic Ni-Fe hydrogenase maturation
755 protease; HypA-F – Hydrogenases expression/synthesis accessory proteins; EchABCEF – Ech
756 membrane bound hydrogenase complex; HydAB – Cytosolic Ni-Fe hydrogenases potentially involved
757 in ferredoxin reduction; Hdr? – Missing heterodisulfide reductase CoB-CoM; FdhABC – Formate
758 dehydrogenase. *ENERGY PRODUCTION*: NADH dehydrogenase and ATP synthetase are reported
759 without the names of the single units; Nhe – Sodium/hydrogen symporter. *SULFUR REDUCTION*: Sqr
760 – Putative sulfate quinone reductase involved in sulfur respiration; Nsr? – FAD/NAD nucleotide-
761 disulphide oxidoreductase; Aps – Sulfate adenylyl transferase and kinase involved in assimilation of
762 sulfur. *FLAGELLAR COMPLEX*: for simplicity single unit names are not reported. *REDUCTIVE*

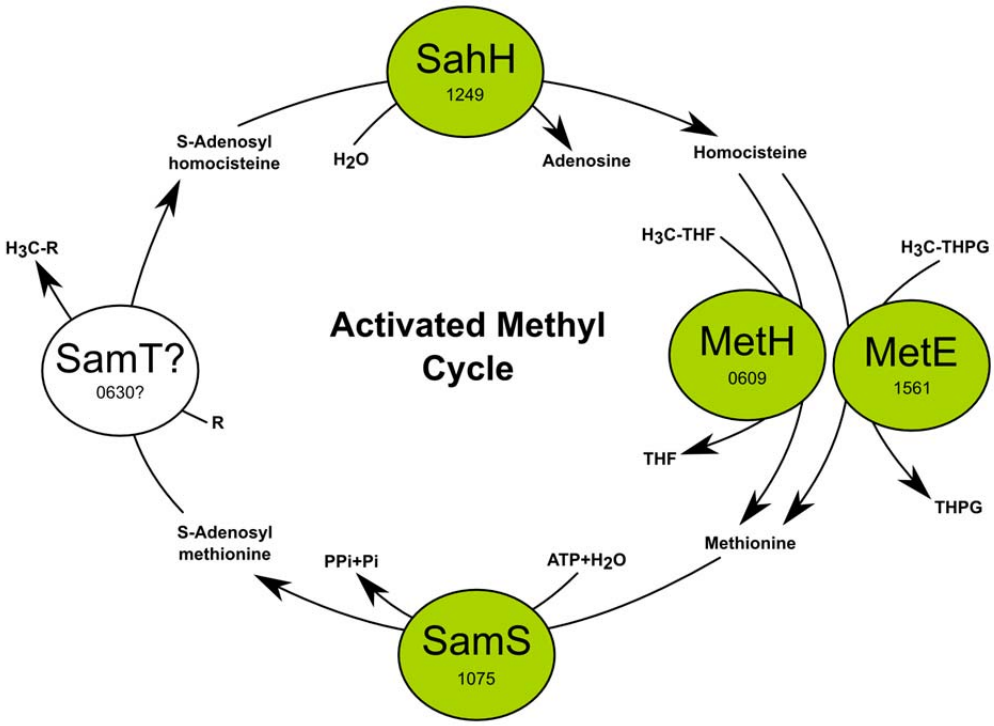
ACETYL-CoA: Fhs – Proposed reverse formyl-THF deformylase; PurHT – Formyltransferases (phosphoribosylaminoimidazolecarboxamide and phosphoribosylglycinamide) FolD – Methenyl-THF cyclohydrolase and dehydrogenase; MetVF – Methylene-THF reductase; MetR – Putative methyl-transferase; CodH – Carbon monoxyde dehydrogenase; AcsA – Acetyl-CoA ligase/synthase; AccABCD – Acetyl-CoA carboxylase. *REDUCTIVE CITRIC ACID CYCLE*: AclAB – ATP-citrate lyase; Mdh – Malate dehydrogenase; FumAB – Fumarate hydratase; FrdAB – Fumarate reductase; SucCD – Succinyl-CoA synthetase; OorABCD – 2-Oxoglutarate synthase; Idh2 – Isocitrate dehydrogenase/2-oxoglutarate carboxylase; AcnB – Aconitate hydratase; PorABDG – Pyruvate synthase; PycAB - Pyruvate carboxylase; PpsA – Phosphoenolpyruvate synthase water dikinase; PyK – Pyruvate:water dikinase. *GLUCONEOGENESIS*: Eno – Enolase; Pgm – Phosphoglycerate mutase; Pgk – Phosphoglycerate kinase; Gapor – Glyceraldehyde-3-phosphate dehydrogenase; Gapdh – Glyceraldehyde 3-phosphate dehydrogenase; Fba – Predicted fructose-bisphosphate aldolase; Tim – Triosephosphate isomerase; Fbp – Fructose-1,6-bisphosphatase I; Pgi – Phosphoglucose isomerase; Pgm – Phosphoglucomutase.



778
779 **Figure 3-figure supplement 1.** Reductive TCA cycle variants found in extant bacterial
780 lineages. (A) two-step variant of the rTCA cycle, also known as symmetric variants, were the
781 carboxylation of 2-oxoglutarate is performed in two steps by the enzymes 2-oxoglutarate carboxylase
782 and oxalosuccinate reductase, and the cleavage of citrate is performed in two step by the citryl-CoA
783 synthetase and citryl-CoA lyase enzymes; (B) one-step variant of the rTCA cycle, also known as
784 asymmetric variants, were the carboxylation of 2-oxoglutarate is performed in a single step by the
785 enzymes isocitrate dehydrogenase, and the cleavage of citrate to acetyl-CoA is performed in a single
786 reaction by the ATP citrate lyase enzyme.

787

788



789

790 **Figure 3-figure supplement 2.** The activated methyl cycle of *T*. *ammonificans* reconstructed

791 from the genome.

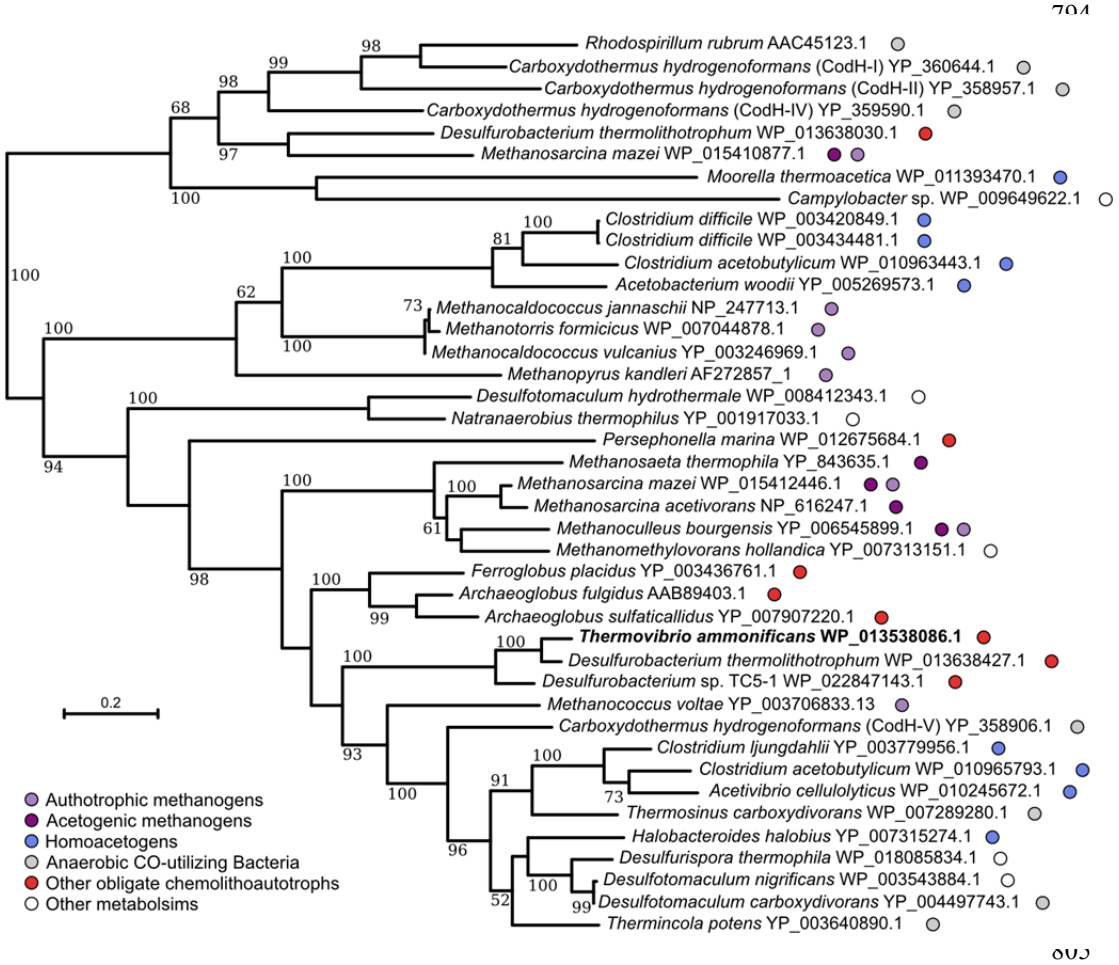
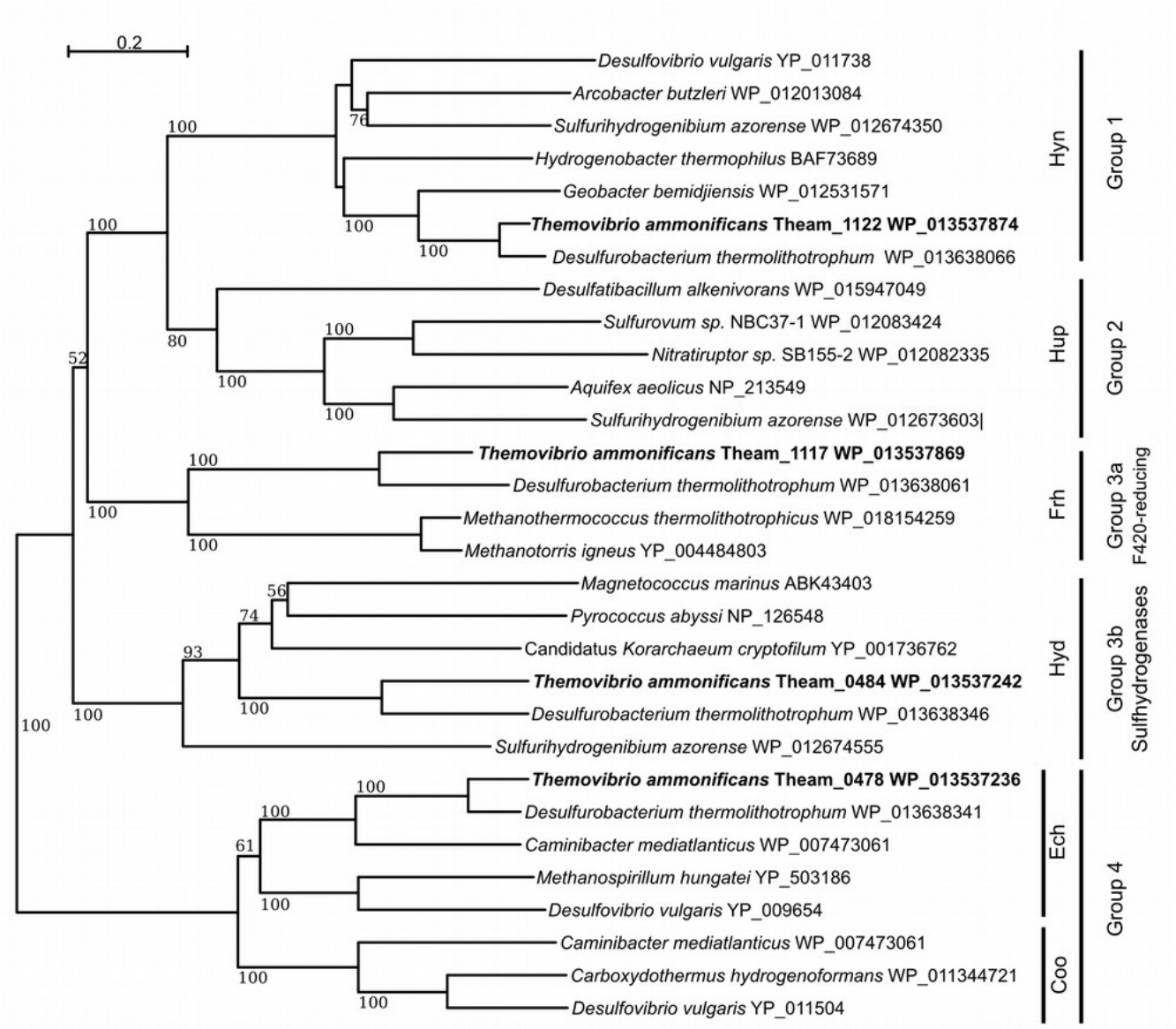


Figure 4. Neighbor-joining tree showing the position of the carbon monoxide dehydrogenase, CodH, of *T. ammonificans* (in bold). Bootstrap values based on 1000 replications are shown at branch nodes. Only bootstrap values above 50% are reported. Bar, 10% estimated substitutions. The metabolism of each organism is reported on the side.



811

812 **Figure 5**. Neighbor-Joining tree showing the position and classification of the [NiFe]-

813 hydrogenases found in the genome of *T*. *ammonificans*. Bootstrap values based on 1000 replication are

814 shown at branch node. Bar, 20% estimated substitution.

815

816

817

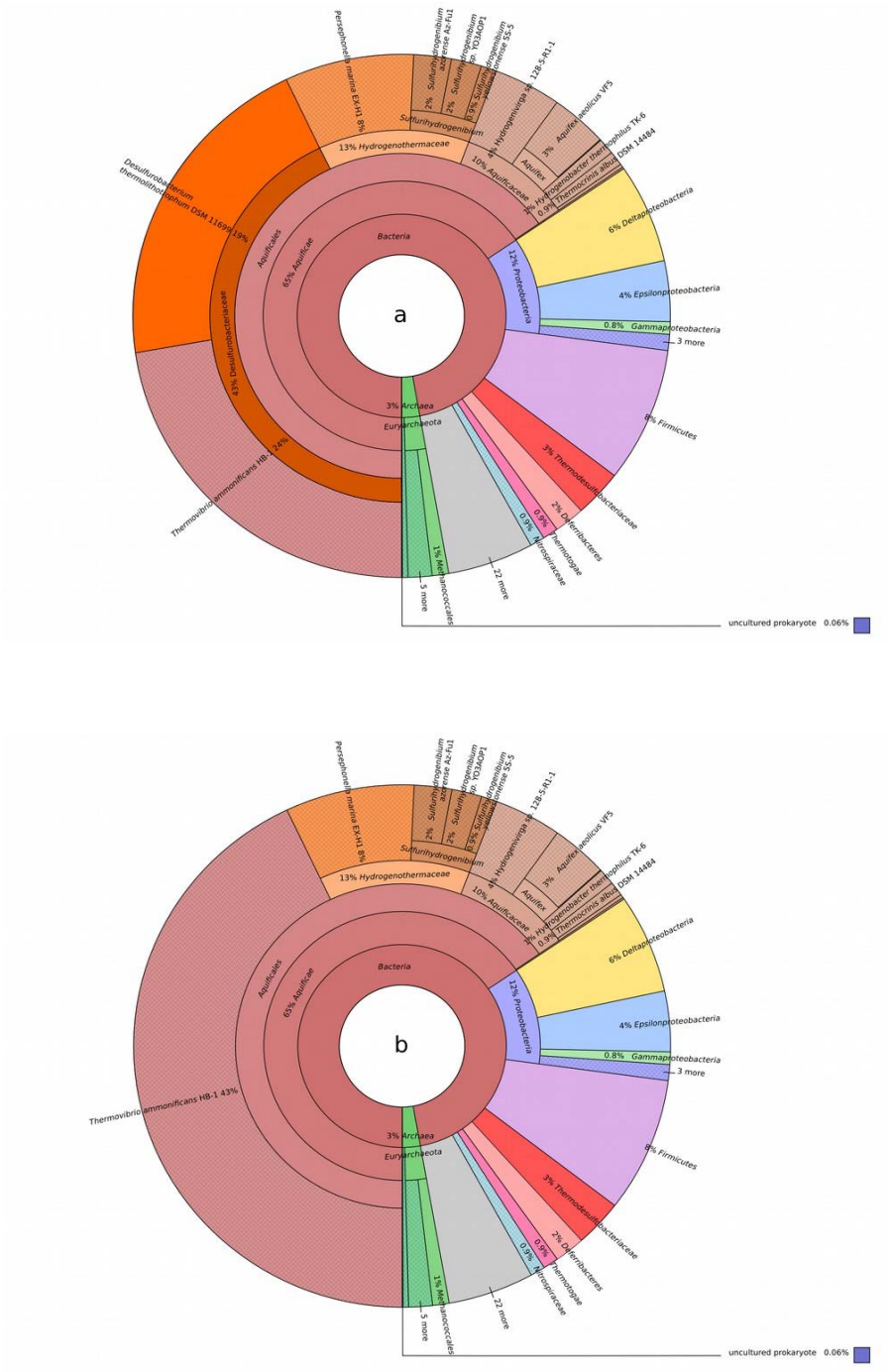
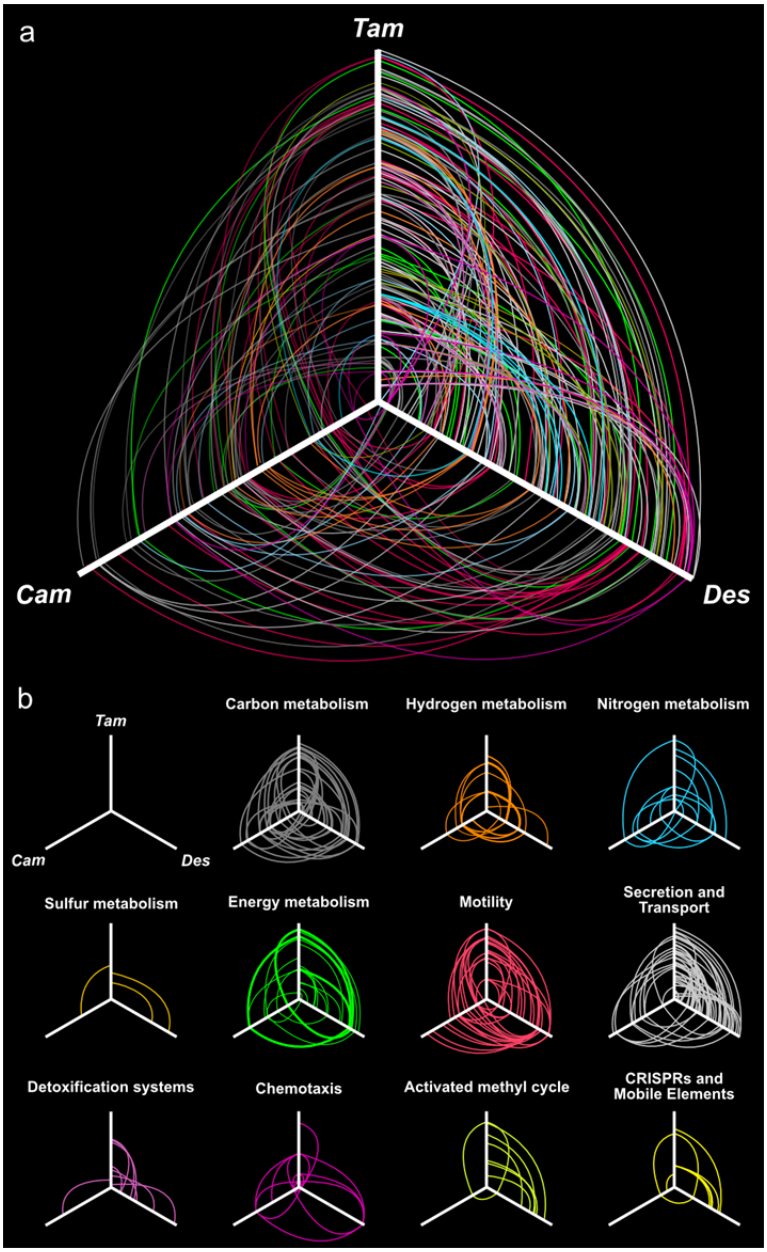


Figure 6. Blastp best hit (cut off at 40% similarity) for the *T*. *ammonificans* CDS: (a) *Desulfurobacterium thermolithotrophum* is included in the database; (b) *D*. *thermolithotrophum* and *Desulfurobacterium* sp. TC5-1 are excluded from the database (best hits are outside of the *Desulfurobacteriaceae* family). The interactive versions of the Krona plots are available for download at DOI: 10.6084/m9.figshare.3178528.



842

843 **Figure 7.** Hives plot presenting the comparative genomic analysis of the *T*. *ammonificans*
844 (Tam) genome with the closest relative available genomes of *D*. *thermolithotrophum* (Des) and the
845 ecologically similar *Epsilonproteobacterium C*. *mediatlanticus* (Cam). The three axes represent the
846 organism's linearized genomes with the origin of replication at the center of the figure. Genome size
847 was normalized for visualization purpose. (A) Localization of homologous genes and syntenic regions
848 among the three genomes. Lines colored according to the general pathway code adopted in Figure 3
849 connect homologous coding sequences. Lines opacity is proportional to similarity between sequences

(darker = higher similarity), line width is proportional to the extent of the syntenic area. (B) Hives panel of the homologous genes divided by pathways. Collinear lastZ alignments and interactive dot-plot replicating the analysis are accessible at the GenomeEvolution website with the following permanent addresses: *T*. *ammonificans* and *D*. *thermolithotrophum* (http://genomevolution.org/r/9vb3); *T*. *ammonificans* and *C*. *mediatlanticus* (http://genomevolution.org/r/9vb4); *D*. *thermolithotrophum* and *C*. *mediatlanticus* (http://genomevolution.org/r/9vb8).

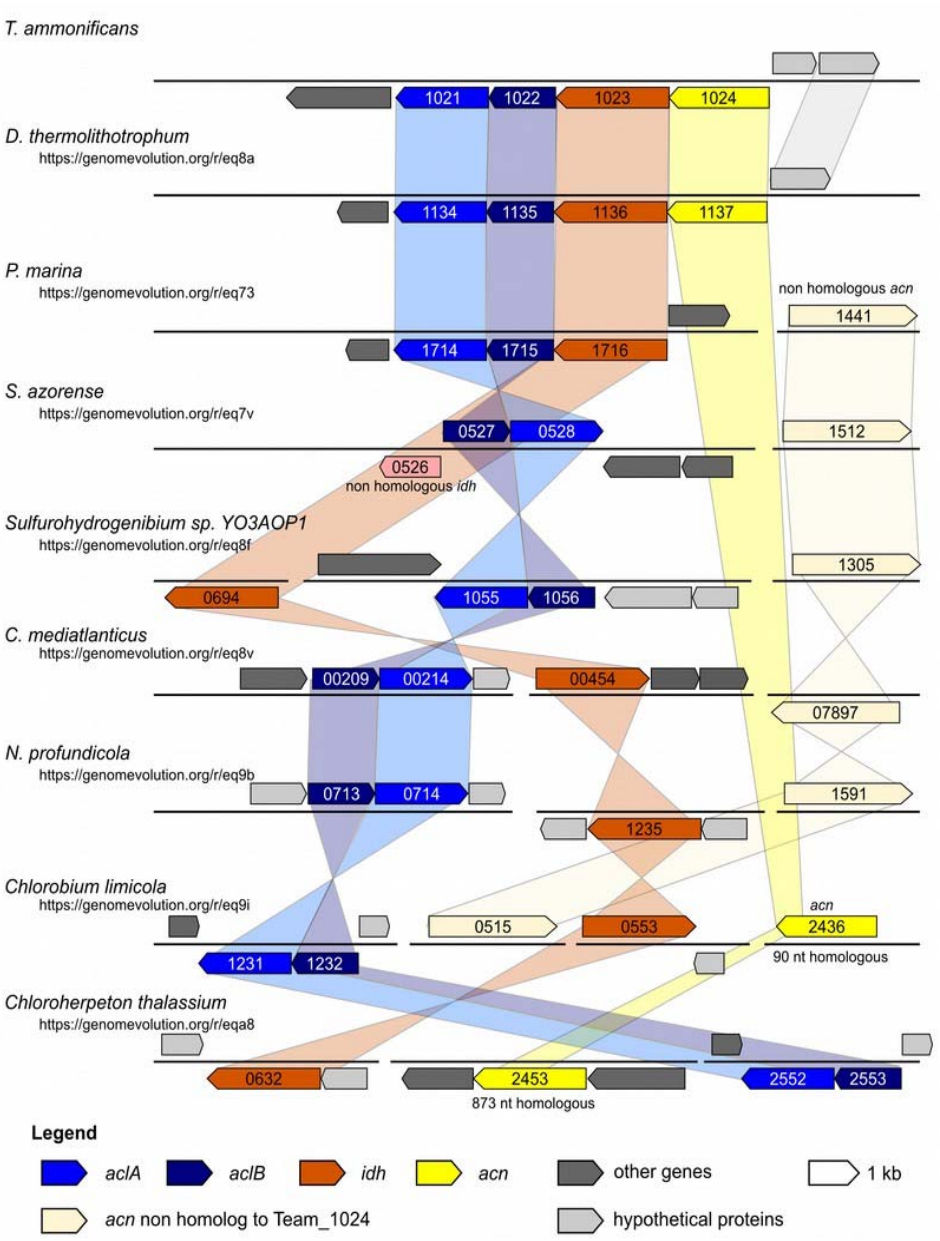
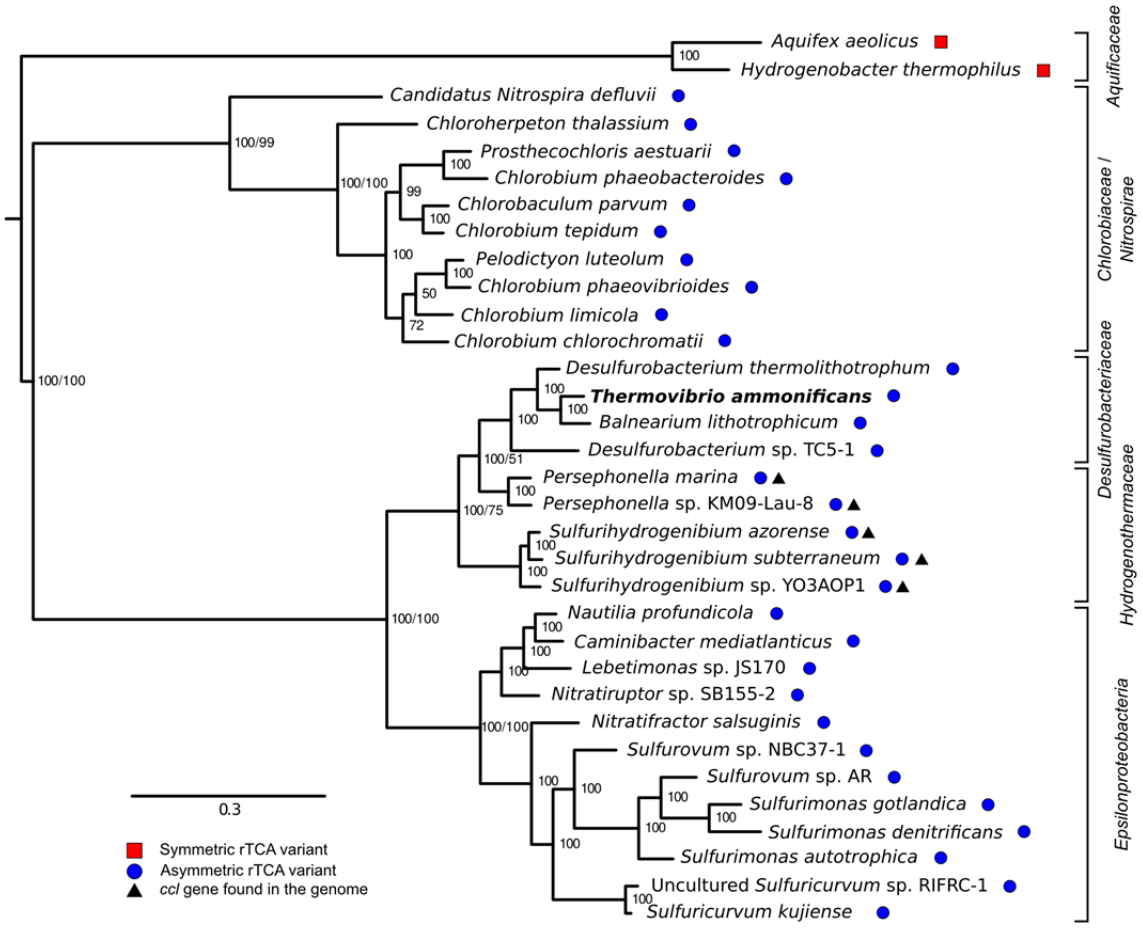
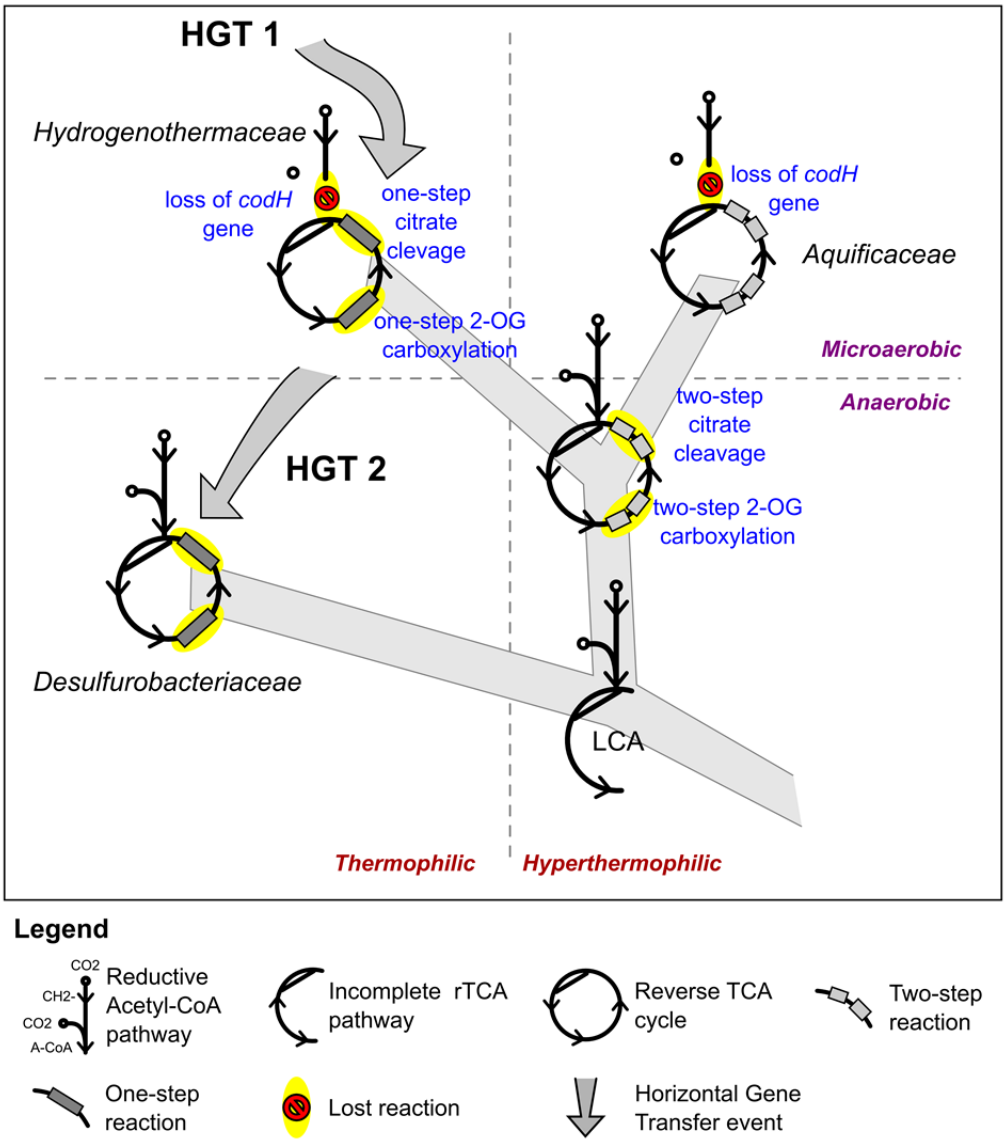


Figure 8. Syntheny diagram presenting the genome organization around the rTCA key enzyme ATP citrate lyase. In *T. ammonificans* the two subunits of the ATP citrate lyase enzyme are organized in a single operon together with the isocitrate dehydrogenase and the aconitate dehydratase. The numbers inside each gene represent the locus number for the organism. Shaded color connects synthenic regions. The website address below each organism name is a permanent link to the pairwise analysis performed on the Genome Evolution server (http://genomevolution.org/) *against* T. ammonificans



867 **Figure 9.** Phylogenetic tree of ATP citrate lyase amino acid sequences. The tree was
868 constructed with Bayesian inference and maximum likelihood methods and presents the phylogenetic
869 relationship among the concatenated subunit of the ATP citrate lyase in different organisms known to
870 use the rTCA cycle. The numbers near the nodes represent the bayesian posterior probability (left
871 number) and the maximum-likelihood confidence values based on 1000 bootstrap replications (right
872 number). Bar, 30% estimated substitutions. Accession numbers are reported in Figure 5-source data 1.
873 Squares – symmetric rTCA cycle variant characterized by the presence of two-step enzyme reactions
874 for the cleavage of citrate and the carboxylation of 2-oxoglutarate. Circles – asymmetric rTCA cycle
875 variant characterized by one-step reaction enzymes catalyzing the above reactions. Triangles –
876 presence of a citryl-CoA lyase in the genome. See Supplementary Materials for details.



877 **Figure 10. Proposed evolution of the carbon fixation pathway in the *Aquificae* phylum.**

878 Proposed evolution of the carbon fixation pathway in the *Aquificae* phylum and reconstruction of the

879 ancestral carbon fixation for the last common ancestor (LCA) based on the results of integrated

880 phylogenetic, comparative genomic and phylometabolic analyses.

Tables

Table 1. Characteristic of representative members of *Aquificae* phylum.

Family	Organism	Growth Temp	Energy Source	Carbon source	Terminal electron acceptor	Isolated from	Genome sequence accession number	References
Aquificaceae	*Aquifex aeolicus* VF5	95 ˚C	H_2	CO_2	O_2	Underwater volcanic vents, Aeolic Islands Sicily, Italy	NC_000918.1	(Deckert et al., 1998; Huber et al., 1992)
	*Hydrogenivirga sp.*128-5-R1-1[1]	75 ˚C	$S_2O_3^{2-}$, S_0	CO_2	NO_3^-, O_2	Lau Basin hydrothermal vent area, Pacific Ocean	NZ_ABHJ00000000	(Nakagawa et al., 2004; Reysenbach et al., 2009)
	Hydrogenobacter thermophilus TK-6	75 ˚C	H_2	CO_2	O_2	Hot springs in Izu and Kyushu, Japan	NC_017161.1	(Arai et al., 2010; Kawasumi et al., 1984)
	Hydrogenobaculum sp. Y04AAS1[1]	-	-	-	-	Marine hydrothermal area, Vulcano Island, Italy	NC_011126.1	(Reysenbach et al., 2009; Stohr et al., 2001)
	Thermocrinis ruber DSM 12173	85 ˚C	H_2, $S_2O_3^{2-}$, S_0	CO_2	O_2	Octopus Spring, Yellowstone National Park, Wyoming, USA	PRJNA75073	(Huber et al., 1998)
Desulfurobacteriaceae	*Balnearium lithotrophicum* 17S	70-75 ˚C	H_2	CO_2	S_0	Deep-sea hydrothermal vent chimney, Suiyo Seamount, Japan	Not sequenced	(Takai et al., 2003b)
	Desulfurobacterium thermolithotrophum DSM 11699	70 ˚C	H_2	CO_2	S_0, $S_2O_3^{2-}$, SO_2^-	Deep-sea hydrothermal chimney, mid-Atlantic ridge, Atlantic Ocean	NC_015185.1	(Goker et al., 2011; L'Haridon et al., 1998)
	Phorcysia thermohydrogeniphila HB-8	75 ˚C	H_2	CO_2	S_0, NO_3^-	Tube of *Alvinella pompejana* tubeworms, deep-sea hydrothermal vents 13 ˚N, East Pacific Rise, Pacific Ocean	Not sequenced	(Pérez-Rodríguez et al., 2012)
	Thermovibrio ammonificans HB-1	75 ˚C	H_2	CO_2	S_0, NO_3^-	Deep sea hydrothermal vents 9˚N, East Pacific Rise, Pacific Ocean	NC_014926.1	(Giovannelli et al., 2012; Vetriani et al., 2004)
Hydrogenothermaceae	*Hydrogenothermus marinus* VM1	65 ˚C	H_2	CO_2	O_2 (1-2%)	Deep sea hydrothermal vents 9˚N, East Pacific Rise, Pacific Ocean	Not sequenced	(Stohr et al., 2001)
	Persephonella marina EX-H1	73 ˚C	S_0	CO_2	O_2, NO_3^-	Deep sea hydrothermal vents 9˚N, East Pacific Rise, Pacific Ocean	NC_012440.1	(Götz et al., 2002; Reysenbach et al., 2009)
	Sulfurihydrogenibium sp. YO3AOP1[1]	70 ˚C	$S_2O_3^{2-}$, S_0	CO_2	O_2	Calcite Hot Springs, Yellowstone National Park, USA	NC_010730.1	(Reysenbach et al., 2009; Takai et al., 2003a)
	Venenivibrio stagnispumantis	70 ˚C	H_2	CO_2	O_2	Terrestrial hot spring Champagne Pool, Waiotapu, New Zealand	Not sequenced	(Hetzer et al., 2008)
Incertae sedis	*Thermosulfidibacter takai* ABI70S6^T	70 ˚C	H_2	CO_2	S_0	deep-sea hydrothermal field at Southern Okinawa Trough , Japan	Not sequenced	(Nunoura et al., 2008)

1 – Strain not formally described whose genome sequence is available. For these strains, the physiological information reported in Table S1 have been collected from MIG associated with the sequencing or from the closest validly published species.

Table 2. Enzymes involved in the putative reductive acetyl-CoA pathway in *T. ammonificans* HB-1, closest relative homolog and homologs within the *Aquificae* phylum.

					Desulfurobacteriaceae		*Hydrogenothermaceae*		*Aquificaceae*				
Reaction	**Putative enzyme**	**Putative gene locus**	**Closest relative**	**Putative origin[c]**	***D. thermolithotrophum***	***Desulfurobacterium sp. TC5-1***	***P. marina***	***Sulfurihydrogenibium sp. YO3AOP1***	***H. thermophilus***	***Hydrogenobaculum sp. Y04AAS1H***	***Hydrogenivirga sp. 128-5-R1-6***	***T. ruber***	***A. aeolicus***
1	Formate dehydrogenase	Theam_1020	*Nitratiruptor* sp. SB155-2 sim. 55% YP_001357016	*Methanogens*	f.e.[d] - 30% YP_00428203[e]	-[f]	56% YP_002730364	f.e.[d] - 32% YP_001930236	f.e.[d] - 33% YP_003433330	-	-	f.e.[d] - 27% YP_003474076	-
2[a]	5-formyltetrahydrofolate cyclo-ligase	Theam_1206	*Hydrogenivirga* sp. 128-5-R1-1 sim. 41% WP_008285842	*Deltaproteobacteria / Gram +*	41% YP_004281339	44% WP_022846876.1	f.e.[d] - 36% YP_002729868	f.e.[d] - 38% YP_001931834	f.e.[d] - 38% YP_003431710	f.e.[d] - 30% YP_002121539	41% WP_008285842	40% YP_003472981	43% 1SOU_A
	formyltetrahydrofolate deformylase / hydrolase	Theam_0826	*Hydrogenivirga* sp. 128-5-R1-1 sim. 70% WP_008287030	*Bacteria*	82% YP_004281077	77% WP_022846479	-	-	-	-	70% WP_008287030	65% NP_214247	-
	phosphoribosylglycinamide formyltransferase 2	Theam_1211	*P. marina* sim. 76% YP_002731257	*Methanogens / Deltaproteobacteria*	78% YP_004281335	64% WP_022847512	76% YP_002731257	73% YP_001931730	f.e.[d] - 28% YP_003433072	67% YP_007499479	74% WP_008288476	f.e.[d] - 25% YP_003473050	f.e.[d] - 30% NP_213168
	phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase	Theam_0328	*P. marina* sim. 59% YP_002731268	*Aquificae / Bacteria*	86% YP_004281681	78% WP_022847359	59% YP_002731268	56% YP_001931239	53% YP_003433270	50% YP_007499473	57% WP_008288360	52% YP_003474337	56% NP_214344
3	methylenetetrahydrofolate dehydrogenase / methenyltetrahydrofolate cyclohydrolase	Theam_1261	*A. aeolicus* sim. 60% NP_214304	Aquificae / Gram +	87% YP_004281147	74% WP_022846217	55% YP_002731476	58% YP_001931240	59% YP_005511318	58% YP_001931240	-	58% YP_003473104	60% NP_214304
4	methylenetetrahydrofolate reductase	Theam_0919	*Methanosrcina barkeri* sim. 48% YP_305816	Methanogens	87% YP_004281147	74% WP_022846217	55% YP_002731476	58% YP_001931240	59% YP_005511318	58% YP_002121558	f.e.[d] - 24% WP_008286783	58% YP_003473104	60% NP_214304
5	5-methyltetrahydrofolate corrinoid/iron sulfur protein methyltransferase	Theam_1184	*Clostridium glycolicum* sim. 42% WP_018589788	Gram +	82% YP_004281484	72% WP_022846603	f.e.[d] - 31% YP_002729885	f.e.[d] - 38% YP_001930381	f.e.[d] - 37% YP_003432344	f.e.[d] - 34% YP_002122031	f.e.[d] - 38% WP_008287326	f.e.[d] - 36% YP_003473438	f.e.[d] - 34% NP_213375
6[b]	Carbon-monoxide dehydrogenase	Theam_1337	*Candidatus* Methanoperedens sp. BLZ1 sim. 56% KPQ43483	Archaea / Gram +	87% WP_013638427 38% WP_013638030	76% WP_022847143	41% YP_002729904	-	-	-	-	-	-

a – Reactions are numbered according to Figure 1. For reaction 2 we reported the possible enzymes that could substitute the missing 10-fomyl-THF synthetase (Fhs). The enzymes are listed in order of decreasing likelihood of their involvement in the reaction based on putative substrate affinity.

b – Reaction 6 is catalyzed by the Acs/CodH complex, reported here separately.

c – Putative origin of the gene was calculated as consensus taxonomic assignment of the first 100 bastp hits against the nr database. Double assignment implies equal number of assignments to the two taxonomic groups.

d – f.e. = functional equivalent annotated in the genome with similarity below 40% with *T. ammonificans* equivalent gene. Not considered a true homolog in the present study.

e – Pairwise similarity to *T. ammonificans* translated gene and accession number of the homologs.

f – Missing homolog or functional annotated equivalent in the genome.

907

908 **Table 3.** List of the genomes belonging to the *Aquificae* phylum used for comparative genomic
909 analyses.
910

Organism	Genome Acc. Number	Genome Length	GC Content	Num. Genes
Aquifex aeolicus VF5	NC_000918	1,590,791	43%	1,782
Desulfurobacterium sp. TC5-1	NZ_ATXC01000001	1,653,625	40%	1,680
Desulfurobacterium thermolithotrophum DSM 11699	NC_015185	1,541,968	35%	1,561
Hydrogenivirga sp. 128-5-R1-1	NZ_ABHJ01000551	3,038,240	44%	3,756
Hydrogenobacter thermophilus TK-6	NC_013799	1,743,135	44%	1,897
Hydrogenobaculum sp. Y04AAS1	NC_011126	1,559,514	35%	1,631
Persephonella marina EX-H1	NC_012440	1,983,966	37%	2,067
Persephonella sp. IF05-L8	NZ_JNLJ01000001	1,828,858	35%	1,920
Sulfurihydrogenibium azorense Az-Fu1	NC_012438	1,640,877	33%	1,722
Sulfurihydrogenibium sp. YO3AOP1	NC_010730	1,838,442	32%	1,832
Sulfurihydrogenibium subterraneum DSM 15120	NZ_JHUV01000001	1,610,181	32%	1,701
Sulfurihydrogenibium yellowstonense SS-5	NZ_ABZS01000228	1,534,471	33%	1,637
Thermocrinis albus DSM 14484	NC_013894	1,500,577	47%	1,631
Thermocrinis ruber DSM 23557	NZ_CP007028	1,521,037	45%	1,625
Thermocrinis sp. GBS	NZ_JNIE01000001	1,315,625	41%	1,417
Thermosulfidibacter takaii ABI70S6	NZ_AP013035	1,816,670	43%	1,844
Thermovibrio ammonificans HB-1	NC_014926	1,759,526	52%	1,820

911

912

Source Data

Figure 3-source data 1. List of the proteins identified in the proteome of *T. ammonificans* grown under nitrate reducing conditions.

Locus	Product name	*Gene Name*	Accession number	MW (kDa)	NSAF %
Theam_0001	anthranilate synthase component I	*trpE*	YP_004150616.1	55	**0.09**
Theam_0005	glycosyl transferase group 1	*alg1*	YP_004150620.1	64	**0.09**
Theam_0007	DNA polymerase III, beta subunit	*dnaN*	YP_004150622.1	39	**0.25**
Theam_0010	Polyprenyl synthetase	*ggpps*	YP_004150625.1	32	**0.10**
Theam_0011	deoxyxylulose-5-phosphate synthase	*dxs*	YP_004150626.1	68	**0.02**
Theam_0025	phage protein		YP_004150640.1	40	**0.04**
Theam_0045	hypothetical protein		YP_004150660.1	9	**0.02**
Theam_0050	hypothetical protein		YP_004150665.1	35	**0.03**
Theam_0051	hypothetical protein		YP_004150666.1	33	**0.01**
Theam_0069	Rhodanese domain protein		YP_004150684.1	14	**1.06**
Theam_0070	regulatory protein ArsR	*arsR*	YP_004150685.1	14	**0.04**
Theam_0071	carbonic anhydrase	*cah*	YP_004150686.1	25	**0.04**
Theam_0073	methylthioadenosine phosphorylase	*mtaP*	YP_004150688.1	32	**0.58**
Theam_0074	50S ribosomal Protein L19 (hypothetical protein)	*rpL19*	YP_004150689.1	11	**0.09**
Theam_0075	imidazoleglycerol phosphate synthase, cyclase subunit	*hisF*	YP_004150690.1	27	**0.06**
Theam_0076	thiamine biosynthesis protein ThiC	*thiC*	YP_004150691.1	48	**0.23**
Theam_0077	thiazole biosynthesis enzyme	*thi1*	YP_004150692.1	29	**1.13**
Theam_0078	hypothetical protein		YP_004150693.1	48	**0.06**
Theam_0079	amidohydrolase	*ah*	YP_004150694.1	45	**0.02**
Theam_0080	orotate phosphoribosyltransferase	*pyrE*	YP_004150695.1	22	**0.10**
Theam_0081	Polyprenyl synthetase	*idsB*	YP_004150696.1	36	**0.11**
Theam_0084	NAD-dependent epimerase/dehydratase		YP_004150699.1	37	**0.05**
Theam_0085	chromosome segregation protein SMC	*smc*	YP_004150700.1	134	**0.01**
Theam_0088	ribosome recycling factor	*frr*	YP_004150703.1	21	**0.03**
Theam_0089	uridylate kinase	*pyrH*	YP_004150704.1	26	**0.19**
Theam_0090	translation elongation factor Ts	*tsf*	YP_004150705.1	22	**0.46**
Theam_0091	ribosomal protein S2	*rpsB_bact*	YP_004150706.1	33	**0.46**
Theam_0092	5-oxoprolinase (ATP-hydrolyzing)	*opla*	YP_004150707.1	56	**0.03**
Theam_0093	5-oxoprolinase (ATP-hydrolyzing)	*opla*	YP_004150708.1	71	**0.11**
Theam_0094	ubiquinone/menaquinone biosynthesis methyltransferase	*men_ubi*	YP_004150709.1	24	**0.01**
Theam_0095	hypothetical protein		YP_004150710.1	44	**0.00**
Theam_0098	apurinic endonuclease Apn1	*nfo*	YP_004150713.1	32	**0.02**
Theam_0100	hypothetical protein		YP_004150715.1	13	**0.10**
Theam_0101	hypothetical protein		YP_004150716.1	15	**0.02**
Theam_0102	hypothetical protein		YP_004150717.1	22	**0.02**
Theam_0103	hypothetical protein		YP_004150718.1	12	**0.05**
Theam_0104	beta-lactamase domain protein		YP_004150719.1	32	**0.09**
Theam_0105	Roadblock/LC7 family protein		YP_004150720.1	13	**0.01**
Theam_0106	pyrimidine-nucleoside phosphorylase	*pynp*	YP_004150721.1	46	**0.05**
Theam_0108	AMMECR1 domain protein		YP_004150723.1	22	**0.03**
Theam_0122	outer membrane assembly lipoprotein YfiO	*yfiO*	YP_004150737.1	37	**0.01**
Theam_0126	protein of unknown function DUF481		YP_004150741.1	25	**0.03**
Theam_0129	histone family protein DNA-binding protein		YP_004150744.1	11	**0.13**
Theam_0132	non-canonical NTP pyrophosphatase rdgB/HAM1 family	*rdgB*	YP_004150747.1	22	**0.12**
Theam_0133	ribonuclease PH	*RNasePH*	YP_004150748.1	26	**0.06**

Locus	Product name	*Gene Name*	Accession number	MW (kDa)	NSAF %
Theam_0136	ATP-dependent Clp protease, proteolytic subunit ClpP	*clpP*	YP_004150751.1	22	**0.31**
Theam_0137	3,4-dihydroxy-2-butanone 4-phosphate synthase	*ribA*	YP_004150752.1	48	**0.03**
Theam_0139	diguanylate phosphodiesterase CDS		YP_004150753.1	77	**0.00**
Theam_0140	shikimate 5-dehydrogenase	*aroE*	YP_004150754.1	30	**0.05**
Theam_0142	ribosomal protein L27	*rpl27*	YP_004150756.1	9	**0.17**
Theam_0143	ribosomal protein L21	*rpl21*	YP_004150757.1	11	**0.46**
Theam_0145	inosine-5'-monophosphate dehydrogenase	*impdh*	YP_004150759.1	53	**0.53**
Theam_0146	argininosuccinate lyase	*argH*	YP_004150760.1	52	**0.15**
Theam_0147	Transketolase domain-containing protein		YP_004150761.1	32	**0.11**
Theam_0148	Transketolase central region		YP_004150762.1	34	**0.36**
Theam_0149	carboxyl-terminal protease	*prc*	YP_004150763.1	49	**0.02**
Theam_0151	hypothetical protein		YP_004150765.1	37	**0.02**
Theam_0155	Radical SAM domain protein		YP_004150769.1	40	**0.00**
Theam_0157	chemotaxis sensory transducer		YP_004150771.1	68	**0.06**
Theam_0159	dihydroorotate dehydrogenase family protein	*pyrD1*	YP_004150773.1	32	**0.11**
Theam_0161	Tetratricopeptide TPR_1 repeat-containing protein		YP_004150775.1	67	**0.01**
Theam_0163	peptidase M17 leucyl aminopeptidase domain protein		YP_004150777.1	52	**0.03**
Theam_0165	chemotaxis sensory transducer		YP_004150779.1	86	**0.02**
Theam_0168	phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase	*hisA*	YP_004150782.1	27	**0.15**
Theam_0170	quinolinate synthetase complex, A subunit	*nadA*	YP_004150784.1	35	**0.01**
Theam_0171	GMP synthase, large subunit	*guaA*	YP_004150785.1	59	**0.12**
Theam_0179	adenylosuccinate synthetase	*purA*	YP_004150793.1	48	**0.09**
Theam_0180	histidyl-tRNA synthetase 2	*hisS2*	YP_004150794.1	48	**0.01**
Theam_0181	threonine synthase	*thrC*	YP_004150795.1	38	**0.16**
Theam_0182	dihydrodipicolinate synthase	*dapA*	YP_004150796.1	32	**0.35**
Theam_0183	hypothetical protein		YP_004150797.1	40	**0.00**
Theam_0184	dihydrodipicolinate reductase	*dapB*	YP_004150798.1	28	**0.26**
Theam_0188	hypothetical protein		YP_004150802.1	101	**0.01**
Theam_0190	homoserine dehydrogenase	*ak1h*	YP_004150804.1	47	**0.09**
Theam_0191	aminotransferase class I and II		YP_004150805.1	46	**0.09**
Theam_0195	TonB-dependent receptor plug		YP_004150809.1	67	**0.08**
Theam_0197	Citryl-CoA lyase	*citE*	YP_004150811.1	37	**0.06**
Theam_0198	LysR substrate-binding	*lysR*	YP_004150812.1	34	**0.01**
Theam_0201	3-isopropylmalate dehydratase, large subunit	*haco*	YP_004150815.1	46	**0.13**
Theam_0205	nicotinate (nicotinamide) nucleotide adenylyltransferase		YP_004150819.1	25	**0.03**
Theam_0206	Phosphoglycerate kinase	*pgk*	YP_004150820.1	45	**0.17**
Theam_0210	Peptidoglycan-binding lysin domain		YP_004150824.1	45	**0.00**
Theam_0215	methionyl-tRNA formyltransferase	*fmt*	YP_004150829.1	35	**0.02**
Theam_0217	protein of unknown function DUF89		YP_004150831.1	34	**0.03**
Theam_0218	phosphoribosylaminoimidazole carboxylase, catalytic subunit	*purE*	YP_004150832.1	17	**0.21**
Theam_0220	translation elongation factor P	*efp*	YP_004150834.1	21	**0.06**
Theam_0221	acetyl-CoA carboxylase, biotin carboxyl carrier protein	*accB*	YP_004150835.1	16	**0.23**
Theam_0222	acetyl-CoA carboxylase, biotin carboxylase	*accC*	YP_004150836.1	49	**0.03**
Theam_0223	Like-Sm ribonucleoprotein core		YP_004150837.1	17	**0.04**
Theam_0227	aminotransferase class V - serine glyoxylate	*sgt*	YP_004150841.1	42	**0.17**
Theam_0228	peptidylprolyl isomerase FKBP-type		YP_004150842.1	16	**0.05**
Theam_0229	flavin reductase domain protein FMN-binding		YP_004150843.1	21	**0.11**
Theam_0230	UbiD family decarboxylase	*ubid*	YP_004150844.1	56	**0.00**
Theam_0231	NIF3 containing protein of unknown function DUF34		YP_004150845.1	28	**0.04**
Theam_0233	ATP-dependent protease La	*lon*	YP_004150847.1	92	**0.03**

Locus	Product name	*Gene Name*	Accession number	MW (kDa)	NSAF %
Theam_0234	UTP-glucose-1-phosphate uridylyltransferase	*galU*	YP_004150848.1	34	**0.12**
Theam_0235	protein of unknown function Met10		YP_004150849.1	45	**0.02**
Theam_0237	carbamoyl-phosphate synthase, small subunit	*cpsS*	YP_004150851.1	42	**0.08**
Theam_0238	periplasmic solute binding protein		YP_004150852.1	32	**0.03**
Theam_0239	glutamine amidotransferase of anthranilate synthase	*trpG_papA*	YP_004150853.1	21	**0.02**
Theam_0244	dihydrouridine synthase DuS	*dus*	YP_004150858.1	35	**0.03**
Theam_0245	ribosomal protein S15	*S15*	YP_004150859.1	11	**0.09**
Theam_0246	polyribonucleotide nucleotidyltransferase	*pnp*	YP_004150860.1	79	**0.19**
Theam_0247	deoxyUTP pyrophosphatase	*dut*	YP_004150861.1	32	**0.06**
Theam_0250	2-isopropylmalate synthase	*leuA*	YP_004150864.1	56	**0.20**
Theam_0252	ribosomal protein S12	*rpsL*	YP_004150866.1	14	**0.27**
Theam_0253	ribosomal protein S7	*rpsG*	YP_004150867.1	18	**0.07**
Theam_0254	translation elongation factor G	*efG*	YP_004150868.1	78	**0.88**
Theam_0256	ribosomal protein S10	*rpsJ*	YP_004150870.1	12	**0.96**
Theam_0257	50S ribosomal protein L3	*rpl3*	YP_004150871.1	22	**0.34**
Theam_0258	ribosomal protein L4/L1e	*rpsE*	YP_004150872.1	24	**0.14**
Theam_0259	Ribosomal protein L25/L23	*rplL*	YP_004150873.1	12	**0.44**
Theam_0260	ribosomal protein L2	*rplB*	YP_004150874.1	30	**0.26**
Theam_0261	ribosomal protein S19	*rpsS*	YP_004150875.1	11	**0.08**
Theam_0262	ribosomal protein L22	*rplV*	YP_004150876.1	15	**0.38**
Theam_0263	ribosomal protein S3	*rpsC*	YP_004150877.1	27	**0.29**
Theam_0264	ribosomal protein L16	*rplP*	YP_004150878.1	16	**0.36**
Theam_0266	30S ribosomal protein S17	*rps17*	YP_004150880.1	12	**0.18**
Theam_0267	ribosomal protein L14	*rplN*	YP_004150881.1	13	**1.18**
Theam_0268	ribosomal protein L24	*rplX*	YP_004150882.1	13	**0.13**
Theam_0269	50S ribosomal protein L5	*rpl5*	YP_004150883.1	21	**0.56**
Theam_0271	ribosomal protein S8	*rps8*	YP_004150885.1	16	**0.44**
Theam_0272	ribosomal protein L6	*rpl6*	YP_004150886.1	20	**0.23**
Theam_0273	ribosomal protein L18	*rpl18*	YP_004150887.1	14	**0.39**
Theam_0274	ribosomal protein S5	*rpsE*	YP_004150888.1	20	**0.34**
Theam_0275	ribosomal protein L30	*rpmD*	YP_004150889.1	7	**0.19**
Theam_0276	ribosomal protein L15	*rplO*	YP_004150890.1	17	**0.04**
Theam_0278	adenylate kinase	*adk*	YP_004150892.1	21	**0.27**
Theam_0279	methionine aminopeptidase, type I	*map1*	YP_004150893.1	28	**0.02**
Theam_0282	30S ribosomal protein S13	*rps13*	YP_004150896.1	14	**0.37**
Theam_0283	30S ribosomal protein S11	*rps11*	YP_004150897.1	14	**0.16**
Theam_0284	ribosomal protein S4	*rpsD*	YP_004150898.1	25	**0.24**
Theam_0285	DNA-directed RNA polymerase, alpha subunit	*rpoA*	YP_004150899.1	35	**0.22**
Theam_0286	ribosomal protein L17	*rpl17*	YP_004150900.1	15	**0.24**
Theam_0288	septum site-determining protein MinD	*minD*	YP_004150902.1	29	**0.01**
Theam_0291	spermidine synthase	*speE*	YP_004150905.1	33	**0.14**
Theam_0292	protein of unknown function DUF43		YP_004150906.1	40	**0.30**
Theam_0295	purine or other phosphorylase family 1		YP_004150909.1	33	**0.00**
Theam_0297	ribosomal 5S rRNA E-loop binding protein Ctc/L25/TL5	*ctc*	YP_004150911.1	22	**0.10**
Theam_0298	peptidyl-tRNA hydrolase	*pth*	YP_004150912.1	21	**0.03**
Theam_0299	ribosomal protein S6	*S6*	YP_004150913.1	15	**0.34**
Theam_0301	ribosomal protein S18	*S18*	YP_004150915.1	10	**0.02**
Theam_0303	ribosomal protein L9	*rpl9*	YP_004150917.1	17	**0.13**
Theam_0304	S-adenosylmethionine decarboxylase related	*speD*	YP_004150918.1	20	**0.21**
Theam_0306	endoribonuclease L-PSP	*lpsp*	YP_004150920.1	14	**0.06**
Theam_0307	regulatory protein ArsR	*arsR*	YP_004150921.1	12	**0.06**

Locus	Product name	*Gene Name*	Accession number	MW (kDa)	NSAF %
Theam_0308	acetyl-CoA carboxylase, carboxyl transferase, beta subunit	*accD*	YP_004150922.1	31	**0.01**
Theam_0310	transcription-repair coupling factor	*mfd*	YP_004150924.1	119	**0.00**
Theam_0311	Ankyrin		YP_004150925.1	26	**0.01**
Theam_0313	hypothetical protein		YP_004150927.1	43	**0.01**
Theam_0315	excinuclease ABC, A subunit	*uvra*	YP_004150929.1	105	**0.00**
Theam_0316	DsrE family protein		YP_004150930.1	13	**0.24**
Theam_0324	CMP/dCMP deaminase zinc-binding		YP_004150938.1	14	**0.07**
Theam_0325	response regulator receiver		YP_004150939.1	26	**0.03**
Theam_0326	hypothetical protein		YP_004150940.1	22	**0.06**
Theam_0327	Glutamate synthase (ferredoxin)	*gltA*	YP_004150941.1	159	**0.17**
Theam_0328	phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase	*purH*	YP_004150942.1	59	**0.14**
Theam_0331	alanine racemase	*alr*	YP_004150945.1	39	**0.01**
Theam_0332	ribosomal L11 methyltransferase	*rpl11*	YP_004150946.1	30	**0.01**
Theam_0333	PHP domain protein		YP_004150947.1	31	**0.01**
Theam_0334	Protein of unknown function DUF2081		YP_004150948.1	74	**0.00**
Theam_0337	heat shock protein HslVU, ATPase subunit HslU	*hslU*	YP_004150951.1	55	**0.01**
Theam_0338	peptidase M24	*pM24*	YP_004150952.1	38	**0.03**
Theam_0342	CRISPR-associated protein, Csx11 family		YP_004150956.1	123	**0.00**
Theam_0344	CRISPR-associated RAMP protein, Cmr1 family		YP_004150958.1	50	**0.00**
Theam_0347	3-isopropylmalate dehydratase, small subunit	*leuA*	YP_004150961.1	18	**0.19**
Theam_0348	3-isopropylmalate dehydrogenase	*leuB*	YP_004150962.1	40	**0.38**
Theam_0349	aspartate-semialdehyde dehydrogenase	*dhaS*	YP_004150963.1	37	**0.23**
Theam_0350	hypothetical protein		YP_004150964.1	14	**0.02**
Theam_0351	metal-dependent hydrolase		YP_004150965.1	25	**0.18**
Theam_0354	Peptidase M23	*pM23*	YP_004150968.1	39	**0.00**
Theam_0356	TrkA-N domain protein		YP_004150970.1	48	**0.03**
Theam_0357	cytochrome c-type biogenesis protein CcsB	*ccsB*	YP_004150971.1	38	**0.01**
Theam_0358	ResB family protein		YP_004150972.1	50	**0.00**
Theam_0360	metal dependent phophohydrolase	*hdig*	YP_004150974.1	55	**0.01**
Theam_0362	ATPase AAA-2 domain protein		YP_004150976.1	94	**0.01**
Theam_0363	protein of unknown function DUF558		YP_004150977.1	26	**0.07**
Theam_0365	Appr-1-p processing domain protein		YP_004150979.1	19	**0.12**
Theam_0373	2-isopropylmalate synthase/homocitrate synthase family protein	*leu1*	YP_004150987.1	60	**0.16**
Theam_0374	aspartate kinase	*aspK*	YP_004150988.1	44	**0.40**
Theam_0375	molybdenum cofactor synthesis domain protein	*moeA*	YP_004150989.1	44	**0.00**
Theam_0376	glyceraldehyde-3-phosphate dehydrogenase	*gapor*	YP_004150990.1	75	**0.01**
Theam_0377	N-acetyl-gamma-glutamyl-phosphate reductase	*argC*	YP_004150991.1	39	**0.09**
Theam_0378	ribosomal protein S9	*rps9*	YP_004150992.1	15	**0.44**
Theam_0379	ribosomal protein L13	*rplM*	YP_004150993.1	16	**0.02**
Theam_0385	ribulose-phosphate 3-epimerase	*rpe*	YP_004150999.1	24	**0.13**
Theam_0386	ribose-phosphate pyrophosphokinase	*rpppk*	YP_004151000.1	34	**0.73**
Theam_0389	pyridoxal phosphate biosynthetic protein PdxJ	*pdxJ*	YP_004151003.1	27	**0.22**
Theam_0392	acyl carrier protein	*acp*	YP_004151006.1	9	**0.10**
Theam_0393	3-oxoacyl-(acyl-carrier-protein) reductase	*fabG*	YP_004151007.1	26	**0.07**
Theam_0397	hypothetical protein		YP_004151011.1	24	**0.12**
Theam_0398	phosphodiesterase, MJ0936 family		YP_004151012.1	18	**0.08**
Theam_0400	Conserved TM helix repeat-containing protein		YP_004151014.1	41	**0.01**
Theam_0404	HI0933 family protein		YP_004151017.1	57	**0.01**
Theam_0410	regulatory protein LuxR	*luxR*	YP_004151023.1	22	**0.01**
Theam_0412	thymidylate kinase	*kthY*	YP_004151025.1	24	**0.05**

Locus	Product name	*Gene Name*	Accession number	MW (kDa)	NSAF %
Theam_0414	TrkA-C domain protein		YP_004151027.1	18	**0.01**
Theam_0415	hypothetical protein		YP_004151028.1	23	**0.01**
Theam_0417	PSP1 domain protein		YP_004151030.1	35	**0.01**
Theam_0418	hypothetical protein		YP_004151031.1	52	**0.19**
Theam_0423	NapA nitrate reductase periplasmic molybdopterin oxidoreductase	*napA*	YP_004151036.1	101	**0.10**
Theam_0429	arginine biosynthesis bifunctional protein ArgJ	*argJ*	YP_004151042.1	41	**0.20**
Theam_0430	tRNA (guanine-N1-)-methyltransferase	*trmD*	YP_004151043.1	41	**0.01**
Theam_0431	cysteine synthase	*cysK*	YP_004151044.1	33	**0.10**
Theam_0433	hypothetical protein		YP_004151046.1	15	**0.04**
Theam_0436	tryptophan synthase, alpha subunit	*trpA*	YP_004151049.1	29	**0.20**
Theam_0437	twin-arginine translocation protein, TatA/E family subunit	*tatAE*	YP_004151050.1	9	**0.70**
Theam_0439	transcription termination factor Rho	*rho*	YP_004151052.1	48	**0.04**
Theam_0440	phosphoheptose isomerase	*gmhA*	YP_004151053.1	22	**0.09**
Theam_0441	predicted fructose-bisphosphate aldolase	*fba*	YP_004151054.1	28	**1.71**
Theam_0442	Radical SAM domain protein		YP_004151055.1	63	**0.00**
Theam_0445	ABC transporter related for Fe-S assembly	*sufC*	YP_004151058.1	27	**0.14**
Theam_0446	SufBD protein for Fe-S assembly	*sufB*	YP_004151059.1	35	**0.08**
Theam_0447	Putative Superoxide reductase (Desulfoferrodoxin ferrous iron-binding region)	*sor*	YP_004151060.1	13	**0.15**
Theam_0449	Ppx/GppA phosphatase	*gppA*	YP_004151062.1	34	**0.04**
Theam_0450	nucleotide sugar dehydrogenase UDP glucose type	*udg*	YP_004151063.1	49	**0.06**
Theam_0452	2,5-diamino-6-hydroxy-4-(5-phosphoribosylamino)pyrimidine 1-reductase	*ribD*	YP_004151065.1	24	**0.13**
Theam_0456	regulatory protein ArsR	*arsR*	YP_004151069.1	12	**0.08**
Theam_0459	ketol-acid reductoisomerase	*ilvC*	YP_004151072.1	37	**0.71**
Theam_0460	acetolactate synthase, small subunit	*acolacS*	YP_004151073.1	20	**0.05**
Theam_0461	acetolactate synthase, large subunit, biosynthetic type	*acolacL*	YP_004151074.1	64	**0.18**
Theam_0464	KpsF/GutQ family protein	*kpsF*	YP_004151077.1	29	**0.03**
Theam_0467	[Glutamate--ammonia-ligase] adenylyltransferase	*glnE*	YP_004151080.1	102	**0.05**
Theam_0470	ribosomal protein L34	*rpmH*	YP_004151083.1	7	**0.03**
Theam_0473	membrane protein insertase, YidC/Oxa1 family	*yidC*	YP_004151086.1	55	**0.01**
Theam_0478	NADH dehydrogenase (quinone) Ni-Fe subunit III	*echE*	YP_004151091.1	41	**0.02**
Theam_0480	NADH ubiquinone oxidoreductase kDa subunit	*echC*	YP_004151093.1	20	**0.02**
Theam_0481	respiratory-chain NADH dehydrogenase subunit 1	*echB*	YP_004151094.1	31	**0.01**
Theam_0483	cyclic nucleotide-binding		YP_004151096.1	19	**0.01**
Theam_0484	nickel-dependent hydrogenase large subunit	*hydA*	YP_004151097.1	47	**0.01**
Theam_0485	NADH ubiquinone oxidoreductase 20 kDa subunit	*fqhD*	YP_004151098.1	28	**0.01**
Theam_0487	cytochrome-c3 hydrogenase alpha chain	*hydB*	YP_004151100.1	41	**0.02**
Theam_0488	pantetheine-phosphate adenylyltransferase	*coaD*	YP_004151101.1	19	**0.04**
Theam_0489	Aldehyde Dehydrogenase	*ald*	YP_004151102.1	52	**0.28**
Theam_0490	metallophosphoesterase		YP_004151103.1	26	**0.01**
Theam_0491	tRNA synthetase class II (D K and N)		YP_004151104.1	35	**0.03**
Theam_0492	3-octaprenyl-4-hydroxybenzoate carboxy-lyase	*ubiX*	YP_004151105.1	21	**0.08**
Theam_0493	acetylglutamate kinase	*argB*	YP_004151106.1	32	**0.12**
Theam_0494	cytochrome bd ubiquinol oxidase subunit I	*cydA*	YP_004151107.1	52	**0.01**
Theam_0495	cytochrome d ubiquinol oxidase, subunit II	*cydB*	YP_004151108.1	44	**0.01**
Theam_0500	RNA chaperone Hfq	*hfq*	YP_004151113.1	9	**0.05**
Theam_0502	glutamyl-tRNA synthetase	*gltXb*	YP_004151115.1	56	**0.08**
Theam_0505	Uroporphyrinogen III synthase HEM4		YP_004151118.1	27	**0.01**
Theam_0506	porphobilinogen deaminase	*hemC*	YP_004151119.1	35	**0.14**
Theam_0507	glutamyl-tRNA reductase	*hemA*	YP_004151120.1	47	**0.00**

Locus	Product name	*Gene Name*	Accession number	MW (kDa)	NSAF %
Theam_0509	3-dehydroquinate dehydratase, type I	*aroD*	YP_004151122.1	28	**0.14**
Theam_0510	protein of unknown function DUF190		YP_004151123.1	13	**0.02**
Theam_0511	type IV-A pilus assembly ATPase PilB	*pilB*	YP_004151124.1	64	**0.01**
Theam_0512	hypothetical protein		YP_004151125.1	24	**0.14**
Theam_0514	3-phosphoshikimate 1-carboxyvinyltransferase	*aroA*	YP_004151127.1	47	**0.08**
Theam_0520	Porphobilinogen synthase	*hemC*	YP_004151133.1	37	**0.87**
Theam_0521	hypothetical protein		YP_004151134.1	13	**0.30**
Theam_0522	beta-lactamase domain-containing protein		YP_004151135.1	31	**0.07**
Theam_0526	acetylornithine and succinylornithine aminotransferase	*argD*	YP_004151139.1	43	**0.11**
Theam_0534	dihydroxy-acid dehydratase	*ilvD*	YP_004151147.1	59	**0.17**
Theam_0537	peptidase M22 glycoprotease	*pM22*	YP_004151150.1	21	**0.02**
Theam_0542	hypothetical protein		YP_004151155.1	6	**0.05**
Theam_0544	hypothetical protein		YP_004151157.1	74	**0.01**
Theam_0548	acetate/CoA ligase - Acetyl-CoA Synthase	*acsA*	YP_004151161.1	71	**0.14**
Theam_0549	excinuclease ABC, B subunit	*uvrb*	YP_004151162.1	77	**0.00**
Theam_0551	Tetratricopeptide TPR_1 repeat-containing protein		YP_004151164.1	104	**0.02**
Theam_0552	protein of unknown function DUF507		YP_004151165.1	23	**0.01**
Theam_0553	hypothetical protein		YP_004151166.1	42	**0.01**
Theam_0566	Radical SAM domain protein		YP_004151179.1	25	**0.02**
Theam_0569	1-deoxy-D-xylulose 5-phosphate reductoisomerase	*Dxr*	YP_004151182.1	41	**0.04**
Theam_0570	histidyl-tRNA synthetase	*hisS1*	YP_004151183.1	47	**0.04**
Theam_0571	phosphoribosyltransferase		YP_004151184.1	24	**0.11**
Theam_0574	hypothetical protein		YP_004151187.1	32	**0.04**
Theam_0575	glycosyl transferase family 2		YP_004151188.1	111	**0.01**
Theam_0576	Methyltransferase type 12		YP_004151189.1	30	**0.01**
Theam_0577	glucose-1-phosphate thymidylyltransferase	*rmlA*	YP_004151190.1	33	**0.05**
Theam_0578	dTDP-4-dehydrorhamnose 3,5-epimerase	*rmlC*	YP_004151191.1	21	**0.06**
Theam_0579	dTDP-glucose 4,6-dehydratase	*rmlB*	YP_004151192.1	37	**0.01**
Theam_0581	histone deacetylase superfamily		YP_004151194.1	33	**0.03**
Theam_0582	ribonuclease R	*3xrn*	YP_004151195.1	82	**0.07**
Theam_0586	enolase	*eno*	YP_004151199.1	47	**1.09**
Theam_0589	succinyldiaminopimelate transaminase		YP_004151202.1	44	**0.02**
Theam_0590	hypothetical protein		YP_004151203.1	28	**0.14**
Theam_0591	alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen		YP_004151204.1	18	**0.04**
Theam_0592	ribosomal protein S20	*S20*	YP_004151205.1	11	**0.20**
Theam_0593	thiamine-phosphate pyrophosphorylase	*thiE*	YP_004151206.1	24	**0.03**
Theam_0599	glycosyl transferase family 9		YP_004151212.1	33	**0.00**
Theam_0601	nucleotide sugar dehydrogenase		YP_004151214.1	49	**0.04**
Theam_0602	malate dehydrogenase, NAD-dependent	*mdh*	YP_004151215.1	34	**0.41**
Theam_0603	Fumarate hydratase, Fe-S type, tartrate/fumarate subfamily, alpha subunit	*fumA*	YP_004151216.1	31	**0.27**
Theam_0605	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase	*ispG*	YP_004151218.1	38	**0.03**
Theam_0607	diguanylate cyclase	*ggdef*	YP_004151220.1	42	**0.04**
Theam_0609	homocysteine S-methyltransferase	*metH*	YP_004151222.1	90	**0.04**
Theam_0610	adenine phosphoribosyltransferase	*apt*	YP_004151223.1	20	**1.25**
Theam_0611	3-oxoacyl-[acyl-carrier-protein] synthase 2		YP_004151224.1	44	**0.12**
Theam_0612	ribonuclease III		YP_004151225.1	27	**0.02**
Theam_0613	maf protein	*maf*	YP_004151226.1	21	**0.18**
Theam_0614	hypothetical protein		YP_004151227.1	27	**0.04**
Theam_0616	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase		YP_004151229.1	18	**0.01**
Theam_0617	beta-hydroxyacyl-(acyl-carrier-protein) dehydratase FabZ	*fabZ*	YP_004151230.1	17	**0.28**

Locus	Product name	*Gene Name*	Accession number	MW (kDa)	NSAF %
Theam_0618	acyl-[acyl-carrier-protein]--UDP-N-acetylglucosamine O-acyltransferase	*lpxA*	YP_004151231.1	28	**0.09**
Theam_0620	Nucleoside-triphosphatase		YP_004151233.1	20	**0.02**
Theam_0621	Hsp33 protein		YP_004151234.1	37	**0.24**
Theam_0622	peptidase zinc-dependent		YP_004151235.1	19	**0.06**
Theam_0624	Purine-nucleoside phosphorylase		YP_004151237.1	29	**0.03**
Theam_0625	hypothetical protein		YP_004151238.1	63	**0.00**
Theam_0626	ammonium transporter		YP_004151239.1	12	**0.10**
Theam_0629	glutamine synthetase, type I	*glnA*	YP_004151242.1	53	**0.95**
Theam_0631	diaminopimelate decarboxylase	*lysA*	YP_004151243.1	46	**0.05**
Theam_0633	Prephenate dehydrogenase		YP_004151245.1	32	**0.10**
Theam_0634	class II aldolase/adducin family protein		YP_004151246.1	21	**0.04**
Theam_0635	histidine kinase HAMP region domain protein		YP_004151247.1	37	**0.02**
Theam_0636	hypothetical protein		YP_004151248.1	33	**0.02**
Theam_0638	protein of unknown function DUF178		YP_004151249.1	32	**0.01**
Theam_0640	rfaE bifunctional protein	*rfaE*	YP_004151251.1	36	**0.04**
Theam_0642	isoleucyl-tRNA synthetase	*ileS*	YP_004151253.1	108	**0.08**
Theam_0644	Rhodanese domain protein		YP_004151255.1	15	**0.02**
Theam_0645	phosphoglycerate mutase, 2,3-bisphosphoglycerate-independent	*pgm*	YP_004151256.1	58	**0.03**
Theam_0646	tRNA pseudouridine synthase A	*hisT_truA*	YP_004151257.1	30	**0.01**
Theam_0648	glutamate--cysteine ligase GCS2		YP_004151259.1	40	**0.02**
Theam_0649	gamma-glutamyltransferase		YP_004151260.1	54	**0.04**
Theam_0651	protein-P-II uridylyltransferase	*glnD*	YP_004151262.1	102	**0.01**
Theam_0654	UDP-glucose 4-epimerase	*galE*	YP_004151265.1	36	**0.02**
Theam_0655	triosephosphate isomerase	*tim*	YP_004151266.1	27	**0.10**
Theam_0656	ATP-dependent Clp protease, ATP-binding subunit ClpX	*clpX*	YP_004151267.1	46	**0.07**
Theam_0657	N-(5'phosphoribosyl)anthranilate isomerase (PRAI)		YP_004151268.1	23	**0.10**
Theam_0658	Ferritin Dps family protein		YP_004151269.1	19	**0.09**
Theam_0661	aspartyl-tRNA synthetase	*aspSb*	YP_004151271.1	68	**0.11**
Theam_0663	3'(2'),5'-bisphosphate nucleotidase	*cysQ*	YP_004151273.1	29	**0.07**
Theam_0666	sulfate adenylyltransferase	*aps*	YP_004151276.1	44	**0.03**
Theam_0671	glycosyl transferase group 1		YP_004151281.1	40	**0.00**
Theam_0677	RNA polymerase-binding protein DksA		YP_004151286.1	15	**0.02**
Theam_0678	aminotransferase class I and II		YP_004151287.1	44	**0.17**
Theam_0683	EAL domain protein		YP_004151292.1	75	**0.03**
Theam_0686	NAD-dependent glycerol-3-phosphate dehydrogenase domain protein		YP_004151295.1	35	**0.03**
Theam_0688	phosphate transport system regulatory protein PhoU	*phoU*	YP_004151297.1	27	**0.21**
Theam_0691	NADPH-dependent FMN reductase	*fmnr*	YP_004151300.1	21	**0.03**
Theam_0692	ribosome-associated GTPase EngA	*engA*	YP_004151301.1	55	**0.03**
Theam_0693	hypothetical protein		YP_004151302.1	24	**0.04**
Theam_0694	hypothetical protein		YP_004151303.1	21	**0.04**
Theam_0701	lysine 2,3-aminomutase YodO family protein	*yodO*	YP_004151310.1	43	**0.04**
Theam_0703	pseudogene			18	**0.05**
Theam_0704	histidinol-phosphate aminotransferase	*hisC*	YP_004151312.1	42	**0.09**
Theam_0709	3-methyl-2-oxobutanoate hydroxymethyltransferase	*panB*	YP_004151317.1	29	**0.07**
Theam_0710	ribosome small subunit-dependent GTPase A	*rsgA*	YP_004151318.1	34	**0.03**
Theam_0712	ketose-bisphosphate aldolase class-II		YP_004151320.1	53	**0.22**
Theam_0714	L-aspartate oxidase	*nadB*	YP_004151322.1	57	**0.01**
Theam_0718	CTP synthase	*PyrG*	YP_004151326.1	60	**0.04**
Theam_0719	3-deoxy-D-manno-octulosonate cytidylyltransferase	*kdsB*	YP_004151327.1	27	**0.02**

Locus	Product name	*Gene Name*	Accession number	MW (kDa)	NSAF %
Theam_0720	phosphoribosylformylglycinamidine synthase I	*fgams*	YP_004151328.1	25	**0.14**
Theam_0721	phosphoribosylformylglycinamidine synthase, purS	*purS*	YP_004151329.1	10	**0.16**
Theam_0725	RNA polymerase sigma factor RpoD	*rpoD*	YP_004151333.1	66	**0.01**
Theam_0727	Peroxiredoxin		YP_004151335.1	22	**0.76**
Theam_0730	Fumarate hydratase, Fe-S type, tartrate/fumarate subfamily, beta subunit	*fumB*	YP_004151338.1	21	**0.25**
Theam_0734	NADH dehydrogenase I, D subunit	*nuoD*	YP_004151342.1	45	**0.02**
Theam_0736	NADH-quinone oxidoreductase, F subunit	*nuoF*	YP_004151344.1	46	**0.00**
Theam_0737	NADH:ubiquinone oxidoreductase, subunit G, iron-sulphur binding	*nuoG*	YP_004151345.1	80	**0.06**
Theam_0738	putative transcriptional regulator, Crp/Fnr family		YP_004151346.1	18	**0.04**
Theam_0746	L-seryl-tRNA selenium transferase	*selA*	YP_004151354.1	52	**0.01**
Theam_0747	TonB-dependent receptor		YP_004151355.1	76	**0.00**
Theam_0751	Peptidoglycan-binding lysin domain		YP_004151359.1	27	**0.14**
Theam_0752	hypothetical protein		YP_004151360.1	13	**0.10**
Theam_0753	acriflavin resistance protein		YP_004151361.1	121	**0.03**
Theam_0754	efflux transporter, RND family, MFP subunit	*mfp*	YP_004151362.1	42	**0.05**
Theam_0755	glutamine amidotransferase class-I	*gmps*	YP_004151363.1	26	**0.07**
Theam_0756	peptidase U62 modulator of DNA gyrase	*tldD*	YP_004151364.1	50	**0.09**
Theam_0757	NAD+ synthetase	*nadE*	YP_004151365.1	31	**0.11**
Theam_0758	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase		YP_004151366.1	32	**0.09**
Theam_0763	PhoH family protein		YP_004151371.1	36	**0.08**
Theam_0765	SMC domain protein		YP_004151373.1	105	**0.00**
Theam_0768	Radical SAM domain protein		YP_004151376.1	35	**0.02**
Theam_0770	phosphoesterase RecJ domain protein		YP_004151378.1	37	**0.10**
Theam_0772	peptidase U62 modulator of DNA gyrase	*tldD*	YP_004151380.1	48	**0.11**
Theam_0773	thymidylate synthase, flavin-dependent		YP_004151381.1	23	**0.04**
Theam_0778	NHL repeat containing protein		YP_004151386.1	29	**0.04**
Theam_0779	hypothetical protein		YP_004151387.1	27	**0.01**
Theam_0784	hypothetical protein		YP_004151392.1	50	**0.00**
Theam_0785	molybdenum cofactor synthesis domain protein		YP_004151393.1	43	**0.01**
Theam_0786	molybdopterin binding domain		YP_004151394.1	38	**0.02**
Theam_0790	mannose-1-phosphate guanylyltransferase/mannose-6-phosphate isomerase		YP_004151398.1	53	**0.03**
Theam_0797	nucleotide sugar dehydrogenase		YP_004151405.1	50	**0.03**
Theam_0800	DNA polymerase I	*dna2*	YP_004151408.1	94	**0.01**
Theam_0804	putative lipoprotein		YP_004151412.1	16	**0.03**
Theam_0812	DNA helicase		YP_004151420.1	83	**0.02**
Theam_0814	hypothetical protein		YP_004151422.1	31	**0.01**
Theam_0815	Nucleoside-diphosphate kinase		YP_004151423.1	16	**0.20**
Theam_0820	protein-export membrane protein SecD	*secD*	YP_004151428.1	61	**0.00**
Theam_0824	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	*murA*	YP_004151432.1	47	**0.00**
Theam_0825	ATP phosphoribosyltransferase	*hisG*	YP_004151433.1	24	**0.06**
Theam_0826	formyltetrahydrofolate deformylase	*PurU*	YP_004151434.1	33	**0.04**
Theam_0827	glucosamine/fructose-6-phosphate aminotransferase, isomerizing	*glmS*	YP_004151435.1	68	**0.05**
Theam_0828	tRNA modification GTPase TrmE		YP_004151436.1	52	**0.04**
Theam_0830	protein of unknown function DUF583		YP_004151438.1	13	**0.20**
Theam_0836	HAD-superfamily hydrolase, subfamily IIA		YP_004151444.1	29	**0.12**
Theam_0838	GTP-binding protein YchF	*ychF*	YP_004151446.1	41	**0.02**
Theam_0843	response regulator receiver	*cheY*	YP_004151451.1	15	**0.08**
Theam_0844	response regulator receiver	*cheV*	YP_004151452.1	36	**0.03**

Locus	Product name	*Gene Name*	Accession number	MW (kDa)	NSAF %
Theam_0845	chemotaxis sensory transducer	*mcp*	YP_004151453.1	72	**0.09**
Theam_0846	CheW domain protein	*cheW*	YP_004151454.1	19	**0.02**
Theam_0847	ATP-binding region ATPase domain protein - histidine kinase	*cheA*	YP_004151455.1	74	**0.03**
Theam_0848	putative myosin-2 heavy chain, non muscle	*cheZ*	YP_004151456.1	20	**0.07**
Theam_0849	cell division protein FtsZ	*ftsZ*	YP_004151457.1	39	**0.04**
Theam_0852	D-alanine/D-alanine ligase		YP_004151460.1	32	**0.01**
Theam_0854	UDP-N-acetylmuramyl-tripeptide synthetase	*murE*	YP_004151462.1	53	**0.00**
Theam_0855	hypothetical protein		YP_004151463.1	11	**0.04**
Theam_0856	hypothetical protein		YP_004151464.1	10	**0.02**
Theam_0862	outer membrane efflux protein		YP_004151470.1	48	**0.01**
Theam_0873	DNA ligase, NAD-dependent	*dnlj*	YP_004151481.1	82	**0.05**
Theam_0876	molybdenum cofactor biosynthesis protein C	*moa*	YP_004151484.1	33	**0.36**
Theam_0877	hypothetical protein		YP_004151485.1	21	**0.02**
Theam_0878	molybdopterin biosynthesis MoaE protein	*moaE*	YP_004151486.1	13	**0.01**
Theam_0884	signal recognition particle-docking protein FtsY	*ftsY*	YP_004151492.1	54	**0.02**
Theam_0885	anthranilate phosphoribosyltransferase	*trpD*	YP_004151493.1	38	**0.04**
Theam_0887	metalloendopeptidase, glycoprotease family	*gcp*	YP_004151495.1	35	**0.03**
Theam_0889	transglutaminase domain-containing protein		YP_004151497.1	72	**0.00**
Theam_0890	phosphoglucosamine mutase	*glmM*	YP_004151498.1	49	**0.07**
Theam_0893	dihydropteroate synthase	*dhps*	YP_004151501.1	44	**0.01**
Theam_0894	Polynucleotide adenylyltransferase region		YP_004151502.1	98	**0.01**
Theam_0896	general secretion pathway protein D	*gspD*	YP_004151504.1	71	**0.04**
Theam_0903	type II and III secretion system protein	*gspC*	YP_004151511.1	72	**0.02**
Theam_0906	reverse gyrase	*rgy*	YP_004151514.1	137	**0.04**
Theam_0907	iron-containing alcohol dehydrogenase		YP_004151515.1	41	**0.09**
Theam_0909	transcription antitermination factor NusB	*nusB*	YP_004151517.1	16	**0.03**
Theam_0910	6,7-dimethyl-8-ribityllumazine synthase	*ribH*	YP_004151518.1	17	**0.27**
Theam_0911	enoyl-(acyl-carrier-protein) reductase II	*fabK*	YP_004151519.1	34	**0.07**
Theam_0912	3-oxoacyl-(acyl-carrier-protein) synthase III	*fabH*	YP_004151520.1	34	**0.05**
Theam_0913	fatty acid/phospholipid synthesis protein PlsX	*plsX*	YP_004151521.1	36	**0.00**
Theam_0919	methylenetetrahydrofolate reductase F	*metF*	YP_004151527.1	32	**0.04**
Theam_0920	undecaprenyl diphosphate synthase	*uppS*	YP_004151528.1	27	**0.05**
Theam_0922	hydrogenase expression/formation protein HypE	*hypE*	YP_004151530.1	35	**0.11**
Theam_0924	hydrogenase expression/formation protein HypD	*hypD*	YP_004151532.1	39	**0.06**
Theam_0936	hypothetical protein		YP_004151544.1	38	**0.05**
Theam_0938	Orn/DAP/Arg decarboxylase 2		YP_004151546.1	51	**0.02**
Theam_0940	hypothetical protein		YP_004151548.1	34	**0.10**
Theam_0942	phosphoribosylamine/glycine ligase	*purD*	YP_004151550.1	46	**0.08**
Theam_0948	histidine triad (HIT) protein		YP_004151556.1	19	**0.18**
Theam_0953	Domain of unkown function DUF1786 putative pyruvate format-lyase activating enzyme		YP_004151561.1	37	**0.01**
Theam_0954	cell division protein FtsZ	*ftsZ*	YP_004151562.1	39	**0.03**
Theam_0955	Nicotinamidase	*pncA*	YP_004151563.1	21	**0.10**
Theam_0956	hypothetical protein		YP_004151564.1	33	**0.22**
Theam_0962	peptidase M16 domain protein		YP_004151570.1	46	**0.02**
Theam_0963	hypothetical protein		YP_004151571.1	48	**0.00**
Theam_0964	ADP-ribosylation/Crystallin J1		YP_004151572.1	35	**0.03**
Theam_0965	hypothetical protein		YP_004151573.1	126	**0.01**
Theam_0969	fumarate		YP_004151577.1	42	**0.01**
Theam_0970	Radical SAM domain protein		YP_004151578.1	42	**0.01**
Theam_0972	phosphoenolpyruvate synthase water dikinase	*ppsA*	YP_004151580.1	91	**0.08**

Locus	Product name	*Gene Name*	Accession number	MW (kDa)	NSAF %
Theam_0973	D-isomer specific 2-hydroxyacid dehydrogenase NAD-binding		YP_004151581.1	38	**0.04**
Theam_0974	phosphoglucose isomerase	*pgi*	YP_004151582.1	45	**0.03**
Theam_0984	malonyl CoA-acyl carrier protein transacylase	*fabD*	YP_004151592.1	34	**0.21**
Theam_0985	ribonucleoside-diphosphate reductase, adenosylcobalamin-dependent		YP_004151593.1	63	**0.01**
Theam_0990	twitching motility protein	*pilT*	YP_004151598.1	40	**0.03**
Theam_0998	Glutamate synthase (NADPH)	*gltA*	YP_004151606.1	55	**0.17**
Theam_1000	NAD – Nitrite reductase (FAD-dependent pyridine nucleotide-disulphide oxidoreductase)	*nirA*	YP_004151608.1	47	**0.20**
Theam_1001	glutamine amidotransferase class-II		YP_004151609.1	42	**0.12**
Theam_1002	glutamate synthase alpha subunit domain protein	*glt*	YP_004151610.1	29	**0.24**
Theam_1003	tryptophan synthase, beta subunit	*trpB*	YP_004151611.1	44	**0.10**
Theam_1005	tyrosyl-tRNA synthetase	*tyrS*	YP_004151613.1	47	**0.07**
Theam_1008	lysyl-tRNA synthetase	*lysSb*	YP_004151616.1	61	**0.11**
Theam_1011	Mammalian cell entry related domain protein		YP_004151619.1	58	**0.00**
Theam_1013	??fumarate reductase/succinate dehydrogenase flavoprotein domain protein	*fum?*	YP_004151621.1	20	**0.26**
Theam_1017	adenylosuccinate lyase	*purB*	YP_004151625.1	51	**0.25**
Theam_1020	molybdopterin oxidoreductase fdhA hydrogenase family	*fdhA*	YP_004151628.1	76	**0.04**
Theam_1021	ATP-citrate (pro-S-)-lyase subunit alpha	*aclA*	YP_004151629.1	68	**2.03**
Theam_1022	ATP-citrate (pro-S-)-lyase subunit beta	*aclB*	YP_004151630.1	49	**0.64**
Theam_1023	isocitrate dehydrogenase/2-oxoglutarate carboxylase , NADP-dependent	*idh2*	YP_004151631.1	83	**1.24**
Theam_1024	aconitate hydratase	*acnB*	YP_004151632.1	71	**0.59**
Theam_1027	methyl-accepting chemotaxis protein		YP_004151635.1	38	**0.04**
Theam_1028	hypothetical protein		YP_004151636.1	42	**0.12**
Theam_1029	hypothetical protein		YP_004151637.1	75	**0.63**
Theam_1035	phosphopantothenoylcysteine decarboxylase/phosphopantothenate/cysteine ligase	*coaBC_dfp*	YP_004151643.1	42	**0.01**
Theam_1036	RNA methyltransferase, TrmH family, group 3		YP_004151644.1	26	**0.02**
Theam_1041	carbohydrate kinase, YjeF related protein		YP_004151649.1	57	**0.10**
Theam_1042	peptide deformylase		YP_004151650.1	20	**0.07**
Theam_1043	Xylose isomerase domain-containing protein TIM barrel		YP_004151651.1	28	**0.02**
Theam_1045	Phosphoribosyl-AMP cyclohydrolase		YP_004151653.1	13	**0.11**
Theam_1053	tRNA 2-selenouridine synthase		YP_004151661.1	40	**0.01**
Theam_1055	pyruvate fromate-lyase activating enzyme		YP_004151663.1	38	**0.03**
Theam_1056	PEBP family protein		YP_004151664.1	17	**0.01**
Theam_1062	heat shock protein Hsp20		YP_004151670.1	20	**0.04**
Theam_1063	PHP domain protein		YP_004151671.1	65	**0.02**
Theam_1065	hypothetical protein		YP_004151673.1	41	**0.02**
Theam_1067	3-dehydroquinate synthase	*aroB*	YP_004151675.1	37	**0.32**
Theam_1069	ADP-L-glycero-D-manno-heptose-6-epimerase		YP_004151677.1	36	**0.14**
Theam_1070	regulatory protein MerR	*merR*	YP_004151678.1	14	**0.01**
Theam_1071	ATPase AAA-2 domain protein	*clp?*	YP_004151679.1	95	**0.12**
Theam_1075	Methionine adenosyltransferase	*samS*	YP_004151683.1	44	**0.29**
Theam_1076	hypothetical protein		YP_004151684.1	71	**0.01**
Theam_1078	hypothetical protein		YP_004151686.1	30	**0.02**
Theam_1081	preprotein translocase, SecA subunit	*secA*	YP_004151689.1	100	**0.06**
Theam_1083	protease Do		YP_004151691.1	52	**0.05**
Theam_1087	MotA/TolQ/ExbB proton channel	*motA*	YP_004151695.1	27	**0.01**
Theam_1088	D-lactate dehydrogenase (cytochrome)		YP_004151696.1	49	**0.03**
Theam_1090	polysaccharide export protein		YP_004151698.1	107	**0.01**
Theam_1095	branched-chain amino acid aminotransferase	*ilvE_I*	YP_004151703.1	34	**0.32**

Locus	Product name	*Gene Name*	Accession number	MW (kDa)	NSAF %
Theam_1098	general secretion pathway protein G	*gspG*	YP_004151706.1	16	**0.02**
Theam_1100	hypothetical protein		YP_004151708.1	14	**0.27**
Theam_1106	uroporphyrinogen decarboxylase	*hemE*	YP_004151713.1	39	**0.14**
Theam_1107	exodeoxyribonuclease III Xth	*xth*	YP_004151714.1	30	**0.01**
Theam_1109	protein-L-isoaspartate(D-aspartate) O-methyltransferase		YP_004151716.1	30	**0.03**
Theam_1111	protein of unknown function DUF28		YP_004151718.1	27	**0.01**
Theam_1113	hydrolase, TatD family		YP_004151720.1	53	**0.02**
Theam_1115	arginyl-tRNA synthetase	*argS*	YP_004151722.1	62	**0.09**
Theam_1116	[NiFe] hydrogenase maturation protein HypF	*hypF*	YP_004151723.1	83	**0.01**
Theam_1121	hydrogenase (NiFe) small subunit HydA	*hynA*	YP_004151728.1	39	**0.03**
Theam_1122	nickel-dependent hydrogenase large subunit	*hynB*	YP_004151729.1	65	**0.27**
Theam_1123	Ni/Fe-hydrogenase, b-type cytochrome subunit	*hynC*	YP_004151730.1	26	**0.01**
Theam_1124	hydrogenase 2 maturation protease	*hupD*	YP_004151731.1	20	**0.02**
Theam_1128	hydrogenase accessory protein HypB	*hypB*	YP_004151735.1	28	**0.24**
Theam_1130	outer membrane efflux protein		YP_004151737.1	51	**0.00**
Theam_1133	flavodoxin/nitric oxide synthase not related to nitrogen metabolism		YP_004151740.1	45	**0.09**
Theam_1138	glutamyl-tRNA(Gln) amidotransferase, B subunit	*gatB*	YP_004151743.1	55	**0.20**
Theam_1141	DNA gyrase, A subunit	*gyrA*	YP_004151746.1	91	**0.08**
Theam_1142	seryl-tRNA synthetase	*serS*	YP_004151747.1	49	**0.09**
Theam_1150	transcriptional regulator domain-containing protein		YP_004151755.1	25	**0.01**
Theam_1154	3-deoxy-D-manno-octulosonate 8-phosphate phosphatase, YrbI family		YP_004151759.1	17	**0.02**
Theam_1156	Cytochrome-c peroxidase	*ccpA*	YP_004151761.1	45	**0.02**
Theam_1157	dethiobiotin synthase	*bioD*	YP_004151762.1	22	**0.07**
Theam_1162	flagellin domain protein - flagellin structural protein	*FliC*	YP_004151767.1	32	**1.66**
Theam_1165	glycosyl transferase family 2		YP_004151770.1	47	**0.06**
Theam_1166	riboflavin biosynthesis protein RibD	*ribD*	YP_004151771.1	40	**0.00**
Theam_1168	tryptophanyl-tRNA synthetase	*trpS*	YP_004151773.1	41	**0.10**
Theam_1173	phosphoribosylaminoimidazole-succinocarboxamide synthase	*purC*	YP_004151778.1	28	**0.34**
Theam_1174	amidophosphoribosyltransferase	*purF*	YP_004151779.1	51	**0.04**
Theam_1177	threonyl-tRNA synthetase	*thrS*	YP_004151782.1	76	**0.05**
Theam_1178	translation initiation factor IF-3	*infC*	YP_004151783.1	17	**0.03**
Theam_1180	ribosomal protein L20	*rplT*	YP_004151785.1	14	**0.13**
Theam_1181	hypothetical protein		YP_004151786.1	41	**0.13**
Theam_1183	6-pyruvoyl tetrahydropterin synthase and hypothetical protein		YP_004151788.1	24	**0.18**
Theam_1184	methyltransferase	*metR*	YP_004151789.1	22	**0.08**
Theam_1188	DNA gyrase, B subunit	*gyrB*	YP_004151793.1	91	**0.03**
Theam_1189	cytidylate kinase	*cmk*	YP_004151794.1	24	**0.06**
Theam_1190	nicotinate phosphoribosyltransferase		YP_004151795.1	49	**0.04**
Theam_1191	DNA-(apurinic or apyrimidinic site) lyase		YP_004151796.1	25	**0.01**
Theam_1192	2-dehydro-3-deoxyphosphooctonate aldolase	*kdo*	YP_004151797.1	30	**0.07**
Theam_1195	deoxyribose-phosphate aldolase	*deoC*	YP_004151800.1	24	**0.03**
Theam_1197	GTP-binding protein Era	*era*	YP_004151802.1	35	**0.03**
Theam_1201	hypothetical protein		YP_004151806.1	9	**0.15**
Theam_1203	phenylalanyl-tRNA synthetase, alpha subunit	*pheS*	YP_004151808.1	39	**0.10**
Theam_1204	phenylalanyl-tRNA synthetase, beta subunit	*pheTb*	YP_004151809.1	89	**0.18**
Theam_1206	5-formyltetrahydrofolate cyclo-ligase	*folA*	YP_004151811.1	21	**0.02**
Theam_1207	YmdA/YtgF protein	*hdig*	YP_004151812.1	63	**0.02**
Theam_1208	selenocysteine-specific translation elongation factor	*selB*	YP_004151813.1	71	**0.00**
Theam_1211	phosphoribosylglycinamide formyltransferase 2	*purT*	YP_004151816.1	44	**0.12**

Locus	Product name	*Gene Name*	Accession number	MW (kDa)	NSAF %
Theam_1213	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase		YP_004151818.1	28	**0.03**
Theam_1215	pyrroline-5-carboxylate reductase	*proC*	YP_004151820.1	29	**0.10**
Theam_1217	Phosphomethylpyrimidine kinase		YP_004151822.1	26	**0.01**
Theam_1218	sugar-phosphate isomerase, RpiB/LacA/LacB family	*rpiB*	YP_004151823.1	17	**0.07**
Theam_1219	Glycine hydroxymethyltransferase		YP_004151824.1	46	**0.20**
Theam_1221	chaperonin GroEL		YP_004151826.1	59	**1.37**
Theam_1222	Chaperonin Cpn10		YP_004151827.1	11	**1.03**
Theam_1223	transglutaminase domain-containing protein		YP_004151828.1	35	**0.03**
Theam_1225	phosphate-selective porin O and P		YP_004151830.1	44	**0.13**
Theam_1226	Hydrogenase Mo catalitic subunit - molydopterin dinucleotide-binding region	*hycB2*	YP_004151831.1	131	**0.12**
Theam_1227	hypothetical protein		YP_004151832.1	27	**0.01**
Theam_1228	CBS domain containing protein		YP_004151833.1	15	**0.10**
Theam_1234	ATPase-like, ParA/MinD		YP_004151839.1	32	**0.18**
Theam_1235	inositol monophosphatase		YP_004151840.1	28	**0.13**
Theam_1238	histidine kinase		YP_004151843.1	34	**0.01**
Theam_1240	thioesterase superfamily protein		YP_004151845.1	14	**0.05**
Theam_1241	hypothetical protein		YP_004151846.1	23	**0.01**
Theam_1242	peptide chain release factor 1	*prfA*	YP_004151847.1	41	**0.04**
Theam_1244	Radical SAM domain protein		YP_004151849.1	43	**0.03**
Theam_1246	2,3,4,5-tetrahydropyridine-2,6-carboxylate N-succinyltransferase		YP_004151851.1	30	**0.15**
Theam_1247	methionyl-tRNA synthetase	*metG*	YP_004151852.1	61	**0.09**
Theam_1249	S-Adenosyl homocisteine hydrolase (adenosylhomocysteinase)	*sahH*	YP_004151854.1	47	**0.16**
Theam_1250	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	*lpxD*	YP_004151855.1	36	**0.03**
Theam_1251	outer membrane chaperone Skp (OmpH)		YP_004151856.1	20	**0.01**
Theam_1252	outer membrane protein assembly complex, YaeT protein		YP_004151857.1	87	**0.02**
Theam_1254	glyceraldehyde 3-phosphate dehydrogenase, type I	*gapdh*	YP_004151859.1	36	**0.68**
Theam_1256	nicotinate-nucleotide pyrophosphorylase	*nadC*	YP_004151861.1	32	**0.04**
Theam_1259	dephospho-CoA kinase		YP_004151864.1	21	**0.02**
Theam_1260	metallophosphoesterase		YP_004151865.1	30	**0.04**
Theam_1261	Methylenetetrahydrofolate dehydrogenase (NADP(+))	*folD*	YP_004151866.1	31	**0.22**
Theam_1265	AMP-dependent synthetase and ligase		YP_004151869.1	61	**0.01**
Theam_1266	diaminopimelate epimerase	*DapF*	YP_004151870.1	30	**0.01**
Theam_1267	phage SPO1 DNA polymerase-related protein	*spo1*	YP_004151871.1	27	**0.07**
Theam_1270	CoB--CoM heterodisulfide reductase / Fumarate reductase ascciated protein A?	*hdrA*	YP_004151874.1	32	**0.09**
Theam_1271	fumarate reductase iron-sulfur protein	*frdB1*	YP_004151875.1	43	**0.11**
Theam_1272	fumarate reductase, flavoprotein subunit	*frdB*	YP_004151876.1	62	**0.18**
Theam_1273	fumarate reductase, flavoprotein subunit	*frdA1*	YP_004151877.1	63	**0.07**
Theam_1274	fumarate reductase iron-sulfur protein	*frdA*	YP_004151878.1	36	**0.03**
Theam_1275	CoB--CoM heterodisulfide reductase / Fumarate reductase ascciated protein B?	*hdrB*	YP_004151879.1	32	**0.05**
Theam_1276	outer membrane efflux protein		YP_004151880.1	50	**0.02**
Theam_1279	pyruvate kinase	*pvtK*	YP_004151883.1	51	**0.01**
Theam_1282	phosphoglucomutase/phosphomannomutase alpha/beta/alpha domain I		YP_004151886.1	52	**0.20**
Theam_1283	Circadian clock protein KaiC central region		YP_004151887.1	33	**0.11**
Theam_1284	P450 cytochrome, putative		YP_004151888.1	16	**0.07**
Theam_1293	Silent information regulator protein Sir2		YP_004151897.1	30	**0.02**
Theam_1296	putative metal-dependent hydrolase		YP_004151900.1	26	**0.06**
Theam_1297	signal recognition particle protein	*ffh*	YP_004151901.1	50	**0.01**

Locus	Product name	*Gene Name*	Accession number	MW (kDa)	NSAF %
Theam_1298	ribosomal protein S16	*S16*	YP_004151902.1	9	**0.10**
Theam_1299	hypothetical protein		YP_004151903.1	9	**0.20**
Theam_1300	hypothetical protein		YP_004151904.1	22	**0.11**
Theam_1301	transaldolase	*talC*	YP_004151905.1	24	**0.50**
Theam_1303	hypothetical protein		YP_004151907.1	24	**0.06**
Theam_1307	protein of unknown function DUF77		YP_004151911.1	12	**0.02**
Theam_1309	pantoate/beta-alanine ligase		YP_004151913.1	32	**0.02**
Theam_1311	hypothetical protein		YP_004151915.1	13	**0.05**
Theam_1312	glutamate 5-kinase	*proB*	YP_004151916.1	38	**0.01**
Theam_1313	GTP-binding protein Obg/CgtA		YP_004151917.1	38	**0.03**
Theam_1317	hypothetical protein		YP_004151921.1	13	**0.03**
Theam_1320	Ppx/GppA phosphatase		YP_004151924.1	33	**0.02**
Theam_1321	Pyrococcus NSR homolog / FAD-dependent pyridine nucleotide-disulphide oxidoreductase CoB-CoM related	*nsr*	YP_004151925.1	47	**0.11**
Theam_1323	fructose-1,6-bisphosphatase I	*fbp*	YP_004151927.1	34	**0.20**
Theam_1324	radical SAM enzyme, Cfr family		YP_004151928.1	39	**0.02**
Theam_1327	histidinol dehydrogenase	*hisD*	YP_004151931.1	49	**0.07**
Theam_1328	aspartate carbamoyltransferase	*atc*	YP_004151932.1	34	**0.18**
Theam_1330	dihydroorotase, multifunctional complex type	*pyrC*	YP_004151934.1	46	**0.05**
Theam_1331	periplasmic serine protease		YP_004151935.1	33	**0.01**
Theam_1334	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	*ispD*	YP_004151938.1	26	**0.08**
Theam_1337	carbon-monoxide dehydrogenase, catalytic subunit	*codh*	YP_004151941.1	70	**0.41**
Theam_1338	Dinitrogenase iron-molybdenum cofactor biosynthesis protein	*nif*	YP_004151942.1	14	**0.09**
Theam_1342	Phosphoglycerate mutase	*pgm*	YP_004151946.1	24	**0.09**
Theam_1343	glutamate synthase (NADPH), homotetrameric	*gltA*	YP_004151947.1	53	**0.56**
Theam_1344	Protein of unknown function DUF2168		YP_004151948.1	21	**0.06**
Theam_1351	hypothetical protein		YP_004151955.1	20	**0.01**
Theam_1352	fimbrial protein pilin		YP_004151956.1	20	**0.33**
Theam_1354	RNA-metabolising metallo-beta-lactamase		YP_004151958.1	52	**0.01**
Theam_1356	leucyl-tRNA synthetase	*leuS*	YP_004151960.1	105	**0.19**
Theam_1357	hypothetical protein		YP_004151961.1	20	**0.06**
Theam_1359	selenium metabolism protein YedF		YP_004151963.1	22	**0.02**
Theam_1363	Tetratricopeptide TPR_1 repeat-containing protein		YP_004151967.1	30	**0.01**
Theam_1364	GTP cyclohydrolase I	*folE*	YP_004151968.1	21	**0.49**
Theam_1366	hypoxanthine phosphoribosyltransferase	*hgprt*	YP_004151970.1	19	**0.02**
Theam_1368	argininosuccinate synthase	*argG*	YP_004151972.1	45	**0.36**
Theam_1369	trigger factor	*tig*	YP_004151973.1	49	**0.13**
Theam_1372	phosphoribosylformylglycinamidine synthase II	*fgams*	YP_004151976.1	82	**0.24**
Theam_1373	Domain of unknown function DUF1931		YP_004151977.1	17	**0.37**
Theam_1375	Protein of unknown function DUF2148		YP_004151979.1	19	**0.13**
Theam_1378	2-oxoglutarate ferredoxin oxidoreductase subunit alpha	*oorA1*	YP_004151982.1	60	**0.03**
Theam_1379	2-oxoglutarate ferredoxin oxidoreductase subunit beta	*oorB1*	YP_004151983.1	30	**0.02**
Theam_1385	flagellar hook-associated protein FlgK	*flgH*	YP_004151989.1	49	**0.01**
Theam_1387	type IV pilus assembly PilZ		YP_004151991.1	26	**0.04**
Theam_1390	MglA protein		YP_004151994.1	22	**0.15**
Theam_1391	Roadblock/LC7 family protein		YP_004151995.1	18	**0.08**
Theam_1392	pyruvate synthetase / thiamine pyrophosphate TPP-binding domain-containing protein	*porB*	YP_004151996.1	36	**0.43**
Theam_1393	pyruvate synthetase flavodoxin/ferredoxin oxidoreductase domain protein	*porA*	YP_004151997.1	45	**0.46**
Theam_1394	pyruvate synthetase ferredoxin/flavodoxin oxidoreductase, delta subunit	*porD*	YP_004151998.1	11	**0.16**

Locus	Product name	*Gene Name*	Accession number	MW (kDa)	NSAF %
Theam_1395	pyruvate synthetase /ketoisovalerate oxidoreductase, gamma subunit	*porG*	YP_004151999.1	21	**0.51**
Theam_1400	SurA domain		YP_004152004.1	49	**0.00**
Theam_1401	YicC-like domain-containing protein		YP_004152005.1	33	**0.17**
Theam_1402	guanylate kinase		YP_004152006.1	26	**0.03**
Theam_1403	DNA-directed RNA polymerase, omega subunit	*rpoZ*	YP_004152007.1	8	**0.10**
Theam_1404	Dihydroorotate dehydrogenase, electron transfer subunit, iron-sulphur cluster binding domain		YP_004152008.1	28	**0.06**
Theam_1408	succinyl-CoA synthetase, beta subunit	*sucD*	YP_004152012.1	42	**0.58**
Theam_1409	succinyl-CoA synthetase, alpha subunit	*sucC*	YP_004152013.1	32	**0.80**
Theam_1411	2-oxoglutarate synthetase ferredoxin subunit alpha	*oorA*	YP_004152015.1	42	**0.64**
Theam_1412	2-oxoglutarate synthetase ferredoxin subunit beta	*oorB*	YP_004152016.1	32	**0.48**
Theam_1413	2-oxoglutarate synthetase ferredoxin gamma subunit	*oorC*	YP_004152017.1	22	**0.47**
Theam_1418	adenosylmethionine-8-amino-7-oxononanoate aminotransferase	*bioA*	YP_004152022.1	50	**0.04**
Theam_1419	hypothetical protein		YP_004152023.1	18	**0.22**
Theam_1420	UspA domain-containing protein		YP_004152024.1	17	**0.12**
Theam_1421	protein of unknown function DUF814		YP_004152025.1	37	**0.01**
Theam_1422	ribosomal subunit interface protein	*yfiA*	YP_004152026.1	21	**0.11**
Theam_1423	protein of unknown function DUF820		YP_004152027.1	21	**0.04**
Theam_1424	GTP-binding protein LepA	*lepA*	YP_004152028.1	67	**0.01**
Theam_1428	hypothetical protein		YP_004152031.1	151	**0.03**
Theam_1429	helicase domain protein		YP_004152032.1	123	**0.01**
Theam_1430	Inorganic diphosphatase	*ppa*	YP_004152033.1	20	**0.22**
Theam_1435	PHP domain protein		YP_004152038.1	24	**0.03**
Theam_1446	fagellar hook-basal body protein- hook structural protein	*flgE*	YP_004152049.1	57	**0.07**
Theam_1470	Citrate transporter		YP_004152072.1	49	**0.00**
Theam_1472	ribosome biogenesis GTP-binding protein YsxC		YP_004152074.1	22	**0.02**
Theam_1473	Redoxin domain protein		YP_004152075.1	22	**0.10**
Theam_1474	thioredoxin reductase	*trxr*	YP_004152076.1	33	**0.13**
Theam_1475	thioredoxin	*trx*	YP_004152077.1	12	**0.39**
Theam_1476	regulatory protein ArsR		YP_004152078.1	11	**0.04**
Theam_1477	protein of unknown function DUF52		YP_004152079.1	29	**0.11**
Theam_1479	Myo-inositol-1-phosphate synthase		YP_004152081.1	40	**0.20**
Theam_1480	hypothetical protein		YP_004152082.1	15	**0.01**
Theam_1481	tol-pal system protein YbgF		YP_004152083.1	26	**0.05**
Theam_1482	2-nitropropane dioxygenase NPD	*npd*	YP_004152084.1	39	**0.01**
Theam_1484	peptidase M16 domain protein		YP_004152086.1	45	**0.02**
Theam_1486	phosphoribosylformylglycinamidine cyclo-ligase	*purM*	YP_004152088.1	37	**0.13**
Theam_1487	metal dependent phophohydrolase	*hdig*	YP_004152089.1	37	**0.00**
Theam_1493	NADH dehydrogenase subunit I	*nuoI*	YP_004152095.1	22	**0.03**
Theam_1495	NADH dehydrogenase (quinone)	*nuoD*	YP_004152097.1	43	**0.03**
Theam_1497	NADH-quinone oxidoreductase, B subunit	*nuoB*	YP_004152099.1	18	**0.04**
Theam_1499	phosphate ABC transporter, ATPase subunit	*pstB*	YP_004152101.1	29	**0.04**
Theam_1502	DNA polymerase III, alpha subunit	*polc*	YP_004152104.1	131	**0.00**
Theam_1503	acetyl-CoA carboxylase, carboxyl transferase, alpha subunit	*accA*	YP_004152105.1	36	**0.02**
Theam_1505	DEAD/DEAH box helicase domain protein		YP_004152107.1	47	**0.05**
Theam_1510	D-3-phosphoglycerate dehydrogenase	*pgdh*	YP_004152112.1	59	**0.40**
Theam_1512	Holliday junction DNA helicase RuvA	*ruvA*	YP_004152114.1	21	**0.02**
Theam_1517	translation initiation factor, aIF-2BI family	*alF*	YP_004152119.1	39	**0.12**
Theam_1518	UspA domain-containing protein		YP_004152120.1	18	**0.19**
Theam_1519	MazG family protein	*mazG*	YP_004152121.1	31	**0.03**

Locus	Product name	*Gene Name*	Accession number	MW (kDa)	NSAF %
Theam_1520	cysteinyl-tRNA synthetase	*cysS*	YP_004152122.1	55	**0.09**
Theam_1526	Nickel transport complex, NikM subunit, transmembrane		YP_004152128.1	29	**0.12**
Theam_1527	Carbohydrate-selective porin OprB		YP_004152129.1	47	**0.07**
Theam_1539	pyruvate carboxylase, biotin carboxylase	*pycA*	YP_004152140.1	53	**0.09**
Theam_1540	pyruvate carboxylase alpha subunit	*pycB*	YP_004152141.1	69	**0.28**
Theam_1544	chorismate synthase	*aroC*	YP_004152145.1	42	**0.04**
Theam_1545	ribosomal protein L19	*rplS*	YP_004152146.1	14	**0.89**
Theam_1548	translation initiation factor IF-2		YP_004152148.1	99	**0.07**
Theam_1550	1-phosphofructokinase		YP_004152150.1	34	**0.04**
Theam_1552	ribonucleotide reductase		YP_004152152.1	41	**0.16**
Theam_1554	ribonucleoside-diphosphate reductase, alpha subunit		YP_004152154.1	87	**0.17**
Theam_1555	Redoxin domain protein		YP_004152155.1	18	**0.11**
Theam_1559	beta-lactamase domain protein		YP_004152159.1	29	**0.01**
Theam_1560	chemotaxis sensory transducer		YP_004152160.1	28	**0.01**
Theam_1561	5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase	*metE*	YP_004152161.1	82	**0.69**
Theam_1563	pseudogene			79	**0.01**
Theam_1564	ErfK/YbiS/YcfS/YnhG family protein		YP_004152163.1	41	**0.01**
Theam_1566	thioesterase superfamily protein		YP_004152165.1	21	**0.67**
Theam_1567	carbamoyl-phosphate synthase, large subunit	*cpsL*	YP_004152166.1	119	**0.33**
Theam_1568	Adenylate kinase		YP_004152167.1	19	**0.01**
Theam_1569	sun protein	*rsmB*	YP_004152168.1	50	**0.02**
Theam_1570	Metal-dependent hydrolase HDOD		YP_004152169.1	24	**0.04**
Theam_1571	Metal-dependent hydrolase HDOD		YP_004152170.1	25	**0.06**
Theam_1576	Carbonate dehydratase		YP_004152175.1	28	**0.08**
Theam_1578	DNA-directed RNA polymerase, beta' subunit	*rpoB1*	YP_004152177.1	167	**0.34**
Theam_1579	DNA-directed RNA polymerase, beta subunit	*rpoB*	YP_004152178.1	163	**0.38**
Theam_1580	ribosomal protein L7/L12	*L12*	YP_004152179.1	13	**0.38**
Theam_1581	ribosomal protein L10		YP_004152180.1	20	**0.47**
Theam_1582	ribosomal protein L1	*rplA*	YP_004152181.1	26	**0.37**
Theam_1583	ribosomal protein L11	*L11*	YP_004152182.1	15	**0.11**
Theam_1584	transcription termination/antitermination factor NusG	*nusG*	YP_004152183.1	28	**0.04**
Theam_1587	translation elongation factor Tu		YP_004152186.1	44	**2.19**
Theam_1588	hypothetical protein		YP_004152187.1	52	**0.83**
Theam_1590	16S rRNA processing protein RimM		YP_004152188.1	22	**0.01**
Theam_1591	tRNA (guanine-N1)-methyltransferase	*trmD*	YP_004152189.1	29	**0.02**
Theam_1596	chorismate mutase		YP_004152194.1	40	**0.02**
Theam_1598	Imidazoleglycerol-phosphate dehydratase		YP_004152196.1	22	**0.03**
Theam_1600	glutamate-1-semialdehyde-2,1-aminomutase	*hemL*	YP_004152198.1	47	**0.21**
Theam_1605	ATP synthase F0, A subunit	*atpa*	YP_004152203.1	26	**0.01**
Theam_1611	OmpA/MotB domain protein		YP_004152209.1	26	**0.01**
Theam_1612	gamma-glutamyl phosphate reductase	*proA*	YP_004152210.1	47	**0.22**
Theam_1626	hypothetical protein		YP_004152223.1	34	**0.02**
Theam_1627	pyruvate:water dikinase	*pvtk*	YP_004152224.1	100	**0.85**
Theam_1628	glycyl-tRNA synthetase, beta subunit	*glyS*	YP_004152225.1	77	**0.11**
Theam_1629	glycyl-tRNA synthetase, alpha subunit	*glyQ*	YP_004152226.1	33	**0.11**
Theam_1632	exsB protein	*exsB*	YP_004152229.1	26	**0.03**
Theam_1645	hypothetical protein		YP_004152242.1	17	**0.06**
Theam_1647	glutamyl-tRNA(Gln) amidotransferase, A subunit	*gatA*	YP_004152244.1	53	**0.41**
Theam_1648	glutamyl-tRNA(Gln) amidotransferase, C subunit	*gatC*	YP_004152245.1	11	**0.02**
Theam_1656	H+transporting two-sector ATPase B/B' subunit	*atpB*'	YP_004152253.1	17	**0.14**

Locus	Product name	*Gene Name*	Accession number	MW (kDa)	NSAF %
Theam_1657	ATP synthase F0, B subunit	*atpB*	YP_004152254.1	18	**0.08**
Theam_1658	ATP synthase F1, delta subunit	*atpD*	YP_004152255.1	20	**0.05**
Theam_1659	ATP synthase F1, alpha subunit	*atpA*	YP_004152256.1	55	**0.16**
Theam_1660	ATP synthase F1, gamma subunit	*atpC*	YP_004152257.1	32	**0.10**
Theam_1661	ATP synthase F1, beta subunit	*atpD*	YP_004152258.1	54	**0.19**
Theam_1662	ATP synthase F1, epsilon subunit	*atpE*	YP_004152259.1	16	**0.12**
Theam_1667	protein of unknown function DUF62		YP_004152264.1	29	**0.01**
Theam_1668	glucose inhibited division protein A	*gidA*	YP_004152265.1	72	**0.06**
Theam_1669	methyltransferase GidB	*gidB*	YP_004152266.1	24	**0.02**
Theam_1672	homoserine kinase	*thrB*	YP_004152269.1	33	**0.16**
Theam_1676	valyl-tRNA synthetase	*valS*	YP_004152273.1	102	**0.19**
Theam_1677	CMP/dCMP deaminase zinc-binding		YP_004152274.1	17	**0.01**
Theam_1678	recA protein	*recA*	YP_004152275.1	38	**0.02**
Theam_1679	twitching motility protein	*pilT*	YP_004152276.1	41	**0.03**
Theam_1682	alanyl-tRNA synthetase	*alaS*	YP_004152279.1	99	**0.18**
Theam_1684	hypothetical protein		YP_004152281.1	150	**0.02**
Theam_1685	hypothetical protein		YP_004152282.1	50	**0.00**
Theam_1686	PfkB domain protein		YP_004152283.1	36	**0.02**
Theam_1688	transcription termination factor NusA	*NusA*	YP_004152285.1	43	**0.08**
Theam_1689	protein of unknown function DUF150		YP_004152286.1	18	**0.04**
Theam_1691	RNA binding S1 domain protein		YP_004152288.1	39	**0.01**
Theam_1696	S-adenosyl-methyltransferase MraW	*mraW*	YP_004152293.1	34	**0.02**
Theam_1697	ornithine carbamoyltransferase	*otc*	YP_004152294.1	35	**0.14**
Theam_1712	stationary-phase survival protein SurE	*surE*	YP_004152309.1	28	**0.03**
Theam_1725	Indole-3-glycerol-phosphate synthase		YP_004152322.1	29	**0.08**
Theam_1726	UDP-N-acetylglucosamine pyrophosphorylase	*glmU*	YP_004152323.1	51	**0.10**
Theam_1727	NUDIX hydrolase		YP_004152324.1	17	**0.07**
Theam_1728	hypothetical protein		YP_004152325.1	32	**0.01**
Theam_1729	S-adenosylmethionine/tRNA-ribosyltransferase-isomerase	*queA*	YP_004152326.1	39	**0.03**
Theam_1735	hypothetical protein		ADU97691.1	43	**0.03**
Theam_1738	hypothetical protein		ADU97694.1	19	**0.06**
Theam_1739	hypothetical protein		ADU97695.1	23	**0.03**
Theam_1742	hypothetical protein		ADU97698.1	35	**0.01**
Theam_1744	hypothetical protein		ADU97700.1	22	**0.01**
Theam_1768	hypothetical protein		ADU97724.1	30	**0.01**
Theam_1771	ATPase associated with various cellular activities AAA_5		ADU97727.1	31	**0.04**
Theam_1774	trichohyalin		ADU97730.1	21	**0.05**
Theam_1778	hypothetical protein		ADU97734.1	12	**0.12**
Theam_1790	hypothetical protein		ADU97746.1	63	**0.01**
Theam_1795	peptidase C11 clostripain		ADU97751.1	76	**0.01**
Theam_1797	hypothetical protein		ADU97753.1	45	**0.06**
Theam_1799	type II secretion system protein E	*gspE*	ADU97755.1	46	**0.22**
Theam_1801	hypothetical protein		ADU97757.1	19	**0.05**
Theam_1803	hypothetical protein		ADU97759.1	19	**0.07**
Theam_1804	type II secretion system protein E	*gspE*	ADU97760.1	60	**0.01**
Theam_1805	hypothetical protein		ADU97761.1	28	**0.01**
Theam_1818	hypothetical protein		ADU97774.1	18	**0.06**

Accession numbers refer to the *T. ammonificans* proteins in NCBI. NSAF% are normalized spectral abundance factors, i.e. relative abundances for each protein in % of all proteins in the sample. NSAF% are average values (n=3).

Figure 5-source data 1. Accession number of the ATP citrate lyase sequences used for the reconstruction of the phylogenetic history of the enzyme presented in Fig. 2 in the main text.

	Accession numbers	
Organism	**AclB**	**AclA**
Balnearium lithotrophicum 17S	114054980	114054981
Caminibacter mediatlanticus TB2	494739975	494739977
Candidatus Nitrospira defluvii	302036204	302036205
Chlorobaculum parvum NCIB 8327	193212663	193212664
Chlorobium chlorochromatii CaD3	78188767	78188768
Chlorobium limicola DSM 245	189346749	189346748
Chlorobium phaeobacteroides DSM 266	119357118	119357117
Chlorobium phaeovibrioides DSM 265	145219675	145219674
Chlorobium tepidum TLS	21673915	21673914
Chloroherpeton thalassium ATCC 35110	193216250	193216249
Desulfurobacterium sp. TC5-1	551219091	551219090
Desulfurobacterium thermolithotrophum DSM 11699	325295317	325295316
Lebetimonas sp. JS170	640061845	640076131
Nautilia profundicola AmH	224372752	224372753
Nitratifractor salsuginis DSM 16511	319956330	319956331
Nitratiruptor sp. SB155-2	152990401	152990402
Pelodictyon luteolum DSM 273	78186923	78186924
Persephonella marina EX-H1	225851245	225851244
Persephonella sp. KM09-Lau-8	657725816	657725815
Prosthecochloris aestuarii DSM 271	194333911	194333912
Sulfuricurvum kujiense DSM 16994	313681794	313681795
Sulfurihydrogenibium azorense Az-Fu1	225848355	225848356
Sulfurihydrogenibium sp. YO3AOP1	188996978	188996977
Sulfurihydrogenibium subterraneum DSM 15120	114055039	655805784
Sulfurimonas autotrophica DSM 16294	307721497	307721496
Sulfurimonas denitrificans DSM 1251	78776769	78776770
Sulfurimonas gotlandica GD1	495611782	495610702
Sulfurovum sp. AR	495520539	495520532
Sulfurovum sp. NBC37-1	152992137	152992138
Thermovibrio ammonificans HB-1	319789997	319789996
Uncultured *Sulfuricurvum* sp. RIFRC-1	476409849	476409850
Aquifex aeolicus VF5[1]	15606514	15606916 + 15605724
Hydrogenobacter thermophilus TK-6[1]	384129780	384128288 + 384128393

1- Citryl-CoA synthase subunit A (CcsA) was used in place of AclB and citryl-CoA synthase subunit B (CcsB) was manually concatenated to citryl-CoA lyase (Ccl) in place of AclA to reconstruct a hypothetical ancestral ATP citrate lyase enzyme.