

## Higher-level classification of the Archaea: evolution of methanogenesis and methanogens

ÉRIC BAPTESTE<sup>1,2</sup>, CÉLINE BROCHIER<sup>3</sup> AND YAN BOUCHER<sup>4</sup>

<sup>1</sup> Department of Biochemistry and Molecular Biology, Dalhousie University, Sir Charles Tupper Building, Halifax, NS, B3H 4H7, Canada

<sup>2</sup> Corresponding author (eric.bapteste@dal.ca)

<sup>3</sup> EA EGEE (Évolution, Génome, Environnement), Université Aix-Marseille I, Centre Saint-Charles, 13331 Marseille Cedex 3, France

<sup>4</sup> Department of Biological Sciences, Macquarie University, North Ryde, NSW 2109, Australia

Received January 20, 2005; accepted March 29, 2005; published online April 13, 2005

**Summary** We used a phylogenetic approach to analyze the evolution of methanogenesis and methanogens. We show that 23 vertically transmitted ribosomal proteins do not support the monophyly of methanogens, and propose instead that there are two distantly related groups of extant archaea that produce methane, which we have named Class I and Class II. Based on this finding, we subsequently investigated the uniqueness of the origin of methanogenesis by studying both the enzymes of methanogenesis and the proteins that synthesize its specific co-enzymes. We conclude that hydrogenotrophic methanogenesis appeared only once during evolution. Genes involved in the seven central steps of the methanogenic reduction of carbon dioxide (CO<sub>2</sub>) are ubiquitous in methanogens and share a common history. This suggests that, although extant methanogens produce methane from various substrates (CO<sub>2</sub>, formate, acetate, methylated C-1 compounds), these archaea have a core of conserved enzymes that have undergone little evolutionary change. Furthermore, this core of methanogenesis enzymes seems to originate (as a whole) from the last ancestor of all methanogens and does not appear to have been horizontally transmitted to other organisms or between members of Class I and Class II. The observation of a unique and ancestral form of methanogenesis suggests that it was preserved in two independent lineages, with some instances of specialization or added metabolic flexibility. It was likely lost in the Halobacteriales, Thermoplasmatales and Archaeoglobales. Given that fossil evidence for methanogenesis dates back 2.8 billion years, a unique origin of this process makes the methanogenic archaea a very ancient taxon.

*Keywords:* phylogeny, taxonomy.

### Introduction

Methane of biological origin can be found in a wide variety of anaerobic environments, from peat bogs to the digestive tracts of animals and deep-sea hydrothermal vents (McDonald et al. 1999, Takai and Horikoshi 1999, Florin et al. 2000). In all these locations, large quantities of methane originate from

only one type of biological methane producer, archaeal methanogens. There are five phylogenetically divergent orders of the domain Archaea (phylum euryarchaeota) that fall under the appellation “methanogens” (Garrity 2001): Methanobacteriales, Methanopyrales, Methanococcales, Methanomicrobiales and Methanosarcinales. All of these orders contain a wide diversity of taxa that vary greatly in their morphological and physiological characteristics. However, they all have in common an anaerobic lifestyle and the ability to produce methane metabolically.

Soon after it was suggested that the Archaea are a distinct taxonomic group, microbiologists assumed that the domain would be divided along phenotypic lines and that the methanogenic archaea would be monophyletic. Woese and coworkers, using the 16S rRNA gene (Woese and Olsen 1986), demonstrated that this was not the case; the Methanomicrobiales (which at the time included both the current Methanomicrobiales and the Methanosarcinales) were more closely related to extremely halophilic archaea (Halobacteriales) than to other methanogens. Moreover, shortly after the discovery of *Methanopyrus kandleri*, the sequencing of its 16S rRNA gene suggested that it was unrelated to any other methanogens since it emerged at the base of the euryotes, leading the authors to propose that the ancestor of euryotes could have been a methanogen (Burggraf et al. 1991). Recent phylogenies of the archaeal domain based on concatenated ribosomal proteins and on concatenated proteins involved in transcription confirmed the absence of monophyly of methanogens (Matte-Tailliez et al. 2002, Brochier et al. 2004). Such analyses also strongly suggested a close phylogenetic relationship between Methanococcales and Methanobacteriales, but cannot resolve the affiliation of the Methanopyrales (Brochier et al. 2004). Genome trees based on shared gene pairs reconstructed by Slesarev et al. (2002) display a strong monophyletic clustering of *Methanopyrus*, *Methanococcus* and *Methanothermobacter*. The clade that includes these three methanogens is also supported by the RNA polymerase subunit B (rpoB) tree, reconstructed by taking into account the variation of evolutionary rate among sites, and consistent with a shared split event of rpoB

into rpoB' and rpoB" (Brochier et al. 2004). However, the lack of complete genome sequence data for a member of the Methanomicrobiales has prevented the inclusion of this order in concatenated phylogenetic analyses or genome tree reconstructions. The 16S rRNA gene sequences, available from representatives of the Methanomicrobiales, however, support the close relationship of this phylum to the Methanosarcinales (Castro et al. 2004).

Methane can be produced by three different pathways, which vary in the carbon compound used as the substrate, as well as the source of the reducing potential (Figure 1). The hydrogenotrophic pathway is the most widespread, being found in all methanogenic orders. It involves the reduction of CO<sub>2</sub> with H<sub>2</sub> as an electron donor, and is composed of seven central steps (Figure 1) (Reeve et al. 1997). Formate can also be converted to methane through this pathway, acting as a source for CO<sub>2</sub> and reducing potential. Two other pathways are found in the order Methanosarcinales: the acetoclastic pathway and the methylotrophic pathway. In the acetoclastic pathway, acetate is split into a methyl group and CO, the latter being subsequently oxidized to provide electrons (Meuer et al. 2002). The methyl group from the splitting of acetate is linked to methanopterin (or sarcinapterin, for *Methanosarcina*) be-

fore being reduced to methane in two enzymatic reactions, homologous to the last two steps of the hydrogenotrophic pathway. The methylotrophic pathway, also present in one Methanobacteriales genus *Methanosphaera* (van de Wijn-gaard et al. 1991), has several possible variants (Meuer et al. 2002). The best-studied version is that where C-1 compounds such as methyl-amines or methanol can be used as both an electron donor and acceptor. One molecule of C-1 compound is oxidized (running the hydrogenotrophic pathway in the reverse direction from methyl-CoM to CO<sub>2</sub>) to provide electrons for reducing three additional molecules to methane. However, in the presence of methanol and H<sub>2</sub>/CO<sub>2</sub>, some Methanosarcinales can reduce this C-1 compound using only the last step of hydrogenotrophic methanogenesis (methyl-CoM to CH<sub>4</sub>), drawing electrons from H<sub>2</sub>.

*Methanosphaera* presents yet another variation of the methylotrophic pathway. This Methanobacteriales, unlike Methanosarcinales, cannot reduce CO<sub>2</sub> to produce methane (Schwörer and Thauer 1991). *Methanosphaera* requires methanol and H<sub>2</sub> for growth, reducing the former to methane in a process not yet fully understood, but which has been shown to overlap with hydrogenotrophic methanogenesis only in its last step (Schwörer and Thauer 1991).

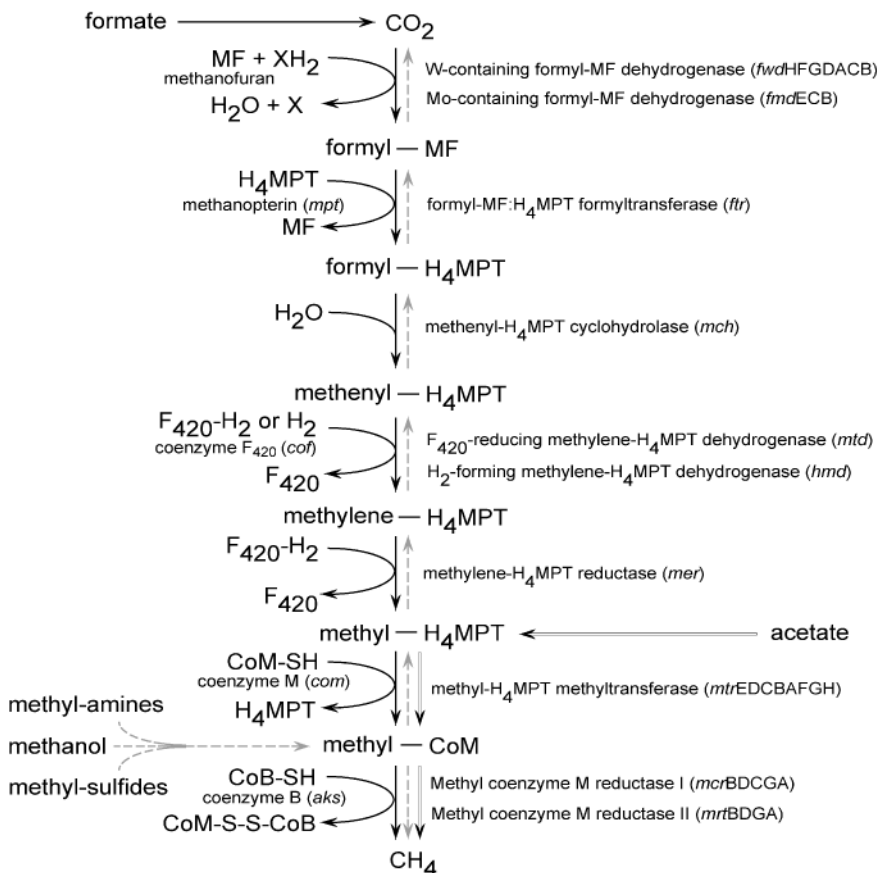


Figure 1. Pathways of methanogenesis: hydrogenotrophic (double-lined arrows), acetoclastic (solid arrows) and methylotrophic (broken gray arrows). The hydrogenotrophic pathway can operate from formate or H<sub>2</sub>/CO<sub>2</sub>, obtaining reducing potential from formate or H<sub>2</sub>, respectively. The acetoclastic pathway obtains electrons from the oxidation of CO produced by the splitting of acetate. The methylotrophic pathway can operate in two ways: (1) the methylated C-1 compound is a source for a methyl group as well as electrons (one molecule oxidized to CO<sub>2</sub> provides electrons for three molecules to be reduced to methane, dotted green arrows); and (2) the methylated C-1 compound is reduced to methane using electrons from H<sub>2</sub>. The names of the proteins studied here are indicated on the pathway. For the synthesis of the cofactors, *mptH* plays a role in the biosynthesis of tetrahydrofolate of *M. jannaschii*, and *mptN* is involved in the biosynthesis of methanopterin. Both tetrahydrofolate and methanopterin are intermediate C1 carriers in methanogenesis. The biosynthesis of the coenzyme F420 rests on the cofactors *cofC*, *cofD*, *cofE*, *cofG* and *cofH*. The *aks* enzymes (*aksA*, *aksD*, *aksE* and *aksF*) participate in

2-oxoacid elongation steps of coenzyme B9 biosynthesis, which partners with coenzyme M in the final step of methanogenesis. This last coenzyme is the smallest known organic cofactor, but an essential terminal methyl carrier during methanogenesis and is produced by the *com* enzymes (*comA*, *comB*, *comC*, *comD* and *comE*) (Graham and White 2002).

Because methanogenesis is found solely in the euryarchaeal branch of the archaeal domain, it most likely originated in that phylum. Biological production of methane requires at least 25 genes (in addition to more than 20 biochemically characterized proteins involved in the synthesis of the coenzymes). Genes encoding different subunits of an enzyme tend to be clustered together in the genome, but these clusters and genes encoding for monomers or homopolymers are scattered around the genome (Reeve et al. 1997). Some Methanosarcinales are estimated to have over 250 genes involved in different aspects of methanogenesis (Galagan et al. 2002). The number of genes involved, as well as their scattered genomic arrangement, makes it unlikely that methanogenesis could be acquired by lateral gene transfers (LGT). However, portions of the pathways involved in this process, along with single genes, very likely have been transferred across vast phylogenetic distances (Galagan et al. 2002). For example, homologs of enzymes catalyzing the first three steps of the methanogenic reduction of CO<sub>2</sub> are used for formaldehyde oxidation in methylotrophic Proteobacteria and Planctomycetes (Vorholt et al. 1999, Chistoserdova et al. 2004). It is extremely unlikely that these enzymes were present in the common ancestor of Archaea and Bacteria and lost in all but a few lineages of prokaryotes. Interdomain transfer(s) is a much more parsimonious possibility (Chistoserdova et al. 1998).

In the present work, we focus on the enzymes catalyzing the seven steps of the hydrogenotrophic methanogenesis pathway that are ubiquitous to methanogens, with the possible exception of *Methanospiraera*, of which its specialized lifestyle might lead to the loss of several methanogenesis enzymes as well as the cofactor methanofuran (van de Wijngaard et al. 1991). This ubiquity strongly suggests that hydrogenotrophic methanogenesis is the ancestral form of methane production, other pathways being subsequent innovations of individual lineages of methanogens. By combining phylogenies of methanogenesis genes and concatenated ribosomal proteins that include data from the partially sequenced genomes of the methanosarcinale *Methanococcoides burtonii* as well as the methanomicrobiale *Methanogenium frigidum*, we set out to determine if the phylogenetic relationships of methanogen orders suggested (i) by genome trees (Methanopyrales + Methanococcales + Methanobacteriales) and (ii) by 16S rRNA phylogenies (Methanomicrobiales + Methanosarcinales) could be considered as monophyletic. This information seems essential because the recently revised Bergey's Manual of Systematic Bacteriology presents a taxonomic grouping (the inclusion of Methanomicrobiales and Methanosarcinales along with Methanococcales in the class Methanococci) that is clearly polyphyletic and places *Methanopyrus kandleri* in a separate class because of its unresolved phylogenetic position. We also show that, contrary to expectation, most of the enzymes involved in methanogenesis have evolved by vertical descent. This finding is important because it is not obvious from previous work if these operational enzymes have evolved by vertical descent in methanogens, or if they have been subjected to extensive LGT and gene loss. Finally, we discuss

briefly why the identification of vertically inherited pathways could serve a purpose in nomenclature, notwithstanding the discovery of an ever-increasing number of laterally transferred operational genes in prokaryotes.

## Materials and methods

### *Constitution of the three data sets and preliminary phylogenetic analyses*

The data sets of individual ribosomal proteins were obtained from Brochier et al. (2004). These 53 ribosomal proteins (rpl1, rpl2, rpl3, rpl4, rpl5, rpl6, rpl10, rpl10e, rpl11, rpl13, rpl14, rpl15, rpl16, rpl18, rpl18e, rpl19e, rpl20a, rpl21e, rpl22, rpl23, rpl24, rpl24e, rpl29, rpl30, rpl31e, rpl32e, rpl34e, rpl35ae, rpl37e, rpl39e, rps2, rps3, rps3ae, rps4, rps4e, rps5, rps6e, rps7, rps8, rps8e, rps9, rps10, rps11, rps13, rps15, rps17, rps17e, rps19, rps19e, rps24e, rps27ae, rps27e and rps28e) were obtained from BLASTP and TBLASTN at the NCBI server (<http://www.ncbi.nlm.nih.gov/>) and aligned using CLUSTALW (Thompson et al. 1994) and the program ED of the MUST package (Philippe et al. 1993). Sequences from the two Methanosarcinales *Methanosarcina mazei* and *Methanosarcina acetivorans* and the Nanoarchaeon *Nanoarchaeum equitans*, for which complete genome sequences are now available, were added to these ribosomal data sets. In addition, we included two methanogens for which the genome has been partially sequenced: *Methanogenium frigidum* and *Methanococcoides burtonii* (Saunders et al. 2003). The first is a psychrophilic euryarchaeon belonging to the order Methanomicrobiales, based on 16S rRNA phylogenetic analysis (Franzmann et al. 1997), whereas the second is a mesophilic euryarchaeon of the Methanosarcinales order (Franzmann et al. 1992). Sequences were retrieved by TBLASTN from genome sequencing web sites for *M. burtonii* and *M. frigidum*, or by using BLASTP in NCBI for *N. equitans*, *M. acetivorans* and *M. mazei* (Altschul et al. 1990). New sequences were manually edited and added to the data sets. Regions where the alignment was ambiguous were removed from each data set. Data sets for proteins of methanogenesis and coenzyme synthesis were obtained at the NCBI. Biochemically characterized enzymes were used as queries to retrieve orthologs using BLASTP. We performed TBLASTN to look for *M. burtonii* and *M. frigidum* (Altschul et al. 1990). Amino acid sequences were aligned with CLUSTALW (default settings). Ambiguously aligned regions were deleted from the alignments.

Maximum likelihood (ML) phylogenetic analyses were performed with PROML with the JTT amino acid substitution matrix (Jones et al. 1992), a rate heterogeneity model with gamma-distributed rates over four categories, with the  $\alpha$  parameter estimated using TREE-PUZZLE, global rearrangements and randomized input order of sequences (10 jumbles). Bootstrap support values represent a consensus (obtained using CONSENSE) of 100 Fitch-Margoliash distance trees (obtained using PUZZLEBOOT and FITCH) from pseudo-replicates (obtained using SEQBOOT) of the original alignment. The settings of PUZZLEBOOT were the same as those used



for PROML, except that global rearrangements and randomized input order of sequences are unavailable in this program. PROML, CONSENSE, FITCH and SEQBOOT are from the PHYLIP package Version 3.6a (<http://evolution.genetics.washington.edu/phylic.html>). We obtained TREE-PUZZLE and PUZZLEBOOT from <http://www.tree-puzzle.de>.

*Concatenations, separate analyses of multiple markers and test for lateral gene transfers in the data sets*

*Phylogenetic analyses of the ribosomal data set and selection of the best tree* Selection of the best tree was based on the concatenation of the ribosomal proteins (called fusion), followed by the separate analyses of 23 of these proteins (rpl2p, rpl15p, rpl18p, rpl22p, rpl23p, rpl30p, rpl37ae, rpl3p, rpl44e, rpl4p, rps10p, rps13, rps15p, rps17e, rps19e, rps19p, rps2p, rps3p, rps4p, rps5p, rps6e, rps7p and rps8e). The concatenation of the 53 markers (6384 positions), representing 20 to 23 species of archaea, was used to calculate the most likely tree by PROML, JTT, and eight categories estimated from TREE-PUZZLE. To further test the relationships between methanogens, as well as the robustness of the best ML tree based on the fusion of the ribosomal proteins, 14 additional alternative topologies were constructed. These topologies were created to test specific hypotheses of relationships. Four topologies were created by local rearrangements of the fusion tree and 10 others were designed independently of it. Briefly, they explore some combinations of relationships between the following euryarchaeal taxa or groups: the (*Ferroplasmatales/Thermoplasmatales*), *Archaeoglobus*, the *Halobacteriales*, the *Pyrococcales* and the methanogens. For instance, in one tree, all these taxa emerge simultaneously; in others, the taxa *M. kandleri*, *M. thermoautotrophicus*, *M. jannaschii* and *M. maripaludis* are associated and considered as either a late or early lineage. Similar test trees explored the late/early emergence of the group of *M. frigidum*, *M. burtonii* and the *Methanosarcinales*. The monophyly of all the methanogens was also investigated (Appendix 1). Likelihoods of these different topologies were compared by an Approximately Unbiased (AU) test (Shimodaira 2002), using CONSEL (Shimodaira and Hasegawa 2001) to identify the best tree for the concatenated data set. This statistical test estimates if the likelihoods of trees harboring the same species, but with different relationships, differ significantly or not. When the AU test associates a *P* value that is < 0.05 to one of the topologies under study, then this tree can be trended as significantly different and worse than the other topologies for a given data set, at a threshold of 5%.

However, concatenation is not the most accurate approach for choosing among topologies when dealing simultaneously with multiple markers (Bapteste et al. 2002). A concatenation enforces a mean rate of evolution for species, and a mean alpha parameter for all sequences; however, not all the markers evolve at the same rate, nor have the same alpha parameter (data not shown). More appropriate than a simple concatenation is a separate analysis of the markers present in all the species. Here the separate analysis consists in testing the

support/rejection for the 15 topologies by 23 individual markers, retaining as the best organismal tree the topology that receives the largest number of individual supports, the smallest number of individual rejections and the highest average *P* value in the AU test.

We also determined if these ribosomal proteins were free of LGT events. To evaluate their phylogenetic signal and to identify potential LGTs, we manually designed a set of 197 topologies, many of which could be explained by LGT. The species part of the groups mentioned above, i.e., the members of the euryotes, of the crenotes, of the *Ferroplasmatales/Thermoplasmatales*, of the *Halobacteriales*, of the *Pyrococcales*, and of each of two groups of methanogens and *Archaeoglobus* were mixed in our test-trees in non-conventional groups to break the accepted relationships, as LGTs between species of these groups would do. Some of these trees were fully dichotomic (entirely resolved), but most of them presented soft polytomies, allowing alternative orders of emergence inside and between the sets of species. These topologies are available from the authors. The AU test was applied to this set of topologies at the level of 5%. If some phylogenetic signal is present in the markers, we expect the trees with deep polytomy (i.e., star phylogenies) to be rejected, and unless recent LGT occurred in our markers, no topology describing an LGT event should be supported.

*Phylogenetic analyses of the enzymes of methanogenesis and of the proteins involved in the synthesis of its coenzymes*

We reconstructed individual ML phylogenies of 20 proteins involved in hydrogenotrophic methanogenesis and of 15 proteins involved in the synthesis of the coenzymes of this pathway. The names and functions of these operational enzymes are listed in Figure 1. Bootstrap values for these individual phylogenies were calculated with PUZZLEBOOT. These analyses allowed the identification of two categories of markers: likely orthologs (presenting a single copy for each species of methanogens) and non-orthologs (presenting more than one copy for some species of methanogens) (Appendix 1). Only unambiguous orthologs were retained for further analyses to test the hypothetical monophyly of the two groups of methanogens on the basis of these two data sets. First, the two largest cores of genes of methanogens were defined. Two data sets were analyzed: (i) seven proteins of coenzyme synthesis (*aksD*, *aksE*, *aksF*, *cofD*, *cofG*, *cofH*, *comD*) and (ii) nine proteins/subunits of hydrogenotrophic methanogenesis (*ftf*, *fwdA*, *mcrA*, *mer*, *mtd*, *mtrB*, *mtrC*, *mtrD*, *mtrE*) (used for the trees in Figures 2B and 2C, respectively). The proteins of each of these data sets were concatenated, not to study the monophyly of methanogens, but to investigate the relationships between these archaea. To test that no alternative tree was more robustly supported than the fusion tree, nor preferred by any individual marker, we rearranged these best fusion trees to allow single species to be located at any alternative position in each of them, generating 65 and 44 rearranged trees, respectively. The rejection/acceptance of these topologies was evaluated by an AU test at the 5% level.

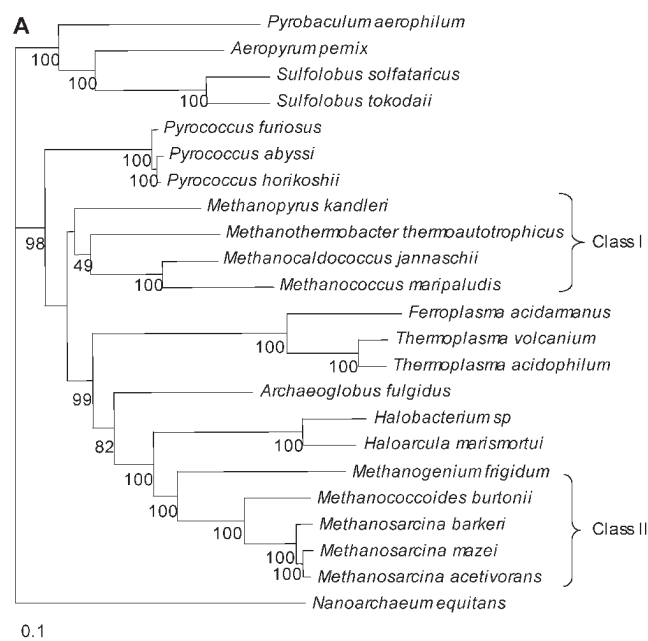


Figure 2

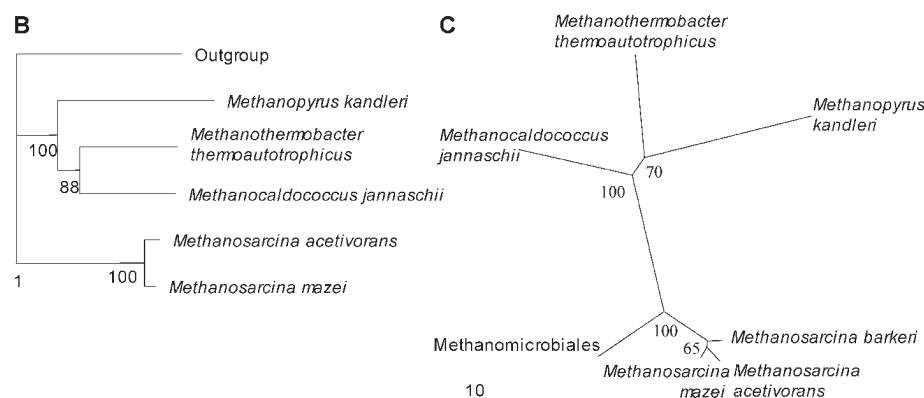


Figure 2. (A) Best maximum likelihood (ML) ribosomal tree obtained from fusion analyses of 53 markers and separate analyses of 23 markers. Two classes of methanogens are proposed. (B) Best ML tree based on the fusion analysis of seven orthologous proteins involved in the synthesis of co-factors (1432 positions). A non-methanogen sequence supposed to be a priori outside these two groups was also retained for the coenzyme data set: either *Archaeoglobus fulgidus* or *Ralstonia solanacearum*, if the former was not present. (C) Best ML tree, based on the fusion of nine orthologous proteins of the hydrogenotrophic methanogenesis (2382 positions). For A, B and C, only bootstrap values > 50% are indicated.

## Results and discussion

### *Methanogens are not monophyletic, but can be separated into two classes*

Methanogens can be defined functionally and ecologically as methane producers. Here we questioned their monophyly, which a priori is unexpected. If their monophyly is supported, we could infer that the production of methane evolved once in their last common ancestor. In contrast, the question of unique versus multiple (and independent) origins of methanogenesis needs to be answered if we conclude that methanogens are paraphyletic or polyphyletic. As in previous studies, our reference organismal tree resulting from the separate analysis of 23 ribosomal proteins rejected the monophyly of methanogens (Figure 2A).

The fusion of the 53 ribosomal proteins significantly favored a tree ( $P$  value of the AU test = 0.874) in which the monophyly of all methanogens was rejected. This tree sup-

ported two monophyletic groups of methanogens: (1) the Methanobacteriales and Methanococcales; and (2) the Methanomicrobiales and Methanosarcinales. *Methanopyrus kandleri* emerged on its own after the Pyrococcales divergence. All alternative topologies under study, except the one retained for Figure 2A ( $P$  value of the AU test = 0.126) were rejected. Moreover, a separate analysis favored this second topology, which proposes two monophyletic groups of methanogens over the basic fusion-tree (16/23 genes preferred it). The difference between the fusion-tree and this tree concerns the positioning of *Methanopyrus kandleri*. In the second tree, it groups with Methanococcales and Methanobacteriales (BV = 50%; Figure 2A), and Methanomicrobiales strongly grouped with Methanosarcinales (BV = 100%, Figure 2A).

The tree shown in Figure 2A may be treated as reasonable reference, because we tested that most of the genes used to build it reject most of the trees with simulated LGT. There is an average of 5% of non-rejected trees by gene, suggesting that

our 23 ribosomal genes contain a significant phylogenetic signal, free of the LGT effect. Only one gene coding for the 50S ribosomal protein L37Ae showed strong support for a unique topology with LGT (between *Haloarcula* and the Methanococcales, as previously described by Matte-Tailliez et al. (2002)), and showed no support for any other tree. Nine other genes (rpl18, rpl22, rps10, rps8e, rps19, rps19e, rps6e, rps17e and rpl23) did not reject several different topologies ( $\geq 5\%$ ) with LGT, and notably rps6e, rps17e and rpl23 did not reject several topologies involving LGT of methanogenic species. This indicates that the hypothesis of LGT involving methanogens cannot be discarded entirely nor proved for these nine markers, although this observation could also be the result of a weakness of the individual phylogenetic signals of these markers. Therefore, we assumed that, apart from rpl37ae, these 23 ribosomal proteins were vertically inherited (i.e., that their histories were free of LGT events), or that, if they contained LGT, these events would have an insignificant impact on our phylogenetic reconstruction. Thus, our analyses strongly rejected the existence of a large clade regrouping all the methanogens (BV = 99%). Similarly, individual phylogenies of proteins involved in coenzyme biosynthesis have never supported the monophyly of a large group of methanogens.

Although the monophyly of methanogens is rejected by our reference tree, taxa producing methane are grouped into two clades: (1) Methanopyrales + Methanobacteriales + Methanococcales, which we call Class I (BV < 50%, but statistically favored by the AU test); and (2) Methanomicrobiales + Methanosarcinales, which we call Class II (strongly supported, BV = 100%), in agreement with the most recent version of the Bergey's taxonomy (Garrity 2001) (see details at <http://141.150.157.80/bergeysoutline/main.htm>).

The monophylies of Class I and Class II are supported by 11 individual ML phylogenies of the hydrogenotrophic methanogenesis enzymes and one coenzyme biosynthesis protein (Appendix 1). The fusion of nine orthologs of the hydrogenotrophic pathway (Figure 2C) and of seven genes of coenzyme synthesis (Figure 2B) also support these two monophylies. The presence of Class II as a group is strongly supported by phylogenetic analyses of ribosomal proteins, methanogenesis enzymes and methanogenesis coenzyme biosynthesis proteins. This result is particularly important because it strengthens the existence of Class II methanogens, which was originally proposed based solely on 16S rRNA phylogenies. In contrast, ribosomal proteins only weakly support the presence of Class I, and individual phylogenies of the genes involved in the synthesis of coenzymes indicate that the paraphyly of Class I is often not rejected (three to six genes can accept a topology with paraphyletic Class I), making this group more suggested than supported by phylogenetic analyses of the coenzyme data set. Yet, methanogenesis enzymes display this group as strongly supported. Furthermore, the presence of five proteins (*hmdI* in methanogenesis + *cofC*, *comA*, *comB* and *comC* in coenzyme synthesis) solely in representatives of Class I suggests that these species may be members of a clade. The presence of these genes in all members of Class I could be

viewed as Class I innovations or Class II losses. In addition, genome trees (based on either gene content or conserved gene pairs) also support the monophyly of Class I (Slesarev et al. 2002). Slesarev et al. (2002) observed that three representatives of Class I methanogens share 59 COGs (Cluster of Ortholog Genes) that are not represented in any other archaea or bacteria and therefore appear to comprise a genomic signature of this group. Two of the three orders found in Class I, Methanopyrales and Methanobacteriales, also share a distinct feature—both have pseudomurein as a major component of their cellular envelope (Konig et al. 1989). Pseudomurein is not found in other organisms and is therefore likely to have originated in the common ancestor of the Methanopyrales and Methanobacteriales, making it a strong shared characteristic. Furthermore, all orders of Class I display hyperthermophiles, whereas none are found in Class II.

We propose restructuring the classification of methanogens at the class level, leaving all other taxonomic categories intact. The monophyletic groups observed in our analyses, Class I and Class II, would represent the only two classes of methanogens. The orders Methanobacteriales, Methanococcales and Methanopyrales would be grouped under Class I, and the orders Methanomicrobiales and Methanosarcinales would stay in the Methanomicrobia or Class II. This division of methanogens into two distinct classes improves on the old classification by being consistent with phylogeny (the old class Methanococci described in Bergey's Manual of Systematic Bacteriology 2001 is polyphyletic and the classes Methanococci and Methanobacteria are unrelated in the 2004 taxonomic update of this manual). Further phylogenomic investigations are needed to validate these groups.

#### *Evolution of the hydrogenotrophic methanogenesis*

It is not obvious from previous work if the enzymes of the hydrogenotrophic methanogenesis and the synthesis of its cofactor have evolved by vertical descent in methanogens, or if they have been subjected to extensive LGT and gene loss among methanogens, as reported for other "operational" proteins that can be a priori exchanged quite easily (Boucher et al. 2003). In fact, the history of this pathway is unlikely to be simple. First, it is obvious that hydrogenotrophic methanogenesis and the synthesis of coenzymes evolved mostly independently (Appendix 1). Enzymes involved in the synthesis of coenzymes are more broadly distributed than those of hydrogenotrophic methanogenesis (they are not restricted to the Archaea). In addition, these proteins seem to have undergone intricate evolutionary processes (duplications/losses and transfers although not between methanogens). The taxonomic distribution of the enzymes of hydrogenotrophic methanogenesis is more restricted, but is not limited to extant methanogens.

At the time that we performed this analysis, five genomes of methanogens had been completely sequenced: *M. jannaschii*, *M. kandleri*, *M. acetivorans*, *M. mazei* and *M. thermotrophicus* (*Methanococcus maripaludis* was completed afterwards). All these species harbor the vast majority of the

proteins involved in hydrogenotrophic methanogenesis and the synthesis of its specific coenzymes (Table 1). The only exceptions are *hmdl*, lacking in Class II, and *mtrF*, absent in *M. kandleri* for methanogenesis, as well as *cofC*, absent in Class II and *M. kandleri*, and *comABC*, absent in Class II for the synthesis of coenzymes. This shows that, although additional pathways are present for methanogenesis in some of these species, their existence was unaccompanied by multiple specific losses of steps of the hydrogenotrophic pathway studied here. The most parsimonious scenario would thus be that proteins for methanogenesis appeared in the context of a unique pathway (the hydrogenotrophic pathway) (Ferry 1999) and were conserved in species producing methane (with the possible exception of the specialized *Methanosphaera*) (van de Wijngaard et al. 1991).

We obtained no evidence that the steps of this pathway were elaborated progressively. We can only report that markers consistent with a unique origin of all methanogens (13 enzymes) are involved in the last two steps of the pathway (Appendix 1). This contrasts with the absence of statistical support for this monophyly for enzymes involved in the first steps of methanogenesis. However, this does not indicate that the first steps would be more ancient than the last steps; it only indicates that all methanogenesis pathways overlap at the last two steps of methane production. A strong evolutionary pressure for the conservation of these final enzymatic steps, while the beginning of the pathway could be tinkered with for metabolic flexibility, likely explains the differences in the phylogenies of genes of methanogenesis.

It thus appears that the methanogen clade exists as a much broader taxon than that usually distinguished as methanogens and includes, in addition to methanogens of Classes I and II, non-methanogenic archaea such as Thermoplasmatales, Archaeoglobales and Halobacteriales. These last three phyla would then represent degenerated methanogens that have lost most of the proteins involved in methanogenesis, especially those of the last two steps. The case of *A. fulgidus* is particularly interesting because it possesses five enzymes involved in methanogenesis, with the exception of the last two steps. It seems that these five enzymes are more likely involved with lactate oxidation (Vorholt et al. 1995) than with methanogenesis (although *A. fulgidus* is capable of reverse methanogenesis, resulting in CO<sub>2</sub> production). The absence of methyl-CoM reductase in this archaeon eliminates the possibility of methane production by conventional pathways (Klenck et al. 1997) and illustrates the idea that genes of methanogenesis are ancient in euryarchaea and were lost independently in various lineages. This vertical descent with differential loss may provide a good explanation for the orthology of hydrogenotrophic methanogenesis enzymes, but needs to be tested further.

Our AU tests indicated that six genes (out of nine) involved in methanogenesis are highly discriminatory and reject all trees other than the fusion-tree in which Class I and Class II are separated. Only three orthologs (*mtrE*, *mtrB* and *mtrC*) fail to reject 3/4/5 alternative trees (in addition to the best tree), where the Class I *M. jannaschii* branches within Class II. In contrast, paralogs a priori excluded from the fusion can accept

several rearranged trees mixing Class I and Class II, depending on which copy was retained, to represent a given lineage. This confirms both our choice of removing the paralogs from fusion to avoid introducing biases, as well as the efficiency of the concatenation analysis. It also strongly suggests that no recent transfer of genes involved in hydrogenotrophic methanogenesis has occurred between Class I and Class II for at least six genes located along the pathway. The analysis of the coenzyme-synthesizing proteins led to a similar conclusion. The monophylies of Class I and Class II are always supported, their polyphyly always rejected, and no recent transfers seem to have occurred between Class I and Class II for at least the seven non-paralogous coenzyme biosynthesis proteins. The paraphyly of Class II (two *Methanosarcina* plus *Methanococcoides* and/or *Methanomicrobium*, depending on available genes) is never supported (except by some combination of paralogs). Lateral gene transfer simply cannot be rejected for the enzymes of the two last steps that underwent duplications, but if we assume that the right paralogs among the duplicated copies can be identified, the number of possible LGT cases decreases to three. Transfers or recruitments sometimes occurred, but only outside the methanogens: *ptr* (enzyme of step 2) is present in four distantly related bacteria, and *mch* (enzyme of step 3) is present in seven distantly related bacteria (four of which are the same as for *ptr*) (Table 1). These bacteria have likely acquired these genes and use them to perform a different function such as formaldehyde oxidation (Vorholt et al. 1999).

In summary, (1) we observed homologous proteins, without evidence of LGT, between Class I and Class II, along the whole hydrogenotrophic pathway (and for the synthesis of some of its specific coenzymes). (2) The non-methanogenic *A. fulgidus* harbors enzymes involved in all but the last two steps of hydrogenotrophic methanogenesis, and Halobacteriales contain some of the proteins involved in methanogenesis. (3) Duplications were observed in seven genes; remarkably, six of them are in the last two steps of hydrogenotrophic methanogenesis and one in the first step. This, as well as a gene specific to Class I in step 4, suggests that some local specialization at crucial points of hydrogenotrophic methanogenesis is possible, but not by LGT, and that the global pathway is highly conserved. The hypothesis of an ancestral methanogenesis is more likely than its transfer between two distantly related and ecologically distinct archaeal groups, especially given that the genes of the hydrogenotrophic pathway are not found in an operon (only genes encoding for subunits of a single enzyme are usually linked).

*On the general interest of testing if pathways have single or multiple origins: toward a new nomenclature?*

The phylogenetic study of the evolution of a pathway is interesting to an evolutionist, especially when, as in the present study, it confirms the unique origin of the pathway. Indeed, in the context of possible LGTs (Doolittle 2000), the evolution of genes in a pathway would show little correlation and the duration of their association would be limited. It is now accepted that the genomes of organisms (especially prokaryotic organ-



Table 1. Distribution of genes coding for proteins involved in the methanogenesis pathway in various prokaryotes.

	Methanogenesis					Coenzyme biosynthesis																		
	<i>fwd</i>	<i>ftr</i>	<i>mch</i>	<i>mtd</i>	<i>hmd</i>	<i>mer</i>	<i>mtr</i>	<i>mcr</i>	<i>aksA</i>	<i>aksD</i>	<i>aksE</i>	<i>aksF</i>	<i>cofC</i>	<i>cofD</i>	<i>cofE</i>	<i>cofG</i>	<i>cofH</i>	<i>comA</i>	<i>comB</i>	<i>comC</i>	<i>comD</i>	<i>comE</i>	<i>mpH</i>	<i>mpN</i>
Euryarchaeota	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Methanosarcinales	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Methanomicrobiales	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Methanobacteriales	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Methanopyrales	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Methanococcales	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Archaeoglobales	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Thermoplasmatales	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Halobacteriales	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Thermococcales	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Crenarchaeota	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sulfobolales	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Thermoproteales	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Desulfurococcales	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Aquificae	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Thermotogae	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Deinococcus-Thermus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Chloroflexi	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cyanobacteria	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Chlorobi	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Proteobacteria	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
alpha	P	P	P	-	-	-	Ph*	-	P	-	P	-	-	-	-	-	-	-	-	-	-	-	-	-
beta	P	P	P	-	-	-	-	-	+	P	+	-	-	-	-	-	-	-	-	-	-	-	-	-
gamma	P	P	P	-	-	-	-	-	-	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-
delta	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
epsilon	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Firmicutes	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Actinobacteria	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Planctomycetes	P	P	P	-	-	-	Ph*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Planctomycetes	P	P	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Chlamydiae	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Spirochaetes	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Bacteroidetes	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Fusobacteria	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

+ = Present in all members of the order; presence is based on completely and partially sequenced genomes.

- = Absent from all members of the order; absence is based on completely sequenced genomes.

P = Present in some members of the order; a blank indicates that complete genome data are unavailable.

\* = Homology detected using inferred ancestral sequences for *mtra* (Pa\*) and *mtrh* (Ph\*).



isms) are mosaics and that the evolution of genomes cannot be accurately described by a unique tree (Doolittle 2000). The descriptive power of a notion such as “monophyly,” which is typically relevant in a tree-like context only, could then be limited and it could be misleading to rely only on it in classification if the molecular parts of organisms have multiple evolutionary histories. This claim has conceptually far-reaching consequences. Monophyly is generally considered the key to the definition of natural groups by phylogeneticists and many taxa have been reclassified accordingly to minimize paraphyly and polyphyly. In the context of LGT, however, natural groups of operational genes can still exist. Sets of molecular characters with a unique origin, which have evolved and have been transmitted as a unit thereafter, could be identified as some of the building blocks from which species are made. The evolution of the hydrogenotrophic methanogenesis described in this paper is an example of such a situation. Identifying this vertically transmitted set of genes allows the re-introduction of some accuracy and naturalness in organismal descriptions, even in the absence of a global Tree of Life on which classifications can be based.

We thus suggest, as a future challenge for evolutionists, to elaborate a partial but “natural and accurate” evolutionary description of organisms based on their long-lasting molecular units. This description would take the form of a chemical formula, in which identified units of genes are considered atoms. These atoms would receive a coefficient index to summarize what we know about their evolutionary history. For instance, coefficients like “ $-i$ ” would indicate that these units have been lost  $i$  times in the ancestors of the species, coefficients like “ $+i$ ” would indicate different degrees  $i$  of evolution of the unit present in the organism, and a zero would indicate that the species never harbored this unit. One advantage of this representation is that it would have a high explanatory power, indicating the origins of organismal properties or atoms (i.e., vertical descent or horizontal transfer). If we take only the example of hydrogenotrophic methanogenesis, denoted as “HM,” we could describe species containing the HM atom as follows: Class I *Methanopyrus kandleri* would be  $HM_{+1}$ , Class II *Methanococcoides burtonii* would be  $HM_{+2}$ , the secondarily amethanogenic *Archaeoglobus* would be  $HM_{-1}$ , whereas *Aeropyrum pernix*, which always lacked hydrogenotrophic methanogenesis, would be considered  $HM_0$ . Obviously, the more numerous the molecular units identified, the more complex, accurate and discriminatory the specific formula will be for each organism.

## Conclusion

Hydrogenotrophic methanogenesis is an evolutionary unit that can be lost, as is probably the case for Archaeoglobales and most likely for Halobacteriales and Thermoplasmatales. It can also be tinkered with, as seen in the acquisition of methylo-trophic and acetoclastic methanogenesis by Methanosarcinales and the apparent specialization of *Methanosphaera* in hydro-gen-fueled methanol reduction. We observed no evidence of

transfer events where bacteria, eukarya or other archaea would have recently acquired the core of the hydrogenotrophic methanogenesis. Furthermore, even in the context of LGT, where the exchange of metabolic enzymes between species is said to be frequent, a pathway such as hydrogenotrophic methanogenesis seems to have been mostly maintained by vertical descent. No transfers of enzymes of the hydrogenotrophic methanogenesis pathway seem to have occurred between the two classes of methanogens described here, which is unexpected for operational genes. Here, at maximum, LGT would have a role tinkering with end of this ubiquitous pathway, rather than in facilitating the evolution of prokaryotes, through the exchange of a functional package. We conclude that these archaeal genes, the ribosomal genes and those involved in the hydrogenotrophic methanogenesis, were vertically inherited at higher phylogenetic levels, and that a broader taxonomical monophyletic unit embedding the two classes of methanogens likely had a methanogenic ancestor. We also conclude that hydrogenotrophic methanogenesis has subsequently been lost in several lineages. For these reasons, methanogenesis is likely ancient (as suggested by the fossil record) (Brocks et al. 1999), and if these data are correct, methanogenic archaea are likely ancient as well, based on our conclusion that this process has a unique origin. The phylogenetic study of hydrogenotrophic methanogenesis has led us to propose an alternative “natural” nomenclature, which we hope will stimulate studies of the evolution and origins of biochemical pathways and of their taxonomical distribution, allowing us to identify some minimal and meaningful patterns of gene associations in the flux of genome evolution.

## Acknowledgments and notes

We thank D. Walsh for critical reading of the manuscript. This work was supported by Genome Atlantic and by a grant from CIHR (MOP4467) to W.F. Doolittle. Additional information on the test topologies used for the AU test, and the topologies involving casual gene transfer between methanogens are available from the corresponding authors.

## References

- Altschul, S.F., W. Gish, W. Miller, E.W. Myers and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215: 403–410.
- Baptiste, E., H. Brinkmann, J.A. Lee et al. 2002. The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. *Proc. Natl. Acad. Sci. USA* 99:1414–1419.
- Brochier, C., P. Forterre and S. Gribaldo. 2004. Archaeal phylogeny based on proteins of the transcription and translation machineries: tackling the *Methanopyrus kandleri* paradox. *Genome Biol.* 5:R17.
- Brocks, J.J., G.A. Logan, R. Buick and R.E. Summons. 1999. Archean molecular fossils and the early rise of eukaryotes. *Science* 285:1033–1036.
- Boucher, Y., C.J. Douady, R.T. Papke, D.A. Walsh, M.E. Boudreau, C.L. Nesbo, R.J. Case and W.F. Doolittle. 2003. Lateral gene transfer and the origins of prokaryotic groups. *Annu. Rev. Genet.* 37: 283–328.

- Burggraf, S., K.O. Stetter, P. Rouvière and C.R. Woese. 1991. *Methanopyrus kandleri*: an archaeal methanogen unrelated to all other known methanogens. *Syst. Appl. Microbiol.* 14:346–351.
- Castro, H., A. Ogram and K.R. Reddy. 2004. Phylogenetic characterization of methanogenic assemblages in eutrophic and oligotrophic areas of the Florida Everglades. *Appl. Environ. Microbiol.* 70: 6559–6568.
- Chistoserdova, L., C. Jenkins, M.G. Kalyuzhnaya, C.J. Marx, A. Lapidus, J.A. Vorholt, J.T. Staley and M.E. Lidstrom. 2004. The enigmatic planctomycetes may hold a key to the origins of methanogenesis and methylotrophy. *Mol. Biol. Evol.* 21:1234–1241.
- Chistoserdova, L., J.A. Vorholt, R.K. Thauer and M.E. Lidstrom. 1998. C1 transfer enzymes and coenzymes linking methylotrophic bacteria and methanogenic Archaea. *Science* 281:99–102.
- Doolittle, W.F. 2000. Uprooting the tree of life. *Sci. Am.* 28:90–95.
- Doolittle, W.F., Y. Boucher, C.L. Nesbo, C.J. Douady, J.O. Andersson and A.J. Roger. 2003. How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358:39–58.
- Ferry, J.G. 1999. Enzymology of one-carbon metabolism in methanogenic pathways. *FEMS Microbiol. Rev.* 23:13–38.
- Florin, T.H., G. Zhu, K.M. Kirk and N.G. Martin. 2000. Shared and unique environmental factors determine the ecology of methanogens in humans and rats. *Am. J. Gastroenterol.* 95:2872–2879.
- Franzmann, P.D., Y. Liu, D.L. Balkwill, H.C. Aldrich, E. Conway de Macario and D.R. Boone. 1997. *Methanogenium frigidum* sp. nov., a psychrophilic, H<sub>2</sub>-using methanogen from Ace Lake, Antarctica. *Int. J. Syst. Bacteriol.* 47:1068–1072.
- Franzmann, P.D., N. Springer, W. Ludwig, E. Conway de Macario and M. Rohde. 1992. A methanogenic archaeon from Ace Lake, Antarctica: *Methanococcoides burtonii* sp. nov. *Syst. Appl. Microbiol.* 15:573–581.
- Galagan, J.E., C. Nusbaum, A. Roy et al. 2002. The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res.* 12:532–542.
- Garrity, G. 2001. *Bergey's manual of systematic bacteriology* 721. Springer-Verlag, New York, 721 p.
- Graham, D.E. and R.H. White. 2002. Elucidation of methanogenic coenzyme biosyntheses: from spectroscopy to genomics. *Nat. Prod. Rep.* 19:133–147.
- Jones, D.T., W.R. Taylor and J.M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8:275–282.
- Klenk, H.P., R.A. Clayton, J.F. Tomb et al. 1997. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* 390:364–370.
- König, H., O. Kandler and W. Hammes. 1989. Biosynthesis of pseudomurein: isolation of putative precursors from *Methanobacterium thermoautotrophicum*. *Can. J. Microbiol.* 35:176–181.
- Matte-Tailliez, O., C. Brochier, P. Forterre and H. Philippe. 2002. Archaeal phylogeny based on ribosomal proteins. *Mol. Biol. Evol.* 5:631–639.
- McDonald, I.R. M. Upton, G. Hall, R.W. Pickup, C. Edwards, J.R. Saunders, D.A. Richie and J.C. Murrell. 1999. Molecular ecological analysis of methanogens and methanotrophs in blanket bog peat. *Microb. Ecol.* 38:225–233.
- Meuer, J., H.C. Kuettnner, J.K. Zhang, R. Hedderich and W.W. Metcalf. 2002. Genetic analysis of the archaeon *Methanosarcina barkeri* Fusaro reveals a central role for Ech hydrogenase and ferredoxin in methanogenesis and carbon fixation. *Proc. Natl. Acad. Sci. USA* 99:5632–5637.
- Philippe, H. 1993. MUST, a computer package of management utilities for sequences and trees. *Nucleic Acids Res.* 21:5264–5272.
- Reeve, J.N., J. Nolling, R.M. Morgan and D.R. Smith. 1997. Methanogenesis: genes, genomes, and who's on first? *J. Bacteriol.* 179: 5975–5986.
- Saunders, N.F., T. Thomas, P.M.G. Curmi et al. 2003. Mechanisms of thermal adaptation revealed from the genomes of the Antarctic Archaea *Methanogenium frigidum* and *Methanococcoides burtonii*. *Genome Res.* 13:1580–1588.
- Schwörer, B. and R.K. Thauer. 1991. Activities of formylmethanofuran dehydrogenase, methylenetetrahydromethanopterin dehydrogenase, methylenetetrahydromethanopterin reductase, and heterodisulfide reductase in Methanogenic bacteria. *Arch. Microbiol.* 155:459–465.
- Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* 51:492–508.
- Shimodaira, H. and M. Hasegawa. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17: 1246–1247.
- Slesarev, A.I., K.V. Mezhevaya, K.S. Makarova et al. 2002. The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. *Proc. Natl. Acad. Sci. USA* 99:4644–4649.
- Takai, K. and K. Horikoshi. 1999. Genetic diversity of archaea in deep-sea hydrothermal vent environments. *Genetics* 152: 1285–1297.
- Thompson, J.D., D.G. Higgins and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–4680.
- van de Wijngaard, W.M., J. Creemers, G.D. Vogels and C. van der Drift. 1991. Methanogenic pathways in *Methanosphaera stadtmanae*. *FEMS Microbiol. Lett.* 64:207–211.
- Vorholt, J.A., L. Chistoserdova, S.M. Stolyar, R.K. Thauer and M.E. Lidstrom. 1999. Distribution of tetrahydromethanopterin-dependent enzymes in methylotrophic bacteria and phylogeny of methenyl tetrahydromethanopterin cyclohydrolases. *J. Bacteriol.* 181:5750–5757.
- Vorholt, J.A., J. Kunow, K.O. Stetter and R.K. Thauer. 1995. Enzymes and coenzymes of the carbon monoxide dehydrogenase pathway for autotrophic CO<sub>2</sub> fixation in *Archaeoglobus lithotrophicus* and the lack of carbon monoxide dehydrogenase in the heterotrophic *A. profundus*. *Arch. Microbiol.* 163:112–118.
- Woese, C.R. and G.J. Olsen. 1986. Archaeobacterial phylogeny: perspectives on the urkingdoms. *Syst. Appl. Microbiol.* 7:161–177.

Appendix 1. Summary of the phylogenetic information of the operational genes studied.

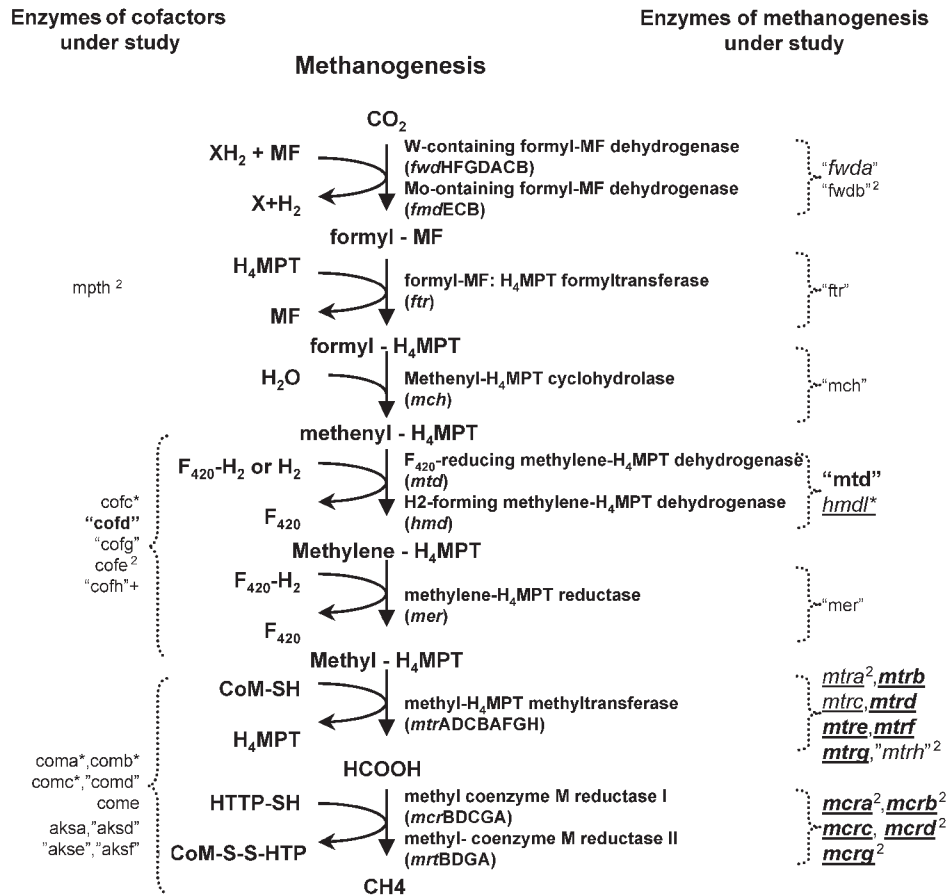
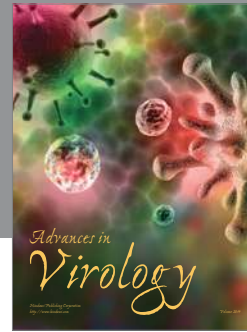
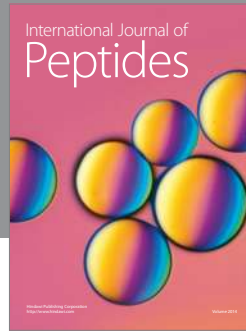
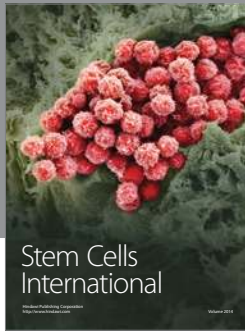


Figure A1. Enzymes are listed along the hydrogenotrophic pathway. Enzymes supporting the monophyly of methanogens are in italics (BV >75%). Enzymes supporting the monophyly of both Class I and Class II are in bold. Enzymes present in methanogens only are underlined. Genes present in Class I only are indicated by an asterisk. Phylogeny supporting the monophyly of Class II only are indicated by a +. Genes without lateral gene transfers between Class I/ Class II are in quotes. Genes with duplication in methanogens, leading to incongruent position of a species in the tree are denoted by a superscript 2. Sixteen genes out of 20 coding for the seven core enzymes of hydrogenotrophic methanogenesis (*cofD*, *fwdA*, *hmdI*, *mtrA*, *mtrB*, *mtrC*, *mtrD*, *mtrE*, *mtrF*, *mtrG*, *mtrH*, *mcrA*, *mcrB*, *mcrC*, *mcrD* and *mcrG*) are consistent with the monophyly of methanogens. In other words, proteins of methanogenesis themselves seem in contradiction with the reference tree based on

ribosomal proteins. In fact, 13 of the 15 proteins consistent with the monophyly of methanogens are exclusively present in methanogens (an additional protein is exclusively found in Class I, see Table 1). Such a restricted taxonomical sampling obviously increases the number of markers with monophyletic methanogens, but does not really help to test this monophyly. A crude interpretation could be that methanogenesis is a plesiomorphic trait, ancestral to most of the euryarchaea. It would have been lost several times independently in the Thermoplasmatales, Archaeoglobales and Halobacteriales lineages.





**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

