PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2006) **361**, 1007–1022
doi:10.1098/rstb.2006.1841
Published online 9 May 2006

The origin and evolution of Archaea: a state of the art

Simonetta Gribaldo[1,*] **and Celine Brochier-Armanet**[2]

[1]*Unité Biologie Moléculaire du Gène chez les Extremophiles, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France*
[2]*EA EGEE (Evolution, Génome, Environnement) Université Aix-Marseille I, Centre Saint-Charles, Case 36, 3 Place Victor Hugo, 13331 Marseille Cedex 3, France*

Environmental surveys indicate that the Archaea are diverse and abundant not only in extreme environments, but also in soil, oceans and freshwater, where they may fulfil a key role in the biogeochemical cycles of the planet. Archaea display unique capacities, such as methanogenesis and survival at temperatures higher than 90 °C, that make them crucial for understanding the nature of the biota of early Earth. Molecular, genomics and phylogenetics data strengthen Woese's definition of Archaea as a third domain of life in addition to Bacteria and Eukarya. Phylogenomics analyses of the components of different molecular systems are highlighting a core of mainly vertically inherited genes in Archaea. This allows recovering a globally well-resolved picture of archaeal evolution, as opposed to what is observed for Bacteria and Eukarya. This may be due to the fact that no rapid divergence occurred at the emergence of present-day archaeal lineages. This phylogeny supports a hyperthermophilic and non-methanogenic ancestor to present-day archaeal lineages, and a profound divergence between two major phyla, the Crenarchaeota and the Euryarchaeota, that may not have an equivalent in the other two domains of life. Nanoarchaea may not represent a third and ancestral archaeal phylum, but a fast-evolving euryarchaeal lineage. Methanogenesis seems to have appeared only once and early in the evolution of Euryarchaeota. Filling up this picture of archaeal evolution by adding presently uncultivated species, and placing it back in geological time remain two essential goals for the future.

Keywords: Archaea; evolution; phylogenomics; methanogenesis; hyperthermophyly; *Nanoarchaeum*

1. INTRODUCTION

An important part of life on Earth has remained unnoticed until about 30 years ago and then rapidly been framed as a curious biota predominant in extreme environments. However, the Archaea are now known to be metabolically diverse organisms coexisting with Bacteria and Eukarya in the majority of Earth environments, both terrestrial and aquatic, including extreme ones, such as high or low pH, low temperature, high salinity or pressure (Rothschild & Mancinelli 2001). Not only are the Archaea very diverse in virtually all environments, but they can also be very abundant. Their predominance in marine plankton, including deep oceans, points to a crucial and still poorly known role in the biogeochemical cycles of our planet (Karner *et al.* 2001; Lopez-Garcia *et al.* 2001). In addition, the Archaea include so far the sole organisms capable of methanogenesis (methane production from H_2 and CO_2), and for this reason they are central to palaeoenvironment and palaeobiology studies. A recent survey of environmental sequences indicates that we know only the tip of the iceberg of archaeal diversity (Schleper *et al.* 2005).

* Author for correspondence (simo@pasteur.fr).

One contribution of 14 to a Discussion Meeting Issue 'Major steps in cell evolution'.

Here, we will review our current understanding on the origin and evolution of the Archaea based on most recent data. We will try to discuss a few questions that remain poorly understood such as: Are the Archaea an ancient lineage or do they derive from within Bacteria? What are the evolutionary relationships between the Archaea and the other two domains of life? What was the nature of the last archaeal ancestor? When did methanogenesis originate and how did it evolve? Did Archaea evolve differently from Bacteria?

2. ARCHAEA ARE A THIRD DOMAIN OF LIFE

The assignment of Archaea to a third domain of life in addition to Bacteria and Eukarya, based on universal small subunit ribosomal RNA (SSU rRNA) and protein trees (Woese *et al.* 1990), has been validated by comparative genomics. The distinctive nature of Archaea that was realized by early studies on molecular mechanisms is still valid today: numerous components of archaeal informational processes are more similar to their eukaryotic than bacterial homologues and sometimes are uniquely shared by Archaea and Eukarya to the exclusion of Bacteria. The most striking case is the DNA replication apparatus: archaeal/eukaryal primases, helicases and replicative polymerases are totally unrelated to their bacterial counterparts (Olsen & Woese 1996). In addition, archaeal molecular systems generally show a level of complexity—in terms of

number of components—halfway through that of bacterial and eukaryal ones. For example, a number of ribosomal proteins are uniquely shared between Archaea and Eukarya, while none is uniquely shared with Bacteria or between Bacteria and Eukarya (Lecompte *et al.* 2002). The same tendency is found in other molecular systems linked to informational processes such as transcription, protein co-translational targeting and RNA metabolism (see below). However, Archaea are also remarkably similar to Bacteria in many respects, such as the size and organization of their chromosomes, the presence of polycistronic transcription units and the utilization of Shine–Dalgarno sequences for the initiation of translation (although not exclusively; Londei 2005). To sum up, the Archaea look like organisms that use eukaryotic-like proteins in a bacterial-like context (Myllykallio *et al.* 2000; Bell & Jackson 2001; Grabowski & Kelman 2003). Nevertheless, a growing number of studies of archaeal molecular biology are unveiling a rather sophisticated level of complexity, such as the combination of multiple origins for chromosome replication (Robinson & Bell 2005), so far not known in any Bacteria, and the possible implication of chromatin proteins in transcription regulation (Bell *et al.* 2002). Finally, the identification of archaeal genomic signatures (i.e. the core of genes exclusively shared by archaeal species) gives a measure of the distinctiveness of Archaea as a coherent group, although these signatures can differ according to the degree of stringency (Makarova & Koonin 2005; Walsh & Doolittle 2005).

Despite their similarities to both bacterial and eukaryal features, the Archaea harbour a unique trait that radically distinguishes them from the other two domains: the stereochemistry of the glycerol backbone of their membrane phospholipids is opposed to that of bacterial and eukaryotic ones. With no exception upto today, all archaeal membrane phospholipids are isoprenoid ethers built on glycerol-1-phosphate (G1P), while bacterial and eukaryal membranes contain fatty acid esters linked to the stereoisomer glycerol-3-phosphate (G3P) (Kates 1993). This idiosyncrasy thus lies in the very origin of Archaea and may have accompanied their diversification (Pereto *et al.* 2004).

3. HOW OLD?

Microbial life can be traced back to the Archaean (greater than 2500 million years ago) based on the ratios of biogenic isotopes distinctive of different metabolisms, but also on microfossils traces and biomarkers. The most ancient reliable biomarkers for bacterial (and possibly eukaryal) life are given by the presence of hopanes and steranes in 2.7 Gyr Archaean shales (Brocks *et al.* 1999; Summons 1999). On the contrary, extended isoprene chains (greater than C_{20}), which are good fossil biomarkers for archaeal lipids, are less stable and have only been found in rocks up to 1.6 Gyr old (Summons *et al.* 1988). However, the isotopic record of ultralight carbon indicates the presence of methane of biological origin (i.e. methanogenesis) at 2.7 Gyr ago (Ga) (Hayes 1994). In addition, evidence of sulphate reduction at 3.4 Ga (Shen *et al.* 2001) suggests that anaerobic consortia of archaeal methanogens and bacterial sulphate reducers, similar to those found in present-day anoxic marine sediments, may have already been in place at that time (Michaelis *et al.* 2002). As today, these consortia may have already included archaeal methanotrophs (Michaelis *et al.* 2002), since no anaerobic methane-oxidizing Bacteria are known (Chistoserdova *et al.* 2005). Both aerobic and anaerobic methanotrophy have been used to explain the highly depleted carbon isotopic values found in 2.8–2.6 Gyr geologic formations. Since oxygen would have still been a trace element in the atmosphere at the time, archaeal anaerobic methanotrophy is likely to have preceded bacterial aerobic methanotrophy.

The antiquity of archaeal fossil traces has been objected (Cavalier-Smith 2002) and possibly needs further confirmation. Moreover, the lack of reliable fossil traces for Archaea may severely affect any attempt to date the origin of this domain by molecular data. Hedges and colleagues recently estimated the divergence between Euryarchaeota and Crenarchaeota as old as 4.1 Gyr, but this was inferred by using the plant/animal divergence as a calibration point (Battistuzzi *et al.* 2004). The techniques to identify archaeal fossil traces in old samples should be developed further in the future and will provide reliable calibration points for the molecular dating of Archaea and prokaryotes in general.

4. ANCIENT OR DERIVED?

The few universal rooted trees that were produced between the end of the 1980s and the end of the 1990s (Gogarten *et al.* 1989; Iwabe *et al.* 1989; Brown & Doolittle 1995; Lawson *et al.* 1996; Gribaldo & Cammarano 1998) nicely converged on a single scenario which is frequently the implicit starting point of discussions on early evolution. In this model, Bacteria derived directly from the last universal common ancestor (LUCA), whereas Archaea and Eukarya share a last common ancestor more recent than LUCA, and are thus sister lineages (Woese *et al.* 1990). This widespread vision of relationships among domains is the reason why the similarities between the informational mechanisms of Archaea and Eukarya are commonly perceived as derived features that appeared in the lineage leading to their common ancestor, while bacterial counterparts represent ancestral traits. This is in agreement with the idea that eukaryotic cells are more complex and derived than prokaryotic ones. However, it is also legitimate to postulate that Archaea and Eukarya have retained ancestral traits, while Bacteria are very derived. Eventually, all three domains may harbour derived traits, and none of them can be traced back to the LUCA. This classical model of the universal tree of life is still surprisingly predominant despite current evidence that molecular saturation renders the phylogenetic signal harboured by such ancient paralogous couples definitely unreliable (Forterre & Philippe 1999*a*,*b*; Gribaldo & Philippe 2002). Consequently, the classical bacterial rooting may be no better than the two other alternatives (i.e. eukaryotic or archaeal rooting), and the exclusive sister relationship between Archaea and Eukarya is far from being established. Finally, chimeric models for the

origin of Eukarya can also explain the presence of archaeal-like genes in this domain (Lopez-Garcia & Moreira 1999; Rivera & Lake 2004).

The question of whether archaeal/eukaryal common traits are ancestral or derived is solved if Archaea, or the archaeal/eukaryal ancestor, arose from within Bacteria (Gupta 1998; Cavalier-Smith 2002). This model, where Archaea (and Eukarya) can be considered as modified Bacteria, predicts the possibility that members of one domain may give rise to another domain. This specifically requires an episode of dramatic evolutionary acceleration in the branch leading to Archaea (Gupta 1998), or to the ancestor of Archaea/Eukarya (Cavalier-Smith 2002). Such an event would mask the real origin of archaeal/eukaryal sequences and distort universal trees, an argument somehow similar to those advanced by proponents of chimeric hypotheses for the origin of Eukarya (Lopez-Garcia & Moreira 1999). The trigger for such a dramatic episode of evolutionary acceleration was proposed to have been selection pressure for antibiotic resistance, or appearance of archaeal/eukaryal type histones to protect DNA against thermal denaturation (Cavalier-Smith 2002). However, the fact that single point mutations are sufficient to produce drug resistant versions of antibiotic targets somehow weakens the first hypothesis. As for the second, bacterial HU histones have replaced the endogenous archaeal counterparts in Thermoplasmatales, without seemingly triggering any drastic evolutionary acceleration at the genome level. Finally, both hypotheses do not easily explain the replacement of the bacterial DNA replication apparatus by the totally unrelated archaeal/eukaryal one, and the change in the stereochemistry of the glycerol backbone of bacterial/eukaryal lipids in the lineage leading to Archaea. Indeed, selection pressure for adaptation to life at high temperature does not seem to be a sufficient trigger for switching from a G3P to G1P glycerol backbone, since hot-loving Bacteria have arisen from mesophilic lineages at least twice in evolution, and adapted their lipids to mimic archaeal ones without changing their backbone stereochemistry.

In conclusion, different explications for the relationships among domains have their own strengths and weaknesses, and at present we do not know where the answer lies (and we may never do; Bapteste & Brochier 2004).

5. TOWARDS A NATURAL HISTORY OF THE ARCHAEA

Is it possible to reconstruct the evolutionary history of the Archaeal domain from molecular data? Shall we expect a demoralizing lack of resolution similar to that shown by current bacterial and eukaryal phylogenies or may we hope to obtain a clearer picture?

As shown in figure 1, SSU rRNA sequences from cultivated Archaea are assigned to the two archaeal phyla currently recognized in the *Bergey's manual of systematic bacteriology* (Boone & Castenholz 2001), the Euryarchaeota and the Crenarchaeota, and to a third recently proposed phylum, the Nanoarchaeota (Huber *et al.* 2002). The number of archaeal phyla is surprisingly small with respect to the 12 bacterial phyla currently recognized in the *Bergey's manual of systematic bacteriology* (Boone & Castenholz 2001) and the 25–35 phyla recognized when divisions of non-cultivated bacteria are considered (Hugenholtz *et al.* 1998; Hugenholtz 2002). Cultivated Crenarchaeota fall into four orders (Thermoproteales, Caldisphaerales, Desulfurococcales and Sulfolobales) within the unique class of *Thermoprotei*, while cultivated Euryarchaeota fall into eight classes (Thermococci, Methanopyri, Methanococci, Methanobacteria, Thermoplasmata, Archaeoglobi, Halobacteria and Methanomicrobia) (Boone & Castenholz 2001).

Despite its poor resolution, the SSU rRNA tree (figure 1) has been until recently the only reference on which the origin and evolution of many archaeal features, together with the very nature of the archaeal ancestor, were based. The grouping of hyperthermophilic phyla at the base of the archaeal SSU rRNA tree (both in Crenarchaeota and Euryarchaeota) suggests that all extant Archaea are derived from a hyperthermophilic ancestor. The basal position of Methanopyrales (the only representative being *Methanopyrus kandleri*) within Euryarchaeota, separated from other methanogens by Thermococcales, has supported for a long time the idea that methanogenesis is an ancestral trait in Euryarchaeota (Burggraf *et al.* 1991). Environmental sequences form different and often large groups that are interspersed among cultivated species. Cultivated species appear to represent a minority in Crenarchaeota, which comprise an important number of sequence groups from marine plankton, freshwater samples, deep sub-surfaces and soil environments (Schleper *et al.* 2005). Their late emergence with respect to hyperthermophilic lineages suggests a single episode of adaptation to mesophilic environments in Crenarchaeota. Regarding Euryarchaeota, a number of environmental sequences expand the diversity of known groups, with a number of groups specifically related to Thermoplasmatales and to Halobacteriales, and with three groups from anaerobic methane oxidizer consortia that appear close to Methanosarcinales (Schleper *et al.* 2005). Two groups of sequences from hyperthermophilic environments do not fall within Crenarchaeota nor Euryarchaeaota: these belong to the Korarchaeota, a phylum proposed nearly 10 years ago (Barns *et al.* 1996) and still including exclusively uncultivated species, and the ancient archaeal group (AAG; Takai & Horikoshi 1999).

Although the small number of completely sequenced archaeal genomes (23 against 231 for Bacteria, as at September 2005) may not look like an ideal starting point to increase the poor resolution of relationships among archaeal phyla provided by SSU rRNA, this situation may hide a positive aspect: in fact, it allows performing accurate phylogenetic analysis in a quasi-exhaustive way by the simultaneous phylogenetic analysis of many genes (also called supermatrix approach) (Delsuc *et al.* 2005). This avoids recurring to whole-genome approaches, whose results can be severely affected by horizontal gene transfer (HGT; Delsuc *et al.* 2005). In fact, nearly all whole-genome trees, whatever the method employed, systematically misplace Halobacteriales and Thermoplasmatales either at the base of the Archaea, or at the base of

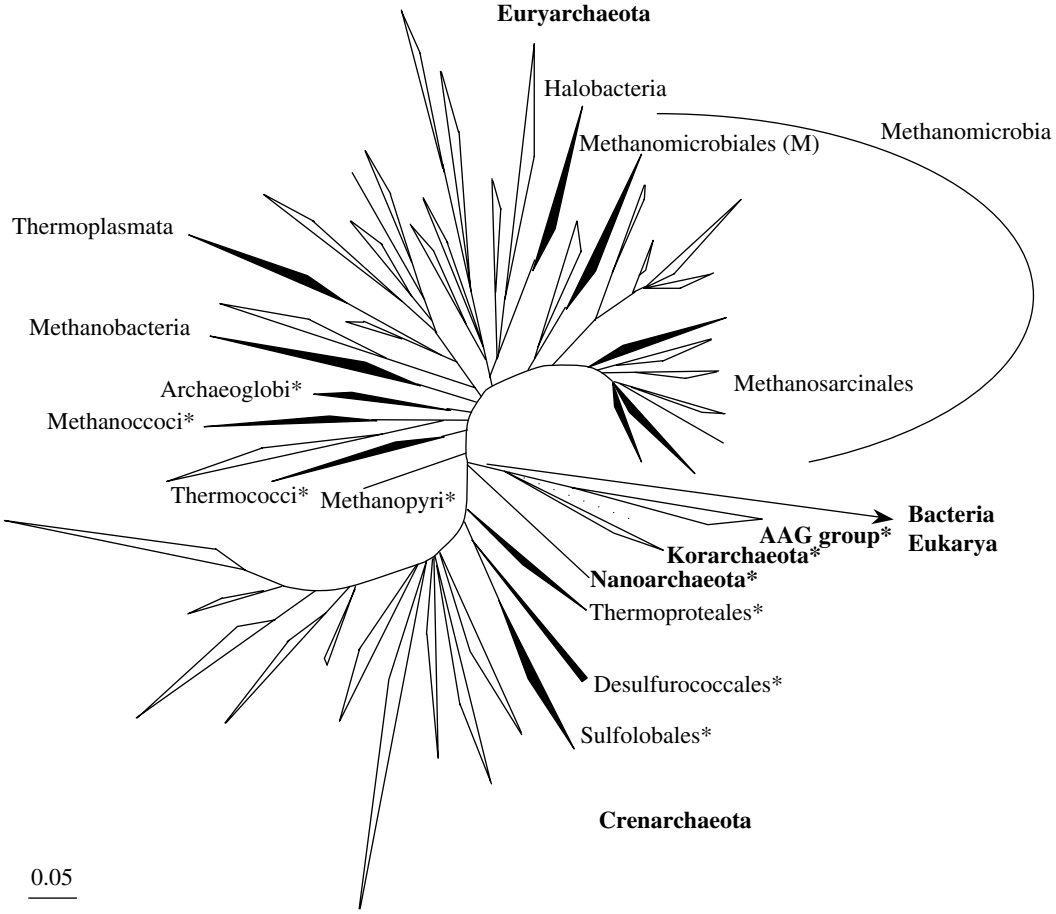

Figure 1. SSU rRNA archaeal phylogeny (adapted from Schleper *et al.* 2005). Triangles are proportional to the diversity of groups. Filled triangles represent groups for which cultivated species are available. Empty triangles indicate groups represented only by environmental sequences. Asterisks indicate groups with hyperthermophilic species.

Euryarchaeota, or even as sister groups of Crenarchaeota (see Slesarev *et al.* 2002; Wolf *et al.* 2002; Gophna *et al.* 2005 and references therein). This is most likely due to a bias introduced by HGTs. In fact, the genome of *Halobacterium* NRC1 contains an important number of genes that were recruited from Bacteria (Kennedy *et al.* 2001). Similarly, the presently available genomes from members of Thermoplasmatales contain a high proportion of genes shared exclusively with Sulfolobales (an archaeal genus inhabiting the same thermoacidophilic environments) (Ruepp *et al.* 2000; Futterer *et al.* 2004; Gophna *et al.* 2005). The basal placement of Thermoplasmatales and Halobacteriales in whole-genome archaeal trees thus likely results from an attraction by the sequences of Sulfolobales and by the usually included bacterial outgroup, respectively. Among all whole-genome tree approaches, supertree methods may be less sensitive to HGT, if these do not occur systematically in one direction (Delsuc *et al.* 2005). For example, a recent universal supertree recovered the late branching of Halobacteriales and Thermoplasmatales observed in the SSU rRNA tree (Daubin *et al.* 2002).

6. THE ARCHAEAL PHYLOGENOMIC CORE

The extent of HGT and whether it may be possible to reconstruct the phylogeny of species by using a core of vertically inherited genes is still a matter of debate (Daubin *et al.* 2003; Bapteste *et al.* 2004; Brochier *et al.* 2005*a*; Ochman *et al.* 2005). Unfortunately, discussions on this issue have started being perceived as defining two separate factions: pro-HGT and anti-HGT scientists. This does not make much sense to us, given that no one can deny that genome evolution—at least in prokaryotes—can be described by a web, and that potentially every gene can be transferred, even informational ones (Brochier *et al.* 2002; Matte-Tailliez *et al.* 2002). However, not every laterally transferred gene has the same probability of fixation. This is in fact affected by a number of factors—such as the selective advantage brought by the new product, the presence of an endogenous homologue performing the same function, the site of insertion in the genome and even chance alone—whose individual impact is difficult to estimate and difficult to know. What we observe in practice is that the transfer of some genes has been more successful than that of others. Consequently, the choice of the genes to use for phylogeny should depend on the level of evolutionary relationships we want to study. Genes that are not often transferred, such as those coding for ribosomal proteins (r-proteins) or rRNA, can be useful to reconstruct ancient phylogenies, while even genes that have experienced transfer can still be used to retrace the phylogeny of a group where no transfer has occurred for that particular gene. A good way of tackling the problem of HGT is then to choose the genes that are best adapted to the evolutionary level under study, and to identify and

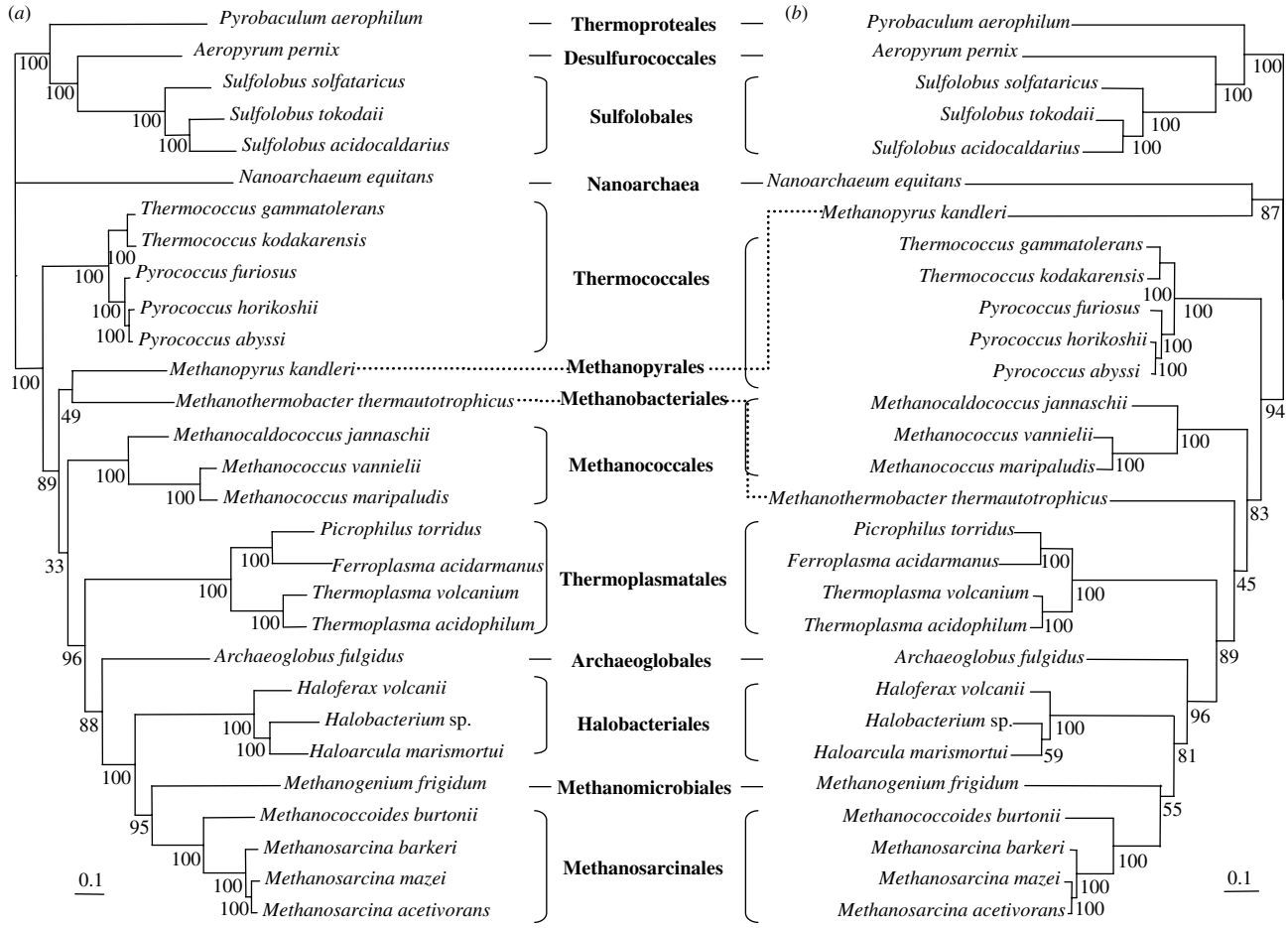

Figure 2. Unrooted maximum-likelihood (ML) trees based on a concatenation of ribosomal proteins (5809 positions) (*a*) and RNA polymerase subunits and transcription factors (2213 positions) (*b*) for which no HGTs were detected in individual analyses. For complete details about the individual and concatenated dataset construction see Brochier *et al.* (2004). The trees were calculated by PHYML (JTT model including a gamma correction (eight discrete classes) with an estimated alpha parameter, and an estimation of the proportion of invariant positions; Guindon & Gascuel 2003). Numbers at nodes are bootstrap values calculated from 1000 replicates by PHYML (Guindon & Gascuel 2003). The scale bar represents the per cent of substitutions per site.

remove before analysis those that have been clearly transferred.

Since the ribosome appears to be one of the most conserved macromolecular machines, the phylogenetic analysis of its component has been an immediate choice for the study of deep evolutionary transitions. The concatenation of r-proteins sequences is becoming quite common and is slowly replacing SSU rRNA-based trees for describing the phylogenetic position of a particular species when its complete genome is published (Slesarev *et al.* 2002; Waters *et al.* 2003). Global archaeal and bacterial phylogenies produced by this approach turned out to be generally congruent with rRNA-based reference trees, but sometimes with some local important differences and often a much higher statistical robustness (Brochier *et al.* 2002; Matte-Tailliez *et al.* 2002). However, since they belong to the same macromolecular machinery, rRNA and r-protein genes may be subject to similar biases that would reinforce eventual phylogenetic misplacements. This hypothesis can be tested by building phylogenies based on the components of different macromolecular systems and verifying whether any consistent evolutionary picture arises. This was indeed proven for the Archaea, where unrooted trees based on the concatenation of components of the transcription machinery (RNA polymerase (RNAP) subunits and transcription factors) were shown to be remarkably congruent with those based on r-protein concatenations (Brochier *et al.* 2004). Moreover, congruence appears to become more pronounced with an improved taxon sampling, indicating that these proteins are part of what may be called a *phylogenomic core* of the Archaea, which can be used to retrace the global history of this domain (Charlebois & Doolittle 2004; Brochier *et al.* 2005*a*)

In figure 2 are two updated versions of these archaeal 'translation' and 'transcription' trees (Brochier *et al.* 2005*a*) following the addition of four species (*Sulfolobus acidocaldarius*, *Thermococcus kodakarensis*, *Picrophilus torridus*, *Methanococcus vannielii*) whose genomes have or are being completely sequenced. Most nodes are well supported statistically and congruent between the two trees, and confirm a number of relationships between major phyla, as well as within phyla, indicated by our previous analyses (Brochier *et al.* 2005*a*). For example, both trees recover the same order of emergence of major groups within Crenarchaeota,

with Thermoproteales representing the first diverging lineage, and a sister relationship between Desulfurococcales and Sulfolobales (figure 2*a*,*b*). Within Euryarchaeota, both trees strongly support the same order of emergence of four large families, Thermoplasmatales, Archaeoglobales, Halobacteriales and Methanomicrobiales/Methanosarcinales, in this order (figure 2*a*,*b*). In the transcription tree, a weak support (bootstrap value (BV) = 55%) is associated to the grouping of *Methanogenium frigidum* and Methanosarcinales (figure 2*b*), likely due to the fact that a number of sequences of RNAP subunits are not yet available for *M. frigidum*. As for the base of Euryarchaeota, the first divergence is represented by Thermococcales in the translation tree (BV = 89%; figure 2*a*), while in the transcription tree, things are rendered more complicated by the fact that *M. kandleri*, the sole representative of Methanopyrales, branches off with *Nanoarchaeum equitans*, the sole representative of Nanoarchaeota (BV = 87%; figure 2*b*). We believe that this incongruence is caused by a long-branch attraction (LBA) artefact in the transcription tree, as in our previously analyses (Brochier *et al.* 2005*a*), and we will discuss it further in §7. Finally, the branching order of Methanobacteriales and Methanococcales is different in the two trees, but the weak support (BV = 33% for the emergence of Methanobacteriales before Methanococcales (figure 2*a*) and that (BV = 45%) for the emergence of Methanococcales prior to Methanobacteriales (figure 2*b*)) suggests that this is due to lack of sufficient signal. Given that our trees are unrooted, the placement of *N. equitans* on a branch that does not emerge within Crenarchaeota nor within Euryarchaeota cannot be taken as support for the proposal of Nanoarchaeota as a third archaeal phylum (Waters *et al.* 2003; Randau *et al.* 2005). This issue will be thoroughly discussed in a §9.

7. ORIGIN AND EVOLUTION OF METHANOGENESIS

Let us focus on the position of *M. kandleri*. Assessing correctly the true place of this species in the archaeal tree is in fact crucial to discussion on the origin of methanogenesis and the nature of the archaeal ancestor. Indeed, in contrast to its basal emergence in SSU rRNA trees, recent whole-genome trees have suggested the grouping of *M. kandleri* with other methanogens, away from the root (Slesarev *et al.* 2002). This placement is also suggested by our unrooted archaeal translation trees (figure 2*a*; Brochier *et al.* 2004, 2005*a*). As we have already suggested, *M. kandleri* is likely misplaced in our archaeal transcription trees (figure 2*b*; Brochier *et al.* 2004, 2005*a*) due to an LBA artefact between its relatively long branch and those leading to *N. equitans* and Crenarchaeaota (Brochier *et al.* 2004). Indeed, we noticed that the components of the transcriptional machinery show higher-than-average evolutionary rates in *M. kandleri* (Brochier *et al.* 2004). This fast evolutionary rate may be due to the fact that the RNAP subunit H of *M. kandleri* was replaced by an orthologue from Thermoplasmatales, and that the important transcription factor S (TFS) is missing in *M. kandleri* (Brochier *et al.* 2004). This lineage displays also other idiosyncrasies, such as the presence of split or fused genes such as those coding for reverse gyrase (Krah *et al.* 1996) and histone (Slesarev *et al.* 1998), and a large proportion of orphan genes or insertions of large foreign elements (Slesarev *et al.* 2002), suggesting that its whole genome has evolved more rapidly than the average. Given the small branch displayed by *M. kandleri* in translation trees (figure 2*b*; Brochier *et al.* 2004, 2005*a*) we believe that ribosomal proteins indicate the correct position for this archaeon (i.e. close to Methanococcales and Methanobacteriales). This is in agreement with a split of the gene coding for RNAP B subunit, as in other methanogens (Klenk *et al.* 1994; Brochier *et al.* 2004), and with the presence of pseudomurein, a character shared with Methanobacteriales (Konig *et al.* 1989). Interestingly, the basal positions of *M. kandleri* in the transcription tree and in the SSU rRNA tree are probably due to different artefacts: a fast-evolving rate of its RNAP subunits, and a high G + C content in its rRNA, a characteristic shared by all hyperthermophiles (Galtier & Lobry 1997), respectively. Indeed, *M. kandleri* exhibits a very short branch in the rRNA tree, but a very high G + C content of its rRNAs compared to its methanogen relatives. The SSU rRNA sequence of *M. kandleri* is likely attracted towards those of other hyperthermophiles due to convergent G or C positions. Consistently with this hypothesis, the analysis of a concatenation of 5S, 16S and 23S rRNA genes but using only transversions shows a grouping of *M. kandleri* with Methanobacteriales, as in the translation tree (data not shown).

The grouping of *M. kandleri* with other methanogens (Methanobacteriales and Methanococcales) in ribosomal protein and genomic gene content trees (Slesarev *et al.* 2002; Bapteste *et al.* 2005*a*) suggests that methanogenesis originated early in Euryarchaeota, but after the divergence of Thermococcales. Moreover, all methanogens (Methanococcales, Methanobacteriales, Methanomicrobiales, Methanosarcinales and Methanopyrales) share the same set of homologous enzymes and cofactors required for the central pathway of methanogenesis (the hydrogenotrophic pathway) (Bapteste *et al.* 2005*a*). A recent analysis has suggested that the genes coding for the enzymes involved in the hydrogenotrophic pathway and in the biosynthesis of coenzymes involved in methanogenesis were never exchanged between methanogens, implying that the methanogenic pathway originated only once in Euryarchaeota (Bapteste *et al.* 2005*a*). Moreover, this finding contradicts the general assumption that metabolic genes are more transferable than informational ones. Indeed, the all-in once acquisition via HGT of the whole pathway may have been prevented by the fact that these genes are interspersed in the genomes of methanogens and that the transfer of single genes apparently did not represent any selective advantage (i.e. no homologous replacements were observed; Bapteste *et al.* 2005*a*).

Interestingly, the fact that two major groups of methanogens, i.e. Methanopyrales, Methanococcales and Methanobacteriales (called Class I methanogens) (Bapteste *et al.* 2005*a*), and Methanomicrobiales and Methanosarcinales (called Class II methanogens) (Bapteste *et al.* 2005*a*) are separated by

non-methanogenic lineages (Thermoplasmatales, Archaeoglobales and Halobacteriales) in the archaeal phylogeny suggests that methanogenesis was lost at least three times in the evolution of Euryarchaeota. Indeed, a few enzymes of the hydrogenotrophic pathway are still found in the genome of *Archaeoglobus fulgidus*. Interestingly, *A. fulgidus* harbours the enzymes responsible for the first five steps of methanogenesis, but lacks those involved in the last two. Since the absence of 2-(methylthio)-ethanesulphonate (methyl-CoM) reductase in this archaeon eliminates the possibility of methane production by conventional pathways (Klenk *et al.* 1997), it was suggested that these five enzymes are likely involved in lactate oxidation (Vorholt *et al.* 1995) or reverse methanogenesis (Hallam *et al.* 2004).

The sudden appearance of the complete set of enzymes of the methanogenic pathway in the early evolution of Euryarchaeota is puzzling. Some hints may arise from the study of methanotrophs, the organisms able to oxidize methane and that form consortia with methanogens. Both archaeal and bacterial methanotrophs exist. Bacteria methanotrophs belong to α and γ proteobacteria and oxidize methane aerobically by a well-described pathway (Chistoserdova *et al.* 2005). On the contrary, methanotrophic Archaea are uncultivated anaerobes whose sequences are closely related to Methanomicrobiales (Schleper *et al.* 2005) and that are normally found in association with sulphate-reducing Bacteria in anoxic deep-sea sediments (Valentine 2002). The hypothesis that these Archaea may oxidize methane by using a *reverse* methanogenic pathway has been strengthened by recent environmental data (Hallam *et al.* 2004). This suggests that anaerobic methanotrophy originated in Archaea from the methanogenic pathway, although its relationship with the bacterial aerobic pathway is presently unknown. Interestingly, homologues of the three enzymes involved in the first steps of methanogenesis are used by Bacterial methanotrophs to oxidize methane, pointing to a possible common origin of the two pathways (Chistoserdova *et al.* 2004). These proteins were also identified in the genomes of Planctomycetales, where they are involved in formaldehyde detoxification (Chistoserdova *et al.* 2004), and we may call them MMF (for methanogenesis, methanotrophy, formaldehyde detoxification). A phylogenetic analysis of MMF proteins has recently put forward the possibility that bacterial homologues were not recruited by HGT from Archaea, but they were already present in the common ancestor of Archaea and Bacteria (Chistoserdova *et al.* 2004). If this is confirmed, methanogenesis and anaerobic methanotrophy in Archaea and aerobic methanotrophy in Bacteria would have originated independently from an ancient formaldehyde detoxification pathway present in their last common ancestor. In agreement with this hypothesis, formaldehyde is presumed to have been very abundant on early Earth (Arrhenius *et al.* 1994). An alternative but provocative hypothesis is that the last common archaeal ancestor might have been itself a methanogen, and that methanogenesis was lost in Crenarchaeota and independently in all non-methanogenic euryarchaeal lineages.

The origin and evolution of methanogenesis is an important issue that requires new analyses and more data. For instance, it will be especially interesting to include in phylogenetic analyses the sequences of MMF proteins that were recently identified from uncultivated archaeal methanotrophs (Hallam *et al.* 2004).

8. A HYPERTHERMOPHILIC LAST COMMON ARCHAEAL ANCESTOR?

The idea that the last common archaeal ancestor was an organism thriving at high temperatures arose very early from the abundance of hyperthermophiles (i.e. having an optimal growth temperature above 80 °C; Stetter 1989) in the archaeal domain and their early branching in rooted archaeal SSU rRNA trees (Woese 1987). However, it was soon understood that the high G+C content of rRNAs in hyperthermophiles reduces the sequence space that can be explored and produces short branches in phylogenetic trees, with homologous positions occupied by G or C bases being possibly due to convergence and not to common ancestry (Woese *et al.* 1991). Nevertheless, hyperthemophilic lineages occupy the most basal positions both in Crenarchaeota and Euryarchaeota also in archaeal protein-based trees (figure 2), supporting the hypothesis of a hyperthemophilic ancestor, if the root is placed in between these two phyla. This may be still due to a specific compositional amino acid bias artificially grouping sequences from hyperthermophilic species, but to our knowledge no systematic study has yet been performed to test this possibility. However, the hypothesis of a hyperthermophilic last common archaeal ancestor is also supported by the evolutionary history and distribution of reverse gyrase. Reverse gyrase is an atypical DNA topoisomerase that produces positive supercoiling into circular DNA *in vitro*, and is formed by the fusion of a classical type I DNA topoisomerase and of a large helicase domain (Declais *et al.* 2000). Although the precise role *in vivo* of this enzyme is still unclear, it is certainly essential for life at high temperature, since it is present in all currently sequences genomes from hyperthermophiles (Forterre 2002), and a *T. kodakarensis* reverse gyrase knock out mutant was recently shown to be unable to grow above 90 °C (Atomi *et al.* 2004). Two lines of evidence have suggested that reverse gyrase first originated in Archaea and was then transferred to Bacteria (Forterre 2002): (i) in a phylogenetic tree of reverse gyrase, bacterial sequences are interspersed within archaeal ones and (ii) the genomic context of reverse gyrase genes in bacterial genomes includes genes of archaeal origin. Moreover, the presence of this enzyme in the last common archaeal ancestor, and thus its hyperthermophilic nature, is consistent with the fact that a reverse gyrase tree containing only archaeal sequences is similar to the phylogeny based on r-protein concatenation (data not shown). However, it cannot be excluded that reverse gyrase may have originated either in Euryarchaeota or Crenarchaeota and been transferred very early between these two domains. Finally, the late emergence of hyperthermophilic Bacteria in very accurate SSU rRNA-based trees (Brochier & Philippe 2002) and

the higher proportion of genes of likely archaeal origin in the genomes of bacterial hyperthermophiles strengthens the hypothesis that survival at hot temperature is a secondary adaptation in Bacteria which was likely helped by an important number of HGT from hyperthermophilic Archaea (including reverse gyrase) (Deckert *et al.* 1998; Nelson *et al.* 1999; Koonin *et al.* 2001*a*).

Curiously, if the last common archaeal ancestor was a hyperthermophile, low temperature environments can be considered as extreme to Archaea. Indeed, adaptation to mesophilic environments in Archaea may have been favoured by HGT from Bacteria and from already adapted mesophilic archaeal lineages (Lopez-Garcia *et al.* 2004; Wiezer & Merkl 2005). The availability of more gene sequences from cold-adapted Crenarchaeota such as *Cenarchaeum symbiosum* and from Korarchaeota and the AAG group will be an invaluable tool to confirm the hypothesis of a hyperthermophilic archaeal ancestor.

9. THE POSITION OF NANOARCHAEA AND THE ROOT OF THE ARCHAEAL TREE

One of the most fascinating recent findings in archaeal research was the description of *Nanoarchaeum equitans*, a tiny hyperthermophile that lives in obligate symbiosis on the surface of *Ignicoccus*—a crenarchaeon—and holds the record of the smallest known living cell (with a volume equal to 1/100 of that of *Escherichia coli*) (Huber *et al.* 2002, 2003). Its very divergent SSU rRNA sequence, with many base changes even in the so-called 'highly conserved regions' that are usually employed as primer targets for SSU rDNA PCR, led to propose *N. equitans* as the first representative of a new archaeal phylum in addition to Crenarchaeota and Euryarchaeota, the Nanoarchaeota (Huber *et al.* 2002). As a consequence, idiosyncrasies observed in the genome of *N. equitans*, such as the presence of split reverse gyrase and tRNA genes, were interpreted as possible ancient traits (Waters *et al.* 2003; Randau *et al.* 2005). Consistently with its lifestyle, *N. equitans* harbours the smallest cellular genome sequenced so far (490 Mb) (Waters *et al.* 2003) and lacks one-third of the genes present in all other archaeal genomes (Makarova & Koonin 2005). The proposal of Nanoarcheaota was supported by the emergence of *N. equitans* prior to the Crenarchaeota/Euryarchaeota divergence in rooted archaeal trees from concatenated ribosomal protein datasets (Waters *et al.* 2003). In our previous unrooted phylogenetic trees based on concatenated datasets of r-proteins and RNAP subunits (Brochier *et al.* 2005*a*) and in the updated ribosomal protein concatenation tree (figure 2*a*), *N. equitans* does not emerge within Crenarchaeota or within Euryarchaeota. However, this cannot be taken as support for Nanoarchaeota, since this placement is also compatible with *Nanoarchaeum* being a basal crenarchaeal or euryarchaeal offshoot. Moreover, we recently showed that this placement is very likely biased by a fast evolutionary rate (Brochier *et al.* 2005*b*), a typical feature of reduced genomes. In fact, in individual ribosomal protein trees, *N. equitans* displays long branches and an unusually unstable placement, emerging very rarely as a separate lineage but rather within Euryarchaeota and less frequently within Crenarchaeota (Brochier *et al.* 2005*b*). Moreover, a tree constructed from the fusion of ribosomal proteins from the small subunit alone showed a weak grouping of *N. equitans* with Thermococcales, within Euryarchaeota (Brochier *et al.* 2005*b*). Importantly, this grouping was again found and strongly supported in the phylogenies of several informational proteins such as elongation factors, DNA topoisomerase VI and tyrosyl-tRNA synthetase (Brochier *et al.* 2005*b*; Moreira & Lopez-Garcia 2005, #2703). Although this may be explained by a specific trend of HGT from Thermococcales to *Nanoarchaeum*, such a trend would rather be expected to occur with the *Ignicoccus* host. Finally, *N. equitans* harbours a number of genes so far considered as distinctive of Euryarchaeota, such as the cell division proteins FtsZ and MinD, replication protein A (RPA), the two subunits of DNA polymerase of the delta family and a eukaryotic-like histone (Waters *et al.* 2003). Moreover, the genome of *N. equitans* encodes five of the nine proteins that form the euryarchaeal core and are not found in any crenarchaeal genome (Makarova & Koonin 2005). However, this may also be consistent with an early divergence of Nanoarchaea (for example, if all these genes were lost in Crenarchaeota).

From the bulk of all these evidences we think that *Nanoarchaeum* may represent a euryarchaeal fast-evolving lineage distantly related to modern Thermococcales that is misplaced in rooted and unrooted phylogenetic trees due to an LBA artefact. A second bias introduced by a possible HGT of some r-proteins from its crenarchaeal host may strengthen the attraction of *N. equitans* towards Crenarchaeota in the translation tree (figure 2*a* and Brochier *et al.* (2005*b*)). The sequencing of the *Ignicoccus* genome will provide important data to test this hypothesis and to get precious insights into the origin and evolution of this so far unique symbiosis between two archaeal species in hyperthermophilic environments.

10. EXPLORING THE ARCHAEAL PHYLOGENOMIC CORE: ADDITIONAL MOLECULAR SYSTEMS

We chose to analyse in a systematic way different archaeal molecular systems at one time. In fact, we are not fond of massive simultaneous analyses, whose results can be severely biased by undetected HGT. We reckon that it is indispensable to analyse each gene separately before analysis. Such an approach is surely time-consuming, but has the advantage to give more reliable results. This permits us to extract additional phylogenetic signal for the evolutionary history of the Archaea, but also to understand the evolutionary history of different molecular machineries. Here, we have analysed the phylogeny of two additional informational molecular systems, the signal recognition particle (SRP) and the exosome.

(a) *The signal recognition particle*

The SRP and its membrane-bound receptor (SR) deliver membrane proteins and secretory proteins to the translocation channel (translocon) in the plasma

membrane (or the endoplasmic reticulum in eukaryotic cells). SRP is a ribonucleoprotein complex that is made up of a 4.5S RNA moiety and a single protein (Srp54) in Bacteria, and of a 7S RNA moiety, Srp54, and five additional non-paralogous subunits (Srp72, Srp68, Srp19, Srp14, Srp9) in Eukarya. The eukaryotic SRP receptor is composed of two subunits, SRα and SRβ, while Bacteria have a receptor composed of a single protein (FtsY), which is homologous to SRα. The archaeal SRP is composed of a 7S RNA molecule remarkably similar in secondary structure to its eukaryotic counterparts, as well as two subunits homologous to the Srp54 and Srp19 subunits and a receptor homologue of FtsY/SRα. The general outline of the SRP pathway is conserved in all three kingdoms of life (Keenan *et al.* 2001). Eukaryal SRP recognizes and binds to the signal sequences of nascent polypeptides on ribosomes via its Srp54 subunit and causes an arrest of elongation. The ribosome–nascent polypeptide–SRP complex is translocated to the membrane of the endoplasmic reticulum where Srp54 contacts the SR receptor. A series of synchronized GTP hydrolysis events then regulates the release of SRP from the complex, leading to resumption of protein translation directly into the translocon pore. In Bacteria, SRP is responsible both for the insertion of some membrane proteins and for secretion. Recent biochemical studies have indicated that the archaeal SRP pathway has many similarities to the eukaryal and bacterial ones, but also has unique aspects (Zwieb & Eichler 2002).

Srp54 and FtsY/SRα are extremely well-conserved ancient paralogues that have allowed rooting the universal tree of life (Gribaldo & Cammarano 1998). We identified Srp54 and FtsY homologues in all completely sequenced archaeal genomes, with the notable exception of *N. equitans*. We found that the gene coding for the Srp19 subunit is missing in all complete archaeal genomes from Thermoplasmatales and Thermococcales, suggesting an ongoing streamlining of archaeal SRP. Comparison of the two individual unrooted archaeal phylogenies based on Srp54 and FtsY showed that branches appeared in general longer for FtsY than Srp54 (data not shown). The dichotomy Crenarchaeota/Euryarchaeota, the monophyly of the major archaeal groups and the relationships among crenarchaeal groups were recovered in both trees. However, basal nodes within Euryarchaeota were less robust. This is very likely due to the limited number of analysed positions available (392 and 249 positions for Srp54 and FtsY, respectively). However, no clear cases of HGT were evident from these analyses, indicating that the genes coding for these components of the SRP machinery may be new members of the archaeal phylogenomic core.

We built an unrooted archaeal phylogeny from a concatenation of the two Srp54 and FtsY datasets (641 positions, hereafter called 'SRP tree'). The phylogeny obtained (figure 3) supports most of the relationships indicated by the translation and transcription trees (figure 2*a*,*b*, respectively). In particular, the SRP tree strongly supports the Crenarchaeota/Euryarchaeota dichotomy (BV=99%) and the monophyly of major archaeal groups (Sulfolobales, Thermococcales, Methanococcales, Halobacteriales, Thermoplasmatales and Methanosarcinales, all with BV=100%) and their internal branching order (figure 3). It also supports the same order of divergence observed in the translation and transcription trees for the internal relationships in Crenarchaeota, and in Thermoplasmatales, Archaeoglobales, Halobacteriales and Methanobacteriales/Methanosarcinales within Euryarchaeota. In contrast, Thermococcales emerge immediately after *M. kandleri*, as in the transcription tree (figure 2*b*), but are sister group of the Methanobacteriales/Methanococcales clade (BV=88%). This discrepancy may result from a lower resolution power due to the limited number of analysed positions, although the hypothesis of an ancient HGT involving the genes coding Srp54 and FtsY components between an ancestor of Thermococcales and an ancestor of Methanobacteriales and Methanococcales cannot be excluded.

(b) *The exosome*

Aspects of RNA metabolism in Bacteria and Eukarya are well studied, but there is still only limited information about RNA processing in Archaea. In Bacteria, as well as in Eukarya, large protein complexes participate in RNA maturation and decay. In Bacteria, the endoribonuclease RNase E is the central component of a protein complex called the degradosome (Py *et al.* 1996; Jager *et al.* 2001). Eukaryotic cells harbour a conserved RNA processing and degrading protein complex called the exosome, which has a central function in the maturation of ribosomal RNA, small nucleolar RNA (snoRNA) and small nuclear RNA (snRNA), as well as in messenger RNA decay (Jacobs *et al.* 1998; Allmang *et al.* 1999*a*). The eukaryotic exosome has a central ring made up of six 3′–5′ exoribonucleases subunits that form two distinct paralogous groups (Rrp41, Rrp 46 and Mtr3 on one side and Rrp42, Rrp43 and Rrp45 on the other side) as well as a number of associated protein factors such as RNA-binding proteins and RNA helicases (Mitchell *et al.* 1997; Allmang *et al.* 1999*b*).

A gene context survey of completely sequenced archaeal genomes, complemented by sequence-profile analysis, recently suggested the existence of an archaeal counterpart of the eukaryotic exosome (Koonin *et al.* 2001*b*). This was confirmed by wet data, which identified the archaeal exosome as composed of a central ring hexamer with a 3′–5′ exonuclease activity and a number of peripheral proteins, as in Eukarya (Evguenieva-Hackenberg *et al.* 2003). In Archaea, only two exosome components are found—Rrp41 (homologous to eukaryotic subunits Rrp41, Rrp46 and Mtr3p) and Rrp42 (homologous to eukaryotic subunits Rrp42, Rrp43 and Rrp45p) (Koonin *et al.* 2001*b*), and the central ring is composed of three Rrp41–Rrp42 heterodimers. The activity of the complex resides within the active sites of the Rrp41 subunits, all three of which face the same side of the hexameric structure, whereas the Rrp42 subunits are inactive but contribute to the structuring of the Rrp41 active site (Lorentzen *et al.* 2005). Archaea possess a unique homologue of Rrp4p and Rrp40p, two paralogous hydrolytic RNAses/RNA binding proteins of the eukaryotic exosome, as well as a homologue of Cs14p, another exosome associated RNA binding protein (Koonin

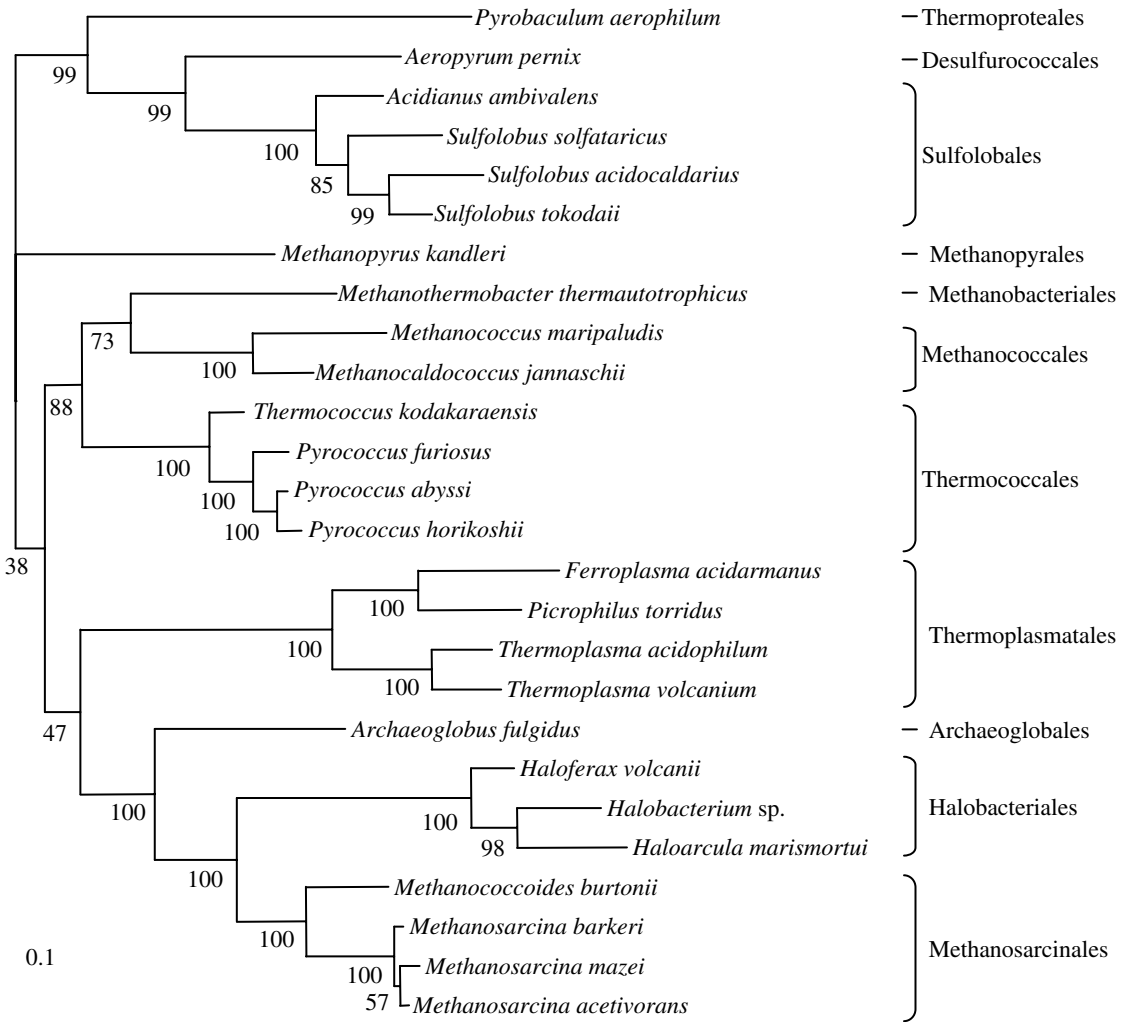

Figure 3. Unrooted ML tree of a concatenation of Srp54 and SRα/FtsY proteins (641 positions). Calculation was made by PHYML as described in the legend to figure 2.

et al. 2001*b*). The high sequence similarity of archaeal and eukaryotic exosome subunits, and their high structural similarity to Bacterial mRNA-degrading polyribonucleotide nucleotidyltransferase support a common strategy for RNA-degrading in all three domains of life (Lorentzen *et al.* 2005). All Archaea except Methanococcales and Halobacteriales encode highly conserved orthologues of Rrp41p, Rrp42p, Rrp4p and Cs14 (Koonin *et al.* 2001*b*). Interestingly, the genes coding for Rrp4 and Rrp42 subunits belong to a very conserved superoperon in Archaea, and appear to have been precisely excised from the superoperon in *Methanococcus jannaschii*, while in Halobacteriales this superoperon is divided into two predicted operons, with the same two exosome subunits missing (Koonin *et al.* 2001*b*). It will be interesting to know how Methanococcales and Halobacteriales cope with the absence of an exosome equivalent, since we could not find any homologues of the bacterial degradosome components in the complete genomes from the members of these archaeal lineages.

The phylogenetic analysis of the four individual archaeal exosome components (Rrp41p, Rrp42p, Rrp4p and Cs14, data not shown) revealed only one single possible HGT event involving Rrp42. In fact, in the Rrp42 tree, *N. equitans* emerged as sister group to Sulfolobales within the Crenarchaeota (not shown). As this may be a transfer from its host *Ignicoccus*, we removed the *N. equitans* sequence from the Rrp42 dataset. Moreover, no homologue of Cs14 could be found in the complete genome of *N. equitans*. The tree obtained from the concatenation of the Rrp41p, Rrp42p, Rrp4p and Cs14 datasets (1224 positions, hereafter called 'Exosome tree') is shown in figure 4. As in the transcription, translation and SRP trees, the exosome tree recovers the divide between Crenarchaeota and Euryarchaeota (BV = 97 and 47%, respectively) as well as the monophyly of major archaeal groups (Sulfolobales, Thermococcales, Thermoplasmatales and Methanosarcinales) and their internal branching order (figure 4). However, the relationships among Euryarchaeal lineages are weakly supported (all BV < 50%, except for the clustering of *Archaeoglobus* with Methanosarcinales, BV = 89%). As for the SRP tree, this weak resolution likely reflects the limited resolving power of the dataset resulting from the restricted number of analysed positions (1224), but also from the specific absence of two major groups (Methanococcales and Halobacteriales). Finally, the emergence of *N. equitans* on a branch separate from Crenarchaeota and Euryarchaeota is to be taken with caution given that only a half of positions was available

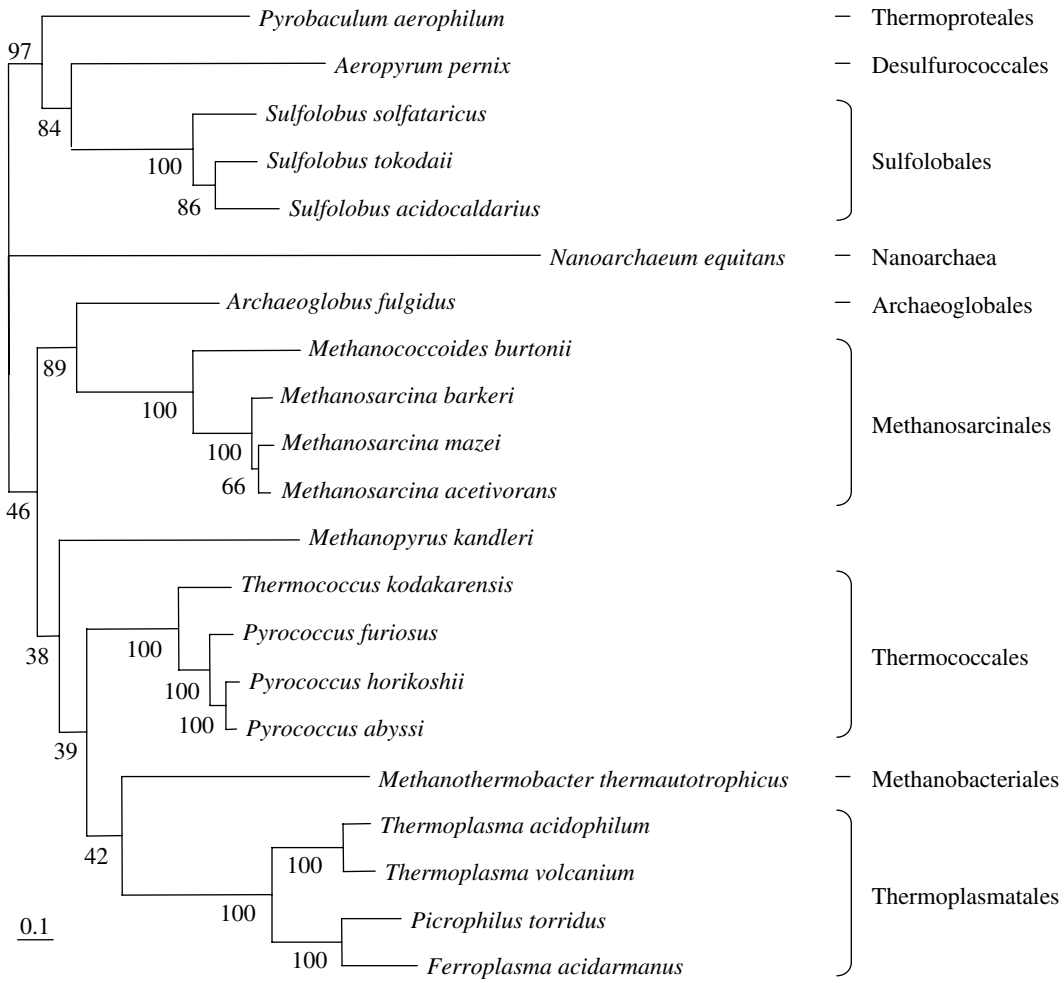

Figure 4. Unrooted ML tree of a concatenation of four exosome subunits (Rrp41, Rrp42, Rrp4 and Cs14, 1224 positions). Calculation was made by PHYML as described in the legend to figure 2.

for this species in the concatenated dataset (the Rrp42 sequence was removed, and the cs14 gene was missing).

The analysis of the archaeal exosome and SRP machineries expands the analysis of the evolution of molecular systems in Archaea. This confirms the rarity of HGT affecting the components of archaeal informational systems. The phylogenies obtained (especially those of SRP components) support several nodes in common with the trees based on transcription and translation datasets, indicating that most components of four informational machineries independently harbour a coherent signal for the phylogeny of the Archaea and are part of the archaeal phylogenomic core. This core of genes that globally supports a specific scenario for the history of Archaea is a good starting point for understanding the evolution of archaeal genes and genomes. This archaeal phylogenomic core should be thought as a 'soft' core of rarely transferred genes rather than as a 'hard' core of strictly congruent and never exchanged genes. Indeed, recent analyses of congruence between markers in their support of different test topologies, either by principal component analysis (PCA) (Brochier *et al.* 2002; Matte-Tailliez *et al.* 2002) or more recently by Heat Maps analysis (Bapteste *et al.* 2005*b*), showed that the majority of these markers did not support a single phylogeny while rejecting all others. However, this is not so unexpected given that the small size of most individual markers does not provide enough resolving power to discriminate between close topologies. Indeed, PCA analyses showed in fact that the discriminating power of a marker is strongly correlated to its size (Brochier *et al.* 2002; Matte-Tailliez *et al.* 2002). Moreover, although for some incongruent markers examination of the corresponding phylogenies revealed that incongruence was clearly due to HGTs, for other markers tree reconstruction artefacts due to rapid evolutionary rates and/or compositional biases could give a better explanation for incongruence (Matte-Tailliez *et al.* 2002, #2570). Interestingly, both PCA and Heat Maps analyses identified overlapping sets of incongruent markers when similar datasets were used (for example, both studies identified ribosomal proteins Rpl7ae and Rpl15e as incongruent markers (Matte--Tailliez *et al.* 2002; Bapteste *et al.* 2005*b*)). Further comparative analyses between the two methods would be of great interest. In particular, refinement will be needed for the choice of the best set of topologies to be used to test congruence between markers.

Figure 5 shows what we think may be the best current picture of the evolutionary history of the archaeal domain based on this soft phylogenomic core and on the critical analysis of the placements of *M. kandleri* and *N. equitans*, although these are not yet definitive. The addition of more sequences from new

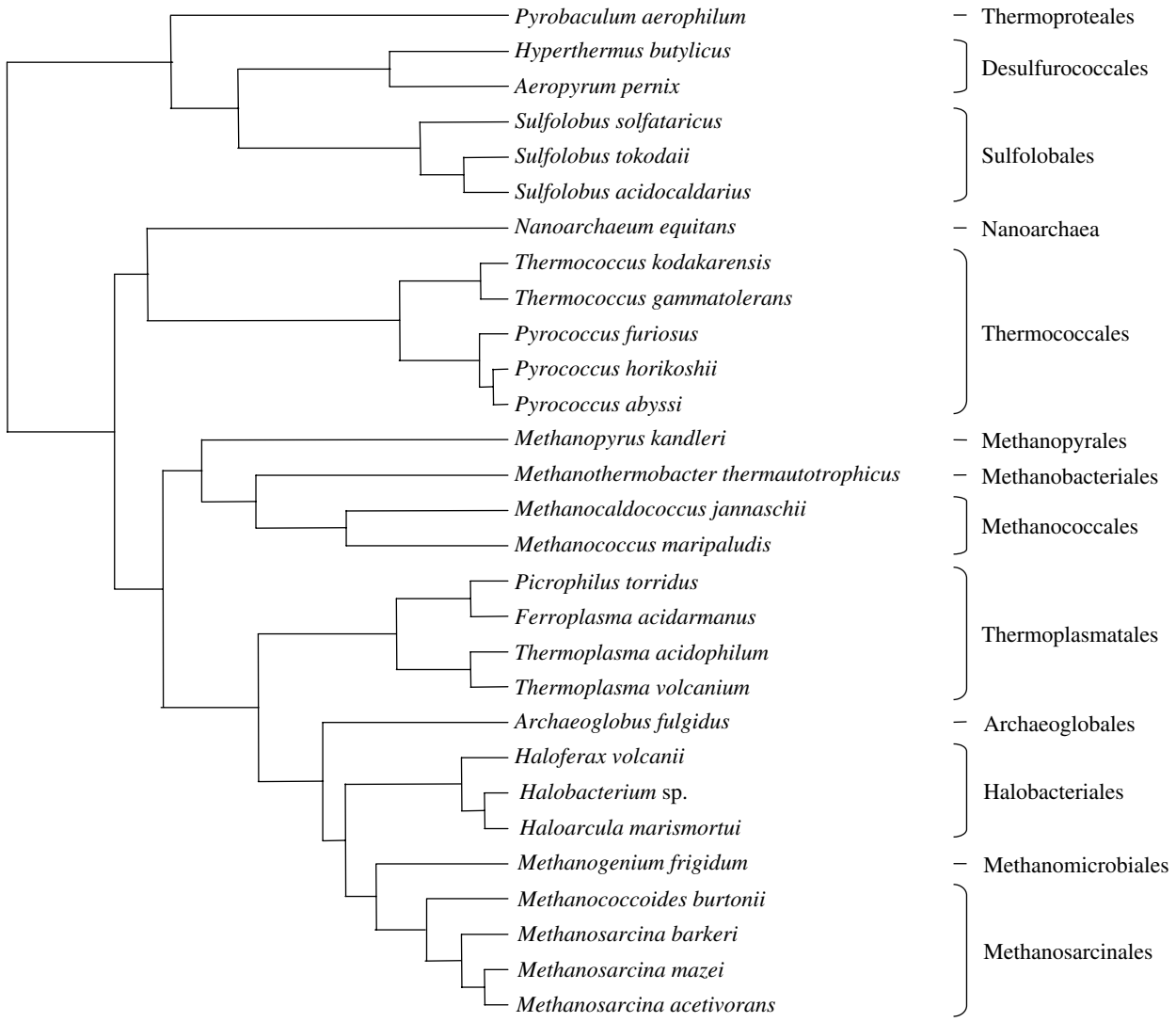

Figure 5. A consensual phylogeny of the Archaea for which complete genome sequences are available, issued from current phylogenomics evidence.

complete genome sequences will surely help refining this tree of the Archaea, by breaking a few long branches (such as those of *N. equitans* or *M. kandleri*), and by improving the detection of HGT and other possible biases. At the same time, the exploration of additional informational as well as operational molecular systems will help refining or confirming currently weakly supported nodes (as the relationships among the lineages belonging to Class I methanogens), or correct some presently unnoticed mistakes. Moreover, the inclusion of novel representatives of known phyla (notably Crenarchaeota) as well as of presently uncultivated groups will expand this picture of archaeal evolution across a more complete sampling of the diversity of this domain of life.

11. DID ARCHAEA EVOLVE DIFFERENTLY?

The possibility of retracing in a rather robust way the divergence between the major archaeal lineages by molecular data is striking when compared to the much higher difficulty to do so for Bacteria, where the order of divergence of phyla is largely unresolved in molecular phylogenies, even with very accurate analyses (Snel *et al.* 1999; Daubin & Gouy 2001; Brochier *et al.* 2002; Daubin *et al.* 2002; Gophna *et al.* 2005). This may simply reflect a partial sampling of archaeal diversity that is currently available for molecular phylogeny reconstruction. In this case, the picture will get more and more blurred as more species are added. However, for the time being this does not seem to be the case, since an increase in taxonomic sampling increases the robustness of archaeal trees based on protein concatenation (Brochier *et al.* 2005*a*). Moreover, the analysis of concatenated ribosomal protein datasets leads to a robust phylogeny for Archaea (Matte-Tailliez *et al.* 2002), but a poorly resolved one for Bacteria (Brochier *et al.* 2002), despite a similar number of lineages examined (13 and 14, respectively). An alternative explanation may thus be that present-day archaeal lineages are younger than bacterial ones, and/or that they did not diverge as rapidly as may have been the case for their prokaryotic cousins.

The statistical confidence associated with the Crenarchaeota/Euryarchaeota divide by virtually all molecular markers is striking and remains unaffected even when additional species are added. The likely misassignment of Nanoarchaeota to a third archaeal phylum is emblematic in this respect. The genomic sequence of representatives of Korarchaeota and the

AAG group will provide valuable data to test the robustness of the Crenarchaeota/Euryarchaeota divide. The profound divergence between Crenarchaeota and Euryarchaeota is also strongly supported by comparative genomics, since a number of genes present in euryarchaeal genomes are missing altogether in crenarchaeal ones, and vice versa (Makarova & Koonin 2005). For example, DNA polymerases of the delta family, eukaryotic-like histones, the replication protein RPA and the cell division proteins FtsZ and MinD appear to be an exclusivity of Euryarchaeota (Uemori *et al.* 1997; Myllykallio *et al.* 2000; Bell & Jackson 2001) (the recent finding of a histone gene from the crenarchaeon *C. symbiosum* (Cubonova *et al.* 2005)) needs further analysis to exclude a transfer from Euryarchaeota). These differences are not trivial, and suggest the use of different molecular strategies for key cellular processes—such as maintenance of chromosome structure, replication, and division—that may have driven the divergence of these two archaeal lineages.

Such a dramatic evolutionary split does not appear to have any equivalent in Bacteria or in Eukarya, and may be more profound than that separating the different bacterial or eukaryotic phyla. Indeed, in the early 1980s, Lake proposed to divide the archaebacteria in two domains on the base of ribosome shape: the Eocytes (namely the Crenarchaeota) and the Archaebacteria (namely the Euryarchaeota) (Henderson *et al.* 1984; Lake *et al.* 1984). The first domain was proposed to be more closely related to Eukarya while the second to be closer to Bacteria (i.e. Archaea would not be monophyletic but paraphyletic) (Henderson *et al.* 1984; Lake *et al.* 1984). However, molecular data brings no support for such a hypothesis, since Crenarchaeota and Euryarchaeota are sister groups in universal phylogenetic trees, be they based on single molecular markers, paralogous couples, or whole genomic data (Woese 1987; Snel *et al.* 1999; Gribaldo & Philippe 2002; Gophna *et al.* 2005). However, the observation of Lake underlined the fact that the differences between Crenarchaeota and Euryarchaeota appear more profound that the ones usually observed at the phylum level. If the division between Crenarchaeota and Euryarchaeota stands the test of time, it may then be more appropriate to consider them as sub-domains rather than phyla (Boone & Castenholz 2001). Beyond purely taxonomic nomenclature issues, this would give a better appreciation of the diversity of archaeal cultivated and non-cultivated lineages, which is comparable to that observed for Bacteria.

We wish to thank the Royal Society for inviting us to submit this manuscript, Patrick Forterre for sharing with us his vast knowledge on Archaea, Purification Lopez-Garcia and David Moreira for insightful discussions, Christa Schleper for kindly providing the tree used in figure 1, and the two anonymous referees for precious comments and advice. All datasets and trees are available on request from S.G. and C.B.A.

REFERENCES

Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E. & Tollervey, D. 1999*a* Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J.* **18**, 5399–5410. (doi:10.1093/emboj/18.19.5399)

Allmang, C., Petfalski, E., Podtelejnikov, A., Mann, M., Tollervey, D. & Mitchell, P. 1999*b* The yeast exosome and human PM–Scl are related complexes of $3' \rightarrow 5'$ exonucleases. *Genes Dev.* **13**, 2148–2158.

Arrhenius, T., Arrhenius, G. & Paplawsky, W. 1994 Archean geochemistry of formaldehyde and cyanide and the oligomerization of cyanohydrin. *Orig. Life Evol. Biosph.* **24**, 1–17. (doi:10.1007/BF01582036)

Atomi, H., Matsumi, R. & Imanaka, T. 2004 Reverse gyrase is not a prerequisite for hyperthermophilic life. *J. Bacteriol.* **186**, 4829–4833. (doi:10.1128/JB.186.14.4829-4833.2004)

Bapteste, E. & Brochier, C. 2004 On the conceptual difficulties in rooting the tree of life. *Trends Microbiol.* **12**, 9–13. (doi:10.1016/j.tim.2003.11.002)

Bapteste, E., Boucher, Y., Leigh, J. & Doolittle, W. F. 2004 Phylogenetic reconstruction and lateral gene transfer. *Trends Microbiol.* **12**, 406–411. (doi:10.1016/j.tim.2004.07.002)

Bapteste, E., Brochier, C. & Boucher, Y. 2005*a* Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. *Archaea* **1**, 353–363.

Bapteste, E., Susko, E., Leigh, J., MacLeod, D., Charlebois, R. L. & Doolittle, W. F. 2005*b* Do orthologous gene phylogenies really support tree-thinking? *BMC Evol. Biol.* **5**, 33. (doi:10.1186/1471-2148-5-33)

Barns, S. M., Delwiche, C. F., Palmer, J. D. & Pace, N. R. 1996 Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc. Natl Acad. Sci. USA* **93**, 9188–9193. (doi:10.1073/pnas.93.17.9188)

Battistuzzi, F. U., Feijao, A. & Hedges, S. B. 2004 A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol. Biol.* **4**, 44. (doi:10.1186/1471-2148-4-44)

Bell, S. D. & Jackson, S. P. 2001 Mechanism and regulation of transcription in archaea. *Curr. Opin. Microbiol.* **4**, 208–213. (doi:10.1016/S1369-5274(00)00190-9)

Bell, S. D., Botting, C. H., Wardleworth, B. N., Jackson, S. P. & White, M. F. 2002 The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation. *Science* **296**, 148–151. (doi:10.1126/science.1070506)

Boone, D. R. & Castenholz, R. W. 2001 *Bergey's manual of systematic bacteriology*. New York: Springer.

Brochier, C. & Philippe, H. 2002 A non-hyperthermophilic ancestor for bacteria. *Nature* **417**, 244. (doi:10.1038/417244a)

Brochier, C., Bapteste, E., Moreira, D. & Philippe, H. 2002 Eubacterial phylogeny based on translational apparatus proteins. *Trends Genet.* **18**, 1–5. (doi:10.1016/S0168-9525(01)02522-7)

Brochier, C., Forterre, P. & Gribaldo, S. 2004 Archaeal phylogeny based on proteins of the transcription and translation machineries: tackling the *Methanopyrus kandleri* paradox. *Genome Biol.* **5**, R17. (doi:10.1186/gb-2004-5-3-r17)

Brochier, C., Forterre, P. & Gribaldo, S. 2005*a* An emerging phylogenetic core of Archaea: phylogenies of transcription and translation machineries converge following addition of new genome sequences. *BMC Evol. Biol.* **5**, 36. (doi:10.1186/1471-2148-5-36)

Brochier, C., Gribaldo, S., Zivanovic, Y., Confalonieri, F. & Forterre, P. 2005*b* Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to Thermococcales? *Genome Biol.* **6**, R42. (doi:10.1186/gb-2005-6-5-r42)

Brocks, J. J., Logan, G. A., Buick, R. & Summons, R. E. 1999 Archean molecular fossils and the early rise of eukaryotes. *Science* **285**, 1033–1036. (doi:10.1126/science.285.5430.1033)

Brown, J. R. & Doolittle, W. F. 1995 Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. *Proc. Natl Acad. Sci. USA* **92**, 2441–2445.

Burggraf, S., Stetter, K. O., Rouviere, P. & Woese, C. R. 1991 *Methanopyrus kandleri*: an archaeal methanogen unrelated to all other known methanogens. *Syst. Appl. Microbiol.* **14**, 346–351.

Cavalier-Smith, T. 2002 The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int. J. Syst. Evol. Microbiol.* **52**, 7–76.

Charlebois, R. L. & Doolittle, W. F. 2004 Computing prokaryotic gene ubiquity: rescuing the core from extinction. *Genome Res.* **14**, 2469–2477. (doi:10.1101/gr.3024704)

Chistoserdova, L., Jenkins, C., Kalyuzhnaya, M. G., Marx, C. J., Lapidus, A., Vorholt, J. A., Staley, J. T. & Lidstrom, M. E. 2004 The enigmatic planctomycetes may hold a key to the origins of methanogenesis and methylotrophy. *Mol. Biol. Evol.* **21**, 1234–1241. (doi:10.1093/molbev/msh113)

Chistoserdova, L., Vorholt, J. A. & Lidstrom, M. E. 2005 A genomic view of methane oxidation by aerobic bacteria and anaerobic archaea. *Genome Biol.* **6**, 208. (doi:10.1186/gb-2005-6-2-208)

Cubonova, L., Sandman, K., Hallam, S. J., Delong, E. F. & Reeve, J. N. 2005 Histones in crenarchaea. *J. Bacteriol.* **187**, 5482–5485. (doi:10.1128/JB.187.15.5482-5485.2005)

Daubin, V. & Gouy, M. 2001 Bacterial molecular phylogeny using supertree approach. *Genome Inform. Ser. Workshop Genome Inform.* **12**, 155–164.

Daubin, V., Gouy, M. & Perriere, G. 2002 A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome Res.* **12**, 1080–1090. (doi:10.1101/gr.187002)

Daubin, V., Moran, N. A. & Ochman, H. 2003 Phylogenetics and the cohesion of bacterial genomes. *Science* **301**, 829–832. (doi:10.1126/science.1086568)

Deckert, G. *et al*. 1998 The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392**, 353–358. (doi:10.1038/32831)

Declais, A. C., Marsault, J., Confalonieri, F., de La Tour, C. B. & Duguet, M. 2000 Reverse gyrase, the two domains intimately cooperate to promote positive supercoiling. *J. Biol. Chem.* **275**, 19 498–19 504. (doi:10.1074/jbc.M910091199)

Delsuc, F., Brinkmann, H. & Philippe, H. 2005 Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* **6**, 361–375. (doi:10.1038/nrg1603)

Evguenieva-Hackenberg, E., Walter, P., Hochleitner, E., Lottspeich, F. & Klug, G. 2003 An exosome-like complex in *Sulfolobus solfataricus*. *EMBO Rep.* **4**, 889–893. (doi:10.1038/sj.embor.embor929)

Forterre, P. 2002 A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. *Trends Genet.* **18**, 236–237. (doi:10.1016/S0168-9525(02)02650-1)

Forterre, P. & Philippe, H. 1999*a* The last universal common ancestor (LUCA), simple or complex? *Biol. Bull.* **196**, 373–375. (discussion 375-7)

Forterre, P. & Philippe, H. 1999*b* Where is the root of the universal tree of life? *Bioessays* **21**, 871–879. (doi:10.1002/(SICI)1521-1878(199910)21:10<871::AID-BIES10>3.0.CO;2-Q)

Futterer, O., Angelov, A., Liesegang, H., Gottschalk, G., Schleper, C., Schepers, B., Dock, C., Antranikian, G. & Liebl, W. 2004 Genome sequence of *Picrophilus torridus* and its implications for life around pH 0. *Proc. Natl Acad. Sci. USA* **101**, 9091–9096. (doi:10.1073/pnas.0401356101)

Galtier, N. & Lobry, J. R. 1997 Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. *J. Mol. Evol.* **44**, 632–636. (doi:10.1007/PL00006186)

Gogarten, J. P. *et al*. 1989 Evolution of the vacuolar H+-ATPase: implications for the origin of eukaryotes. *Proc. Natl Acad. Sci. USA* **86**, 6661–6665.

Gophna, U., Doolittle, W. F. & Charlebois, R. L. 2005 Weighted genome trees: refinements and applications. *J. Bacteriol.* **187**, 1305–1316. (doi:10.1128/JB.187.4.1305-1316.2005)

Grabowski, B. & Kelman, Z. 2003 Archeal DNA replication: eukaryal proteins in a bacterial context. *Annu. Rev. Microbiol.* **57**, 487–516. (doi:10.1146/annurev.micro.57.030502.090709)

Gribaldo, S. & Cammarano, P. 1998 The root of the universal tree of life inferred from anciently duplicated genes encoding components of the protein-targeting machinery. *J. Mol. Evol.* **47**, 508–516. (doi:10.1007/PL00006407)

Gribaldo, S. & Philippe, H. 2002 Ancient phylogenetic relationships. *Theor. Popul. Biol.* **61**, 391–408. (doi:10.1006/tpbi.2002.1593)

Guindon, S. & Gascuel, O. 2003 A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704. (doi:10.1080/10635150390235520)

Gupta, R. S. 1998 What are archaebacteria: life's third domain or monoderm prokaryotes related to Gram-positive bacteria? A new proposal for the classification of prokaryotic organisms. *Mol. Microbiol.* **229**, 695–708. (doi:10.1046/j.1365-2958.1998.00978.x)

Hallam, S. J., Putnam, N., Preston, C. M., Detter, J. C., Rokhsar, D., Richardson, P. M. & DeLong, E. F. 2004 Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* **305**, 1457–1462. (doi:10.1126/science.1100025)

Hayes, J. M. 1994 In *Early life on Earth* (ed. S. Bengtson, J. Bergstrom, V. Gonzalo & A. Knoll) Nobel Symposium, pp. 220–239. New York: Columbia University Press.

Henderson, E., Oakes, M., Clark, M. W., Lake, J. A., Matheson, A. T. & Zillig, W. 1984 A new ribosome structure. *Science* **225**, 510–512.

Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C. & Stetter, K. O. 2002 A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**, 63–67. (doi:10.1038/417063a)

Huber, H., Hohn, M. J., Stetter, K. O. & Rachel, R. 2003 The phylum Nanoarchaeota: present knowledge and future perspectives of a unique form of life. *Res. Microbiol.* **154**, 165–171. (doi:10.1016/S0923-2508(03)00035-4)

Hugenholtz, P. 2002 Exploring prokaryotic diversity in the genomic era. *Genome Biol.* **3**. (doi:10.1186/gb-2002-3-2-reviews0003) (REVIEWS0003)

Hugenholtz, P., Goebel, B. M. & Pace, N. R. 1998 Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* **180**, 4765–4774.

Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. & Miyata, T. 1989 Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl Acad. Sci. USA* **86**, 9355–9359.

Jacobs, J. S., Anderson, A. R. & Parker, R. P. 1998 The 3′ to 5′ degradation of yeast mRNA is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3′ to 5′ exonucleases of the exosome complex. *EMBO J.* **17**, 1497–1506. (doi:10.1093/emboj/17.5.1497)

Jager, S., Fuhrmann, O., Heck, C., Hebermehl, M., Schiltz, E., Rauhut, R. & Klug, G. 2001 An mRNA degrading complex in *Rhodobacter capsulatus*. *Nucleic Acids Res.* **29**, 4581–4588. (doi:10.1093/nar/29.22.4581)

Karner, M. B., DeLong, E. F. & Karl, D. M. 2001 Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**, 507–510. (doi:10.1038/35054051)

Kates, M. 1993 In *The biochemistry of Archaea (Archaebacteria)* (ed. M. Kates, D. J. Kushner & A. T. Matheson), pp. 261–295. Amsterdam: Elsevier.

Keenan, R. J., Freymann, D. M., Stroud, R. M. & Walter, P. 2001 The signal recognition particle. *Annu. Rev. Biochem.* **70**, 755–775. (doi:10.1146/annurev.biochem.70.1.755)

Kennedy, S. P., Ng, W. V., Salzberg, S. L., Hood, L. & DasSarma, S. 2001 Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Res.* **11**, 1641–1650. (doi:10.1101/gr.190201)

Klenk, H. P., Palm, P. & Zillig, W. 1994 DNA-dependent RNA polymerases as phylogenetic marker molecules. *Syst. Appl. Microbiol.* **16**, 638–647.

Klenk, H. P. *et al*. 1997 The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390**, 364–370. (doi:10.1038/37052)

Konig, H., Kandler, O. & Hammes, W. 1989 Biosynthesis of pseudomurein: isolation of putative precursors from *Methanobacterium thermoautotrophicum*. *Can. J. Microbiol.* **35**, 176–181.

Koonin, E. V., Makarova, K. S. & Aravind, L. 2001*a* Horizontal gene transfer in prokaryotes: quantification and classification. *Annu. Rev. Microbiol.* **55**, 709–742. (doi:10.1146/annurev.micro.55.1.709)

Koonin, E. V., Wolf, Y. I. & Aravind, L. 2001*b* Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. *Genome Res.* **11**, 240–252. (doi:10.1101/gr.162001)

Krah, R., Kozyavkin, S. A., Slesarev, A. I. & Gellert, M. 1996 A two-subunit type I DNA topoisomerase (reverse gyrase) from an extreme hyperthermophile. *Proc. Natl Acad. Sci. USA* **93**, 106–110. (doi:10.1073/pnas.93.1.106)

Lake, J. A., Henderson, E., Oakes, M. & Clark, M. W. 1984 Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proc. Natl Acad. Sci. USA* **81**, 3786–3790.

Lawson, F. S., Charlebois, R. L. & Dillon, J. A. 1996 Phylogenetic analysis of carbamoylphosphate synthetase genes: complex evolutionary history includes an internal duplication within a gene which can root the tree of life. *Mol. Biol. Evol.* **13**, 970–977.

Lecompte, O., Ripp, R., Thierry, J. C., Moras, D. & Poch, O. 2002 Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res.* **30**, 5382–5390. (doi:10.1093/nar/gkf693)

Londei, P. 2005 Evolution of translational initiation: new insights from the archaea. *FEMS Microbiol. Rev.* **29**, 185–200. (doi:10.1016/j.femsre.2004.10.002)

Lopez-Garcia, P. & Moreira, D. 1999 Metabolic symbiosis at the origin of eukaryotes. *Trends Biochem. Sci.* **24**, 88–93.

Lopez-Garcia, P., Lopez-Lopez, A., Moreira, D. & Rodriguez-Valera, F. 2001 Diversity of free-living prokaryotes from a deep-sea site at the Antarctic Polar Front. *FEMS Microbiol. Ecol.* **36**, 193–202.

Lopez-Garcia, P., Brochier, C., Moreira, D. & Rodriguez-Valera, F. 2004 Comparative analysis of a genome fragment of an uncultivated mesopelagic crenarchaeote reveals multiple horizontal gene transfers. *Environ. Microbiol.* **6**, 19–34. (doi:10.1046/j.1462-2920.2003.00533.x)

Lorentzen, E., Walter, P., Fribourg, S., Evguenieva-Hackenberg, E., Klug, G. & Conti, E. 2005 The archaeal exosome core is a hexameric ring structure with three catalytic subunits. *Nat. Struct. Mol. Biol.* **12**, 575–581. (doi:10.1038/nsmb952)

Makarova, K. S. & Koonin, E. V. 2005 Evolutionary and functional genomics of the Archaea. *Curr. Opin. Microbiol.* **8**, 586–594. (doi:10.1016/j.mib.2005.08.003)

Matte-Tailliez, O., Brochier, C., Forterre, P. & Philippe, H. 2002 Archaeal phylogeny based on ribosomal proteins. *Mol. Biol. Evol.* **19**, 631–639.

Michaelis, W. *et al*. 2002 Microbial reefs in the Black Sea fueled by anaerobic oxidation of methane. *Science* **297**, 1013–1015. (doi:10.1126/science.1072502)

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. & Tollervey, D. 1997 The exosome: a conserved eukaryotic RNA processing complex containing multiple $3'\rightarrow5'$ exoribonucleases. *Cell* **91**, 457–466. (doi:10.1016/S0092-8674(00)80432-8)

Moreira, D. & Lopez-Garcia, P. 2005 Comment on "The 1.2-megabase genome sequence of Mimivirus". *Science* **308**, 1114.

Myllykallio, H., Lopez, P., Lopez-Garcia, P., Heilig, R., Saurin, W., Zivanovic, Y., Philippe, H. & Forterre, P. 2000 Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* **288**, 2212–2215. (doi:10.1126/science.288.5474.2212)

Nelson, K. E. *et al*. 1999 Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323–329. (doi:10.1038/20601)

Ochman, H., Lerat, E. & Daubin, V. 2005 Examining bacterial species under the specter of gene transfer and exchange. *Proc. Natl Acad. Sci. USA* **102**(Suppl. 1), 6595–6599. (doi:10.1073/pnas.0502035102)

Olsen, G. J. & Woese, C. R. 1996 Lessons from an Archaeal genome: what are we learning from *Methanococcus jannaschii*? *Trends Genet.* **12**, 377–379. (doi:10.1016/0168-9525(96)30092-9)

Pereto, J., Lopez-Garcia, P. & Moreira, D. 2004 Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem. Sci.* **29**, 469–477. (doi:10.1016/j.tibs.2004.07.002)

Py, B., Higgins, C. F., Krisch, H. M. & Carpousis, A. J. 1996 A DEAD-box RNA helicase in the *Escherichia coli* RNA degradosome. *Nature* **381**, 169–172. (doi:10.1038/381169a0)

Randau, L., Pearson, M. & Soll, D. 2005 The complete set of tRNA species in *Nanoarchaeum equitans*. *FEBS Lett.* **579**, 2945–2947. (doi:10.1016/j.febslet.2005.04.051)

Rivera, M. C. & Lake, J. A. 2004 The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* **431**, 152–155. (doi:10.1038/nature02848)

Robinson, N. P. & Bell, S. D. 2005 Origins of DNA replication in the three domains of life. *FEBS J.* **272**, 3757–3766. (doi:10.1111/j.1742-4658.2005.04768.x)

Rothschild, L. J. & Mancinelli, R. L. 2001 Life in extreme environments. *Nature* **409**, 1092–1101. (doi:10.1038/35059215)

Ruepp, A. *et al*. 2000 The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature* **407**, 508–513. (doi:10.1038/35035069)

Schleper, C., Jurgens, G. & Jonuscheit, M. 2005 Genomic studies of uncultivated archaea. *Nat. Rev. Microbiol.* **3**, 479–488. (doi:10.1038/nrmicro1159)

Shen, Y., Buick, R. & Canfield, D. E. 2001 Isotopic evidence for microbial sulphate reduction in the early Archaean era. *Nature* **410**, 77–81. (doi:10.1038/35065071)

Slesarev, A. I., Belova, G. I., Kozyavkin, S. A. & Lake, J. A. 1998 Evidence for an early prokaryotic origin of histones H2A and H4 prior to the emergence of eukaryotes. *Nucleic Acids Res.* **26**, 427–430. (doi:10.1093/nar/26.2.427)

Slesarev, A. I. *et al.* 2002 The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. *Proc. Natl Acad. Sci. USA* **99**, 4644–4649. (doi:10.1073/pnas.032671499)

Snel, B., Bork, P. & Huynen, M. A. 1999 Genome phylogeny based on gene content. *Nat. Genet.* **21**, 108–110. (doi:10.1038/5052)

Stetter, K. O. 1989 In *Autotrophic bacteria* (ed. H. G. Schlegel & B. Bowien), pp. 167–171. Madison, WI/Berlin: Science Tech Publishers/Springer.

Summons, R. 1999 Molecular probing of deep secrets. *Nature* **398**, 752–753. (doi:10.1038/19630)

Summons, R. E., Powell, T. G. & Boreham, C. J. 1988 Petroleum geology and geochemistry of the Middle Proterozoic McArthur Basin, northern Australia. *Geochim. Cosmochim. Acta* **52**, 1747–1763. (doi:10.1016/0016-7037(88)90001-4)

Takai, K. & Horikoshi, K. 1999 Genetic diversity of archaea in deep-sea hydrothermal vent environments. *Genetics* **152**, 1285–1297.

Uemori, T., Sato, Y., Kato, I., Doi, H. & Ishino, Y. 1997 A novel DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus*: gene cloning, expression, and characterization. *Genes Cells* **2**, 499–512. (doi:10.1046/j.1365-2443.1997.1380336.x)

Valentine, D. L. 2002 Biogeochemistry and microbial ecology of methane oxidation in anoxic environments: a review. *Antonie Van Leeuwenhoek* **81**, 271–282.

Vorholt, J. A., Kunow, J., Stetter, K. O. & Thauer, R. K. 1995 Enzymes and coenzymes of the carbon monoxide dehydrogenase pathway for autotrophic CO_2 fixation in *Archaeoglobus fulgidus* and the lack of carbon monoxide dehydrogenase in the heterotrophic *A. profundus*. *Arch. Microbiol.* **163**, 112–118. (doi:10.1007/s002030050179)

Walsh, D. A. & Doolittle, W. F. 2005 The real 'domains' of life. *Curr. Biol.* **15**, R237–R240. (doi:10.1016/j.cub.2005.03.034)

Waters, E. *et al.* 2003 The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc. Natl Acad. Sci. USA* **100**, 12 984–12 988. (doi:10.1073/pnas.1735403100)

Wiezer, A. & Merkl, R. 2005 A comparative categorization of gene flux in diverse microbial species. *Genomics* **86**, 462–475. (doi:10.1016/j.ygeno.2005.05.014)

Woese, C. R. 1987 Bacterial evolution. *Microbiol. Rev.* **51**, 221–271.

Woese, C. R., Kandler, O. & Wheelis, M. L. 1990 Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl Acad. Sci. USA* **87**, 4576–4579.

Woese, C. R., Achenbach, L., Rouviere, P. & Mandelco, L. 1991 Archaeal phylogeny: reexamination of the phylogenetic position of *Archaeoglobus fulgidus* in light of certain composition-induced artifacts. *Syst. Appl. Microbiol.* **14**, 364–371.

Wolf, Y. I., Rogozin, I. B., Grishin, N. V. & Koonin, E. V. 2002 Genome trees and the tree of life. *Trends Genet.* **18**, 472–479. (doi:10.1016/S0168-9525(02)02744-0)

Zwieb, C. & Eichler, J. 2002 Getting on target: the archaeal signal recognition particle. *Archaea* **1**, 27–34.