Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota

Céline Brochier-Armanet, Bastien Boussau‡, Simonetta Gribaldo§ and Patrick Forterre§ ||*

Abstract | The archaeal domain is currently divided into two major phyla, the Euryarchaeota and Crenarchaeota. During the past few years, diverse groups of uncultivated mesophilic archaea have been discovered and affiliated with the Crenarchaeota. It was recently recognized that these archaea have a major role in geochemical cycles. Based on the first genome sequence of a crenarchaeote, *Cenarchaeum symbiosum*, we show that these mesophilic archaea are different from hyperthermophilic Crenarchaeota and branch deeper than was previously assumed. Our results indicate that *C. symbiosum* and its relatives are not Crenarchaeota, but should be considered as a third archaeal phylum, which we propose to name Thaumarchaeota (from the Greek '*thaumas*', meaning wonder).

Hyperthermophile
An organism that has an optimal growth temperature of at least 80°C.

Paraphyletic
A group of organisms or sequences that includes an ancestor and some, but not all, of its descendants.

**Université de Provence, Aix-Marseille I, CNRS, UPR 9043, Laboratoire de Chimie Bactérienne, Institut de Biologie Structurale et de Microbiologie, 13402 Marseille, France.*
‡Université de Lyon, Université Lyon 1, CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, 69622 Villeurbanne, France.
§Biologie Moléculaire du Gène chez les Extrêmophiles (BMGE), Département de Microbiologie, Institut Pasteur, 75015 Paris, France.
||Université Paris-Sud, 91405 Orsay, France.
Correspondence to C.B.
e-mail: celine.brochier@ibsm.cnrs-mrs.fr
doi:10.1038/nrmicro1852

The RNA component of the small subunit of the ribosome (referred to here as SSU rRNA) has been the 'Rosetta stone' of modern evolutionary studies[1]. In particular, the discovery of the archaeal domain and establishment of the evolutionary relationships between archaeal species were based entirely on rRNA studies[2–5]. These analyses led to the proposal that the archaeal domain should be divided into two phyla, the Euryarchaeota (from the Greek '*euryos*', meaning diversity) and the Crenarchaeota (from the Greek '*crenos*', meaning spring or origin)[6]. At that time, the Euryarchaeota included a mixture of methanogens, extreme halophiles, thermoacidophiles and a few hyperthermophiles. By contrast, the Crenarchaeota included only hyperthermophiles (hence their name, which refers to a 'hot origin of life' hypothesis). This division of the Archaea was rapidly accepted, because it had been observed in the early days of archaeal research that Sulfolobales and Thermoproteales (two hyperthermophilic crenarchaeota orders) are fundamentally different to other archaea in terms of their SSU rRNA oligonucleotide catalogues[7] and RNA polymerase structures[8].

More recently, genomic data[9] and gene phylogenies that have been obtained from combined datasets[10–12] have also confirmed the division of the Archaea into two main lineages, although Euryarchaeota are sometimes paraphyletic in whole-genome trees, probably owing to artefacts that have been introduced by horizontal gene transfer (HGT) from bacteria[13,14]. Several genes that are involved in key cellular processes in the Euryarchaeota lack homologues in all hyperthermophilic crenarchaeota for which complete genome sequences are available[15–18]. For example, there are no homologues of the DNA polymerase from the D family[19] and the cell-division protein FtsZ[20] in hyperthermophilic crenarchaeota, both of which are present in all sequenced complete euryarchaeal genomes. Furthermore, this group of organisms lacks homologues of the eukaryotic-like histone[21] and the protein MinD (involved in chromosome and plasmid partitioning[15]), both of which are present in most sequenced euryarchaeal genomes. This indicates that important differences in main cellular processes were established shortly after the speciation of the Euryarchaeota and Crenarchaeota[14].

The discovery of mesophilic crenarchaeota

More than 20 years ago, direct PCR amplification of genes that encode the SSU rRNA from environmental samples gave rise to molecular ecology[22]. One of the major early outcomes of this new discipline was the discovery of many novel lineages of mesophilic or psychrophilic archaea[23,24] (reviewed in REFS 25,26). The first environmental archaeal sequences were detected in marine environments, and were clearly separated into two groups (named group I and group II) in an SSU rRNA tree that was rooted by a bacterial outgroup[23]. Group I formed a sister group of hyperthermophilic crenarchaeota, whereas group II emerged within the Euryarchaeota[23].

Possibly because they were discovered only 2 years after the generally accepted proposal to divide Archaea into 2 phyla[6], group I was classified as Crenarchaeota[23,24], even though it was only a sister group of hyperthermophilic crenarchaeota and did not branch off within them. The classification of group I archaea as Crenarchaeota was further strengthened by the phylogenetic analysis of a DNA polymerase sequence from *Cenarchaeum symbiosum* (a marine archaeon that inhabits the tissues of a temperate water sponge[27]), which branched within sequences from hyperthermophilic crenarchaeota[28]. Consistent with this, a recent, and widely accepted, SSU rRNA tree that was published by Schleper and colleagues[25,29], and has been widely used to illustrate archaeal phylogeny, shows mesophilic archaea of group I emerging within hyperthermophilic crenarchaeota. This phylogenetic placement is consistent with the assumption that mesophilic crenarchaeota evolved from hyperthermophilic ancestors through adaptation to a mesophilic lifestyle[11,14,30–32]. However, this placement remains controversial, because in most SSU rRNA phylogenies, such as the one recently published by Pace's group[33], group I sequences do not emerge within cultivated hyperthermophilic crenarchaeota and form a distinct lineage. Interestingly, the recent discovery of a eukaryotic-like histone gene that was probably not acquired by HGT in a genomic fragment from *C. symbiosum*[34] suggests that mesophilic crenarchaeota might have genomic features that are substantially different from those of hyperthermophilic crenarchaeota. Indeed, homologues of this gene are present in most euryarchaeal genomes, but never in hyperthermophilic crenarchaeota.

The ecological importance of mesophilic crenarchaeota, an extremely diverse group that is widely distributed in oceans and soils[35], is being increasingly recognized. Indeed, molecular environmental surveys have extended the diversity of mesophilic crenarchaeota by revealing several new lineages that are related to group I sequences, such as SAGMCG-1, FFS, marine benthic groups B and C, YNPFFA and THSC1 (reviewed in REFS 25,26). Some of these crenarchaeota might be moderate thermophiles or psychrophiles, even though the group is still designated as mesophilic crenarchaeota. Mesophilic crenarchaeota comprise organisms that are probably important participants in the global carbon and nitrogen cycles[25,36,37], and might be the most abundant ammonia oxidizers in soil ecosystems[37]. For example, it was reported that *Candidatus* Nitrosopumilus maritimus, a recently isolated mesophilic crenarchaeon, can grow chemolithoautotrophically by aerobically oxidizing ammonia to nitrite, which was the first observation of nitrification in the Archaea[38].

Investigating the phylogenetic position of mesophilic crenarchaeota within the archaeal phylogeny, together with their gene content and genomic features, could, therefore, provide valuable information on the evolution of the Archaea.

Sister groups
In a phylogeny, two lineages that share an exclusive common ancestor.

Monophyletic group
Includes an ancestor and all its descendants.

Can rRNA resolve deep archaeal phylogeny?

The phylogenetic position of mesophilic crenarchaeota is currently based solely on SSU rRNA sequences. The trees that were published by Schleper *et al.*[25] and Robertson *et al.*[33] included a large number of sequences (1,344 and 712 SSU rRNA sequences, respectively), but both showed poor resolution of the relative order of emergence of the different archaeal lineages and it was pointed out that the Crenarchaeota and Euryarchaeota appeared as polytomies (star radiations)[33]. This lack of resolution showed that SSU rRNA sequences do not contain enough phylogenetic signal to resolve the deepest nodes of the archaeal phylogeny, probably owing to their size, which limits the number of nucleotide positions that are available for phylogenetic analyses. However, the number of positions that can be used for phylogenetic analyses can be increased by a combined analysis of SSU and large subunit (LSU) rRNA sequences.

FIGURE 1 shows a maximum likelihood phylogenetic tree that is based on the concatenation of 226 SSU and LSU sequences from complete genomes that are representative of archaeal and bacterial diversity, as well as 18 mesophilic crenarchaeal or euryarchaeal fosmids that contain both types of sequences. Mesophilic crenarchaeal fosmid sequences belong to three distinct subgroups: groups 1.1a and 1.1b[25], and the recently proposed deep-branching HWCG III group[39]. The bacterial part of the tree shows a phylogeny that is consistent with those previously published (that is, high statistical support for the monophyly of most bacterial phyla, but a low resolution of their relative order of emergence (not shown)). For the Archaea, the monophyly of most orders within both Euryarchaeota and Crenarchaeota is robustly recovered (FIG. 1). However, the relationships among most euryarchaeal orders are poorly resolved (bootstrap value (BV) of less than 70%) (FIG. 1), and even the monophyly of the Euryarchaeota is not significantly supported (BV of less than 16%). Importantly, both mesophilic and hyperthermophilic crenarchaeota were recovered as two robust monophyletic groups (BV of 99 and 100%, respectively), which is consistent with the SSU rRNA tree published by Robertson and colleagues[33], but not with the tree that was published by Schleper and colleagues[25]. Moreover, mesophilic and hyperthermophilic crenarchaeota form a sister group, but with low support (BV of 36%), and the node is extremely unstable. For example, using a different evolutionary model, the position of mesophilic crenarchaeota was altered — they branched at the base of the archaeal tree and, therefore, became the sister group of a large group that included Euryarchaeota and hyperthermophilic crenarchaeota — but still with low statistical support (BV of 20%; not shown).

A possible explanation for such poor resolution could be the heterogeneity of G+C content among sequences. Sequences from hyperthermophilic euryarchaeota and crenarchaeota have higher G+C content compared with that of mesophilic organisms. This well-known compositional bias of RNA sequences might blur the genuine phylogenetic signal[40]. To investigate this possibility, we used a recently developed phylogenetic method that reduces the biases that are due to convergent G+C content (nhPHYML[41]). We tested three possible deep placements for mesophilic crenarchaeota, based on the rRNA archaeal phylogeny of FIG. 1: first, as a sister group of hyperthermophilic crenarchaeota; second, as a sister group of a cluster that comprises Euryarchaeota and

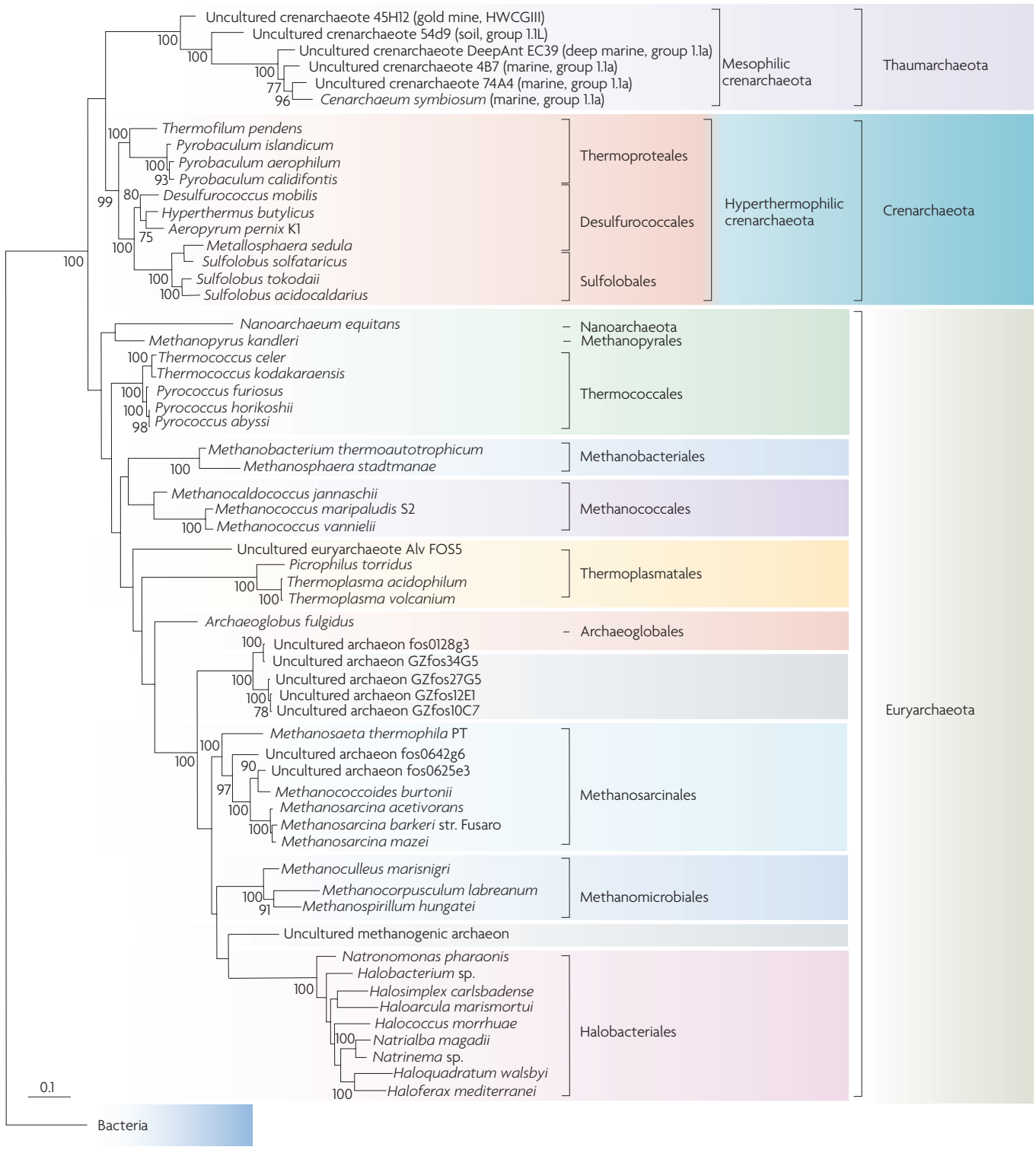

Figure 1 | **Maximum likelihood tree based on the concatenation of 226 SSU and LSU sequences from Archaea and Bacteria.** For clarity, the bacterial part of the tree is not shown. Sequences were aligned using MUSCLE (multiple sequence comparison by log-expectation)[58]. Resulting alignments were manually refined using the MUST (Management Utilities for Sequences and Trees) package[59], and only unambiguously aligned regions were kept for phylogenetic analyses. Concatenation was performed using home-developed software (C.B., unpublished data), which provided a final dataset of 3,305 nucleotide positions. The maximum likelihood tree was computed by PHYML[61], using the general time-reversible model of sequence evolution by including a Γ-correction (eight categories of evolutionary rates, an estimated α-parameter and an estimated proportion of invariant sites). Numbers at nodes represent non-parametric bootstrap values (BVs) that were computed by PHYML[61] (1,000 replications of the original dataset) using the same parameters. For clarity, only BVs of more than 70% are shown. The scale bar represents the average number of substitutions per site. If a different evolutionary model (Hasegawa Kishino Yano) was used, a sister grouping of hyperthermophilic crenarchaeota and euryarchaeota, and a basal branching of mesophilic crenarchaeota was recovered, albeit with weak statistical support (BV of 20%).

hyperthermophilic crenarchaeota; and, third, as a sister group of Euryarchaeota (Supplementary information S1 (table)). All six tests significantly rejected the third topology, whereas only two tests rejected the second topology. This means that the tests discard the third topology, but do not allow discarding the second topology in favour of the first topology. It is likely that the phylogenetic signal which is carried by rRNA sequences is too weak to confidently resolve the position of mesophilic crenarchaeota in the archaeal phylogeny, even if the number of positions is increased by combining SSU and LSU rRNA sequences. Nevertheless, the phylogenetic analysis of the rRNAs strongly supports the separation of mesophilic and hyperthermophilic crenarchaeota into 2 distinct lineages (BV of 100 and 99% for the monophyly of each lineage, respectively). To clarify the position of mesophilic crenarchaeota in the archaeal tree further, the use of alternative markers thus becomes crucial.

Analysing ribosomal proteins

Although they were first discovered 15 years ago, the isolation and cultivation of representative mesophilic crenarchaeota has proven to be a frustrating task. In fact, the first genome of a member of this group, *C. symbiosum,* which has still not been grown in pure culture, was published only recently[42]. The availability of this genome sequence now permits an investigation of the phylogenetic position of mesophilic crenarchaeota, based on markers other than SSU and LSU rRNA.

Owing to the availability of an increasing number of complete archaeal genomes, large concatenated datasets of ribosomal (R) proteins are now widely used as an alternative to SSU rRNA to study archaeal phylogeny[43–45]. Indeed, these proteins have the same evolutionary attributes as rRNA, and their concatenation allows the construction of larger alignments. Although the trees that were obtained using these markers were roughly congruent with the rRNA trees[43], they substantially improved the archaeal phylogeny and resolved a number of important nodes (reviewed in REFS 11,14). In particular, these analyses have helped to clarify the phylogenetic positions of 'lonely' archaeal species (those that lack sequenced relatives), which are often misplaced, especially if they are fast-evolving or have a biased sequence composition (for example, the G+C content of rRNA sequences)[46]. For example, *Nanoarchaeum equitans* was originally proposed to represent a third (and basal) archaeal phylum based on trees that were produced using SSU rRNA[47] and concatenated R proteins[44]. However, a subsequent analysis of R proteins and additional protein markers suggested that this species is not the earliest archaeal offshoot, but is probably a fast-evolving euryarchaeal lineage that is possibly related to Thermococcales[48]. Another example is the hyperthermophilic methanogen *Methanopyrus kandleri*, for which phylogenetic placement is crucial to obtain an understanding of the time of emergence of methanogenesis within Euryarchaeota. In fact, although *M. kandleri* represents the earliest euryarchaeal offshoot in SSU rRNA phylogenies[25,49], in trees that are based on R-protein concatenations it robustly branches off after the non-methanogenic lineage of Thermococcales[10,11,50]. Further, a recent phylogenetic analysis placed this archaeon as a sister group of two other methanogen lineages (Methanococcales and Methanobacteriales)[51], which is in agreement with phylogenomic studies of the genes that are involved in methanogenesis[51] and gene-content analyses[45]. Globally, these analyses indicate that methanogenesis might not be the ancestral metabolism of euryarchaeota.

The examples of *N. equitans* and *M. kandleri* highlight the power of R-protein combined datasets for phylogenetic reconstruction. We therefore applied the same approach to study the placement of *C. symbiosum* in the archaeal phylogeny. FIGURE 2 shows a maximum likelihood phylogeny of the archaeal domain that is based on the concatenation of 53 R-protein sequences from 48 complete archaeal genomes and was rooted using sequences from 16 eukaryotes. The phylogeny includes *C. symbiosum*, 33 Euryarchaeota and 14 hyperthermophilic crenarchaeota, which represents 21 new species (11 Euryarchaeota, and 1 mesophilic and 9 hyperthermophilic crenarchaeota, respectively) with respect to previous similar analyses[11]. This tree is better resolved than the SSU/LSU rRNA tree in FIG. 1 (note the higher BVs at nodes in FIG. 2), and the positions of the newly included archaea are well supported and in agreement with their classification. Consequently, *Thermofilum pendens*, *Caldivirga maquilingensis*, *Pyrobaculum calidifontis*, *Pyrobaculum arsenaticum* and *Pyrobaculum islandicum* are grouped with *Pyrobaculum aerophilum* (group of Thermoproteales; BV of 100%), and *Ignicoccus hospitalis*, *Staphylothermus marinus* and *Hyperthermus butylicus* are grouped with *Aeropyrum pernix* (group of Desulfurococcales; BV of 97%), whereas *Metallosphaera sedula* is grouped with other Sulfolobales (BV of 100%). In Euryarchaeota, *Natronomonas pharaonis*, *Halorubrum lacusprofundi* and *Haloquadratum walsbyi* are grouped with other Halobacteriales (BV of 100%). The four Methanomicrobiales (*Methanocorpusculum labreanum*, *Methanospirillum hungatei*, *Candidatus* Methanoregula boonei and *Methanoculleus marisnigri*) are grouped together (BV of 100%) within a cluster that also contains Methanosarcinales (including their new representative *Methanosaeta thermophila*; BV of 100%) and Halobacteriales (BV of 100%). Finally, *Methanosphaera stadtmanae* emerges as a sister group of the other Methanobacteriale *Methanothermobacter thermautotrophicus* (BV of 100%), whereas *Methanococcus aeolicus* and *Methanococcus vannielii* are grouped with the other Methanococcales (BV of 100%).

In contrast to the tree that is based on SSU and LSU rRNA (FIG. 1), most relationships among the archaeal orders are well resolved and in agreement with previous studies[10], which highlights that R proteins are the phylogenetic markers of choice to study the archaeal phylogeny. Importantly, the monophylies of both hyperthermophilic crenarchaeota and Euryarchaeota are robustly recovered (each has a BV of 100%; FIG. 2). Interestingly, *C. symbiosum* constitutes a deeply branching lineage (BV of 99%), as it is a sister group of a clade that contains both Euryarchaeota (including *N. equitans*) and hyperthermophilic crenarchaeota. We think that this position is genuine and not the consequence of a long-branch attraction artefact, as the branch that leads to *C. symbiosum* is not particularly long

Clade
A monophyletic group.

Long-branch attraction artefact
A phylogenetic artefact that is induced by differences in evolutionary rates, and results in the artificial grouping of lineages that have long branches in a phylogenetic tree.

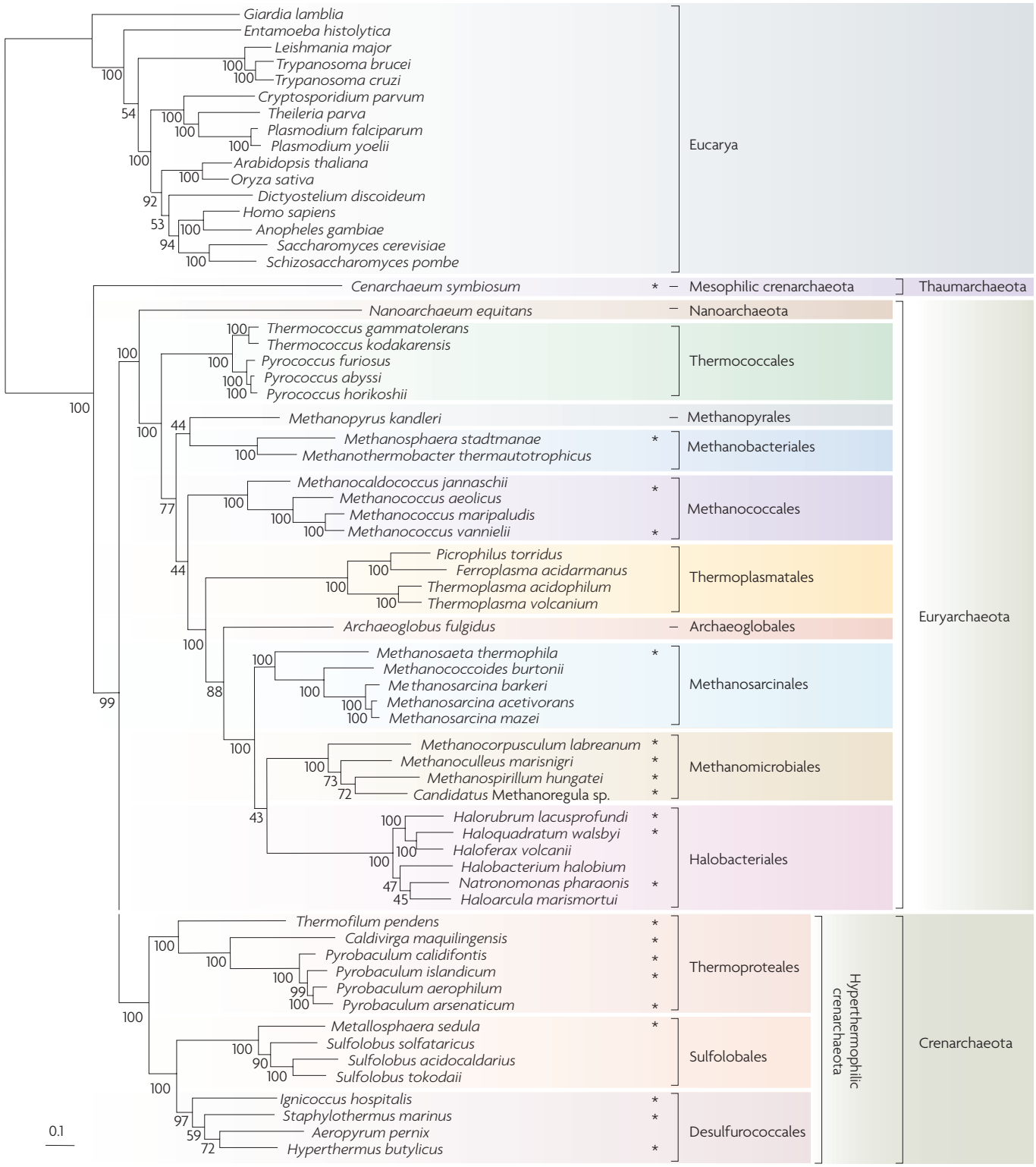

Figure 2 | **Maximum likelihood tree based on the concatenation of 53 R proteins from complete archaeal genomes.** Homologues of each R protein in complete genomes were retrieved by BLASTP and TBLASTN[60]. The concatenation included 53 alignments that harboured sequences from at least 61 of 64 taxa. The maximum likelihood phylogenetic tree was reconstructed using PHYML[61], with the Jones Taylor Thornton model of sequence evolution, by including a Γ-correction (eight categories of evolutionary rates, an estimated α-parameter and an estimated proportion of invariant sites). Numbers at nodes represent non-parametric bootstrap values computed by PHYML[61] (100 replications of the original dataset) using the same parameters. The use of different evolutionary models and methods did not produce differences in the resulting tree topology, at least for the archaeal part of the tree (not shown). Asterisks indicate the 21 new species (1 representative of the mesophilic crenarchaeota, *Cenarchaeum symbiosum*, 9 representatives of hyperthermophilic crenarchaeota and 11 representatives of Euryarchaeota) that were included in this analysis compared with previous work[11]. The scale bar represents the average number of substitutions per site.

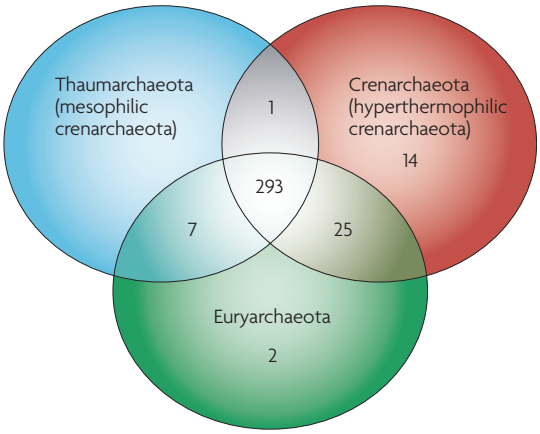

Figure 3 | **Scheme showing the number of proteins shared by Euryarchaeota, mesophilic crenarchaeota and hyperthermophilic crenarchaeota.**

in the tree (Fig. 2) or in individual R-protein trees (not shown), which indicates that its R proteins are not particularly fast-evolving. Moreover, even the fast-evolving Thermoplasmatales and lonely taxon *N. equitans* are not artificially attracted at the base of the tree (FIG. 2). However, a definitive exclusion of a long-branch attraction artefact[52] that could affect the position of *C. symbiosum* in this tree will only be possible by the addition of sequences from its relatives.

In conclusion, in contrast to SSU/LSU rRNA, analysis of R proteins improves the resolution of the deepest nodes in the archaeal phylogeny and suggests that mesophilic crenarchaeota could have diverged before the speciation of Euryarchaeota and hyperthermophilic crenarchaeota.

A conserved crenarchaeal genomic core?

Our SSU/LSU rRNA analysis only weakly suggests that mesophilic and hyperthermophilic crenarchaeota are sister groups (FIG. 1). By contrast, the analysis of R proteins indicates a robust and deeper branching of *C. symbiosum* that occurred before the speciation between Euryarchaeota and hyperthermophilic crenarchaeota (FIG. 2). This placement implies that mesophilic crenarchaeota are not more related to hyperthermophilic crenarchaeota than they are to Euryarchaeota. Thus, we investigated the presence in *C. symbiosum* of genes that seem to be strictly specific to Euryarchaeota (genes that are present in at least one representative of each major order of Euryarchaeota, but are absent from all representatives of Crenarchaeota); strictly specific to hyperthermophilic crenarchaeota (genes that are present in at least one representative of each major order of thermophilic crenarchaeota, but are absent from all representatives of Euryarchaeota); or that are common to Euryarchaeota and thermophilic crenarchaeota (FIG. 3). This criterion might seem stringent, as it excludes the markers that have been secondarily lost from some lineages (for example, histones in Thermoplasmatales). However, it has the advantage of focusing on genes that comprise the strictly conserved genomic core of Euryarchaeota and hyperthermophilic crenarchaeota, but avoiding the introduction of ambiguities that are due to genes with scattered distributions.

Using the NCBI COGs database (see Further information)[53], we identified 12 proteins that are strictly specific to Euryarchaeota, 15 proteins that are strictly specific to hyperthermophilic crenarchaeota (Supplementary information S2 (table)) and 318 proteins that are common to both phyla. Surprisingly, we found that *C. symbiosum* harbours 10 of the 12 euryarchaeal-specific proteins. Because HGTs from Euryarchaeota to mesophilic crenarchaeota were detected in a genome fragment from an uncultivated mesophilic crenarchaeon[32], we carried out a phylogenetic analysis of the ten euryarchaeal-specific proteins that were harboured by *C. symbiosum*. These trees, although generally poorly resolved (not shown), revealed that only three of these proteins might be present owing to HGT, whereas the remaining seven are probably ancestral traits that are common to Euryarchaeota and *C. symbiosum* (FIG. 3; Supplementary information S2 (table)). By contrast, *C. symbiosum* lacks 14 of the 15 hyperthermophilic crenarchaeal-specific proteins (including two R proteins) (FIG. 3; Supplementary information S2 (table)). Thus, with respect to the conserved genomic core, the mesophilic crenarchaeon *C. symbiosum* seems to be more similar to Euryarchaeota than to hyperthermophilic crenarchaeota. Importantly, a few of the euryarchaeal-specific genes that are present in *C. symbiosum* encode proteins that are involved in core cellular processes, such as DNA replication and cell division (Supplementary information S2 (table)), which shows that biologically important differences distinguish this organism, and by extension all mesophilic crenarchaeota, from hyperthermophilic crenarchaeota.

In addition to the presence of most euryarchaeal-specific proteins and absence of most proteins that are specific to hyperthermophilic crenarchaeota, *C. symbiosum* also lacks 25 proteins that are present in both Euryarchaeota and hyperthermophilic crenarchaeota, including the R protein S24e and the type I DNA topoisomerase of the A family (IA) (FIG. 3; Supplementary information S2 (table)). The absence of topoisomerase IA from *C. symbiosum* is surprising, as a protein from this family is present in representatives from the three domains of life[54], including archaea. Finally, *C. symbiosum* lacks the R protein L14e, which is present in all available genomes from hyperthermophilic crenarchaeota and basal euryarchaeota (Methanopyrales, Methanococcales, Methanobacteriales, Thermococcales and *N. equitans*), and the R protein L20a, which is present in all archaeal genomes except Thermoplasmatales. Moreover, we have identified potentially informative insertions and deletions (indels) in two other proteins, the R protein S27ae (hyperthermophilic crenarchaeota harbour a three-amino acid insertion that is absent from Euryarchaeota and mesophilic crenarchaeota) and the elongation factor EF-1α (both hyperthermophilic and mesophilic crenarchaeota harbour a conserved seven-amino acid insertion that is absent from Euryarchaeota). The distribution patterns of the features in the genome of *C. symbiosum* discussed above are puzzling, because they suggest that mesophilic crenarchaeota have a combination of traits that are either specific to hyperthermophilic crenarchaeota or Euryarchaeota.

Similar genome-mining data were recently obtained independently by Makarova, Koonin and co-workers[55], using an updated version of the NCBI COGs database that focused on Archaea. These authors noticed that the genome of *C. symbiosum* includes a much lower proportion of archaeal COGs than other archaeal genomes and groups with Euryarchaeota in a gene-content tree. They concluded from their analysis that "*C. symbiosum* is not a typical crenarchaeon (REF. 55)".

A third archaeal phylum?

Our SSU/LSU rRNA tree (FIG. 1) and analysis of the conserved genomic cores strongly reject the hypothesis that mesophilic crenarchaeota evolved from hyperthermophilic crenarchaeota (BV of 100%, which supports the monophyly of hyperthermophilic crenarchaeota). Moreover, our R-protein concatenation tree (FIG. 2) strongly rejects a sister-group relationship between hyperthermophilic crenarchaeota and *C. symbiosum*. Rather, it favours a deeper branching before the speciation of hyperthermophilic crenarchaeota and Euryarchaeota. The analysis of the genomic cores shows that *C. symbiosum* shares more features with Euryarchaeota than with hyperthermophilic crenarchaeota. This might indicate that *C. symbiosum* and its uncultivated relatives either belong to, or are sister to, Euryarchaeota. However, this is excluded by our phylogenetic analyses. Consistent with the basal emergence of mesophilic crenarchaeota, the genes of the euryarchaeal core that are shared with *C. symbiosum* can be interpreted as being ancestral characters that were present in the ancestor of archaea and were secondarily lost in the branch that led to hyperthermophilic crenarchaeota. We predict that the genomes of other mesophilic crenarchaeota from marine and terrestrial environments[56], such as *Candidatus* N. maritimus, will confirm our results when they become available for analysis. Moreover, this will enable the identification of features that are specific to the group, such as a conserved genomic core. One such feature could be the presence of a type I DNA topoisomerase of the B family (IB), which we detected in the genome of *C. symbiosum*. Whereas members of the topoisomerase IB family have never been identified in archaea, they are almost universal in eukarya and rarely present in bacteria[54]. This probably correlates with the absence from *C. symbiosum* of topoisomerase IA, which is present in all other archaea. Interestingly, the topoisomerase IB of *C. symbiosum* branches as a sister group to eukaryotes (not shown), which suggests that it was not transferred from the sponge host. A topoisomerase IB that was present in the last common ancestor of archaea and eukaryotes could later have been lost in the lineage that led to Euryarchaeota and hyperthermophilic crenarchaeota after their divergence from mesophilic crenarchaeota.

The diversity of mesophilic crenarchaeota based on SSU rRNA sequences[25,26,56,57] is comparable to that of hyperthermophilic crenarchaeota and Euryarchaeota, which suggests that they represent a major lineage that has equal status to Euryarchaeota and hyperthermophilic crenarchaeota. Indeed, environmental SSU rRNA surveys have already revealed several likely order-level subgroups within mesophilic crenarchaeota[25,26,56]. Moreover, the basal placement of one of their representatives in the archaeal phylogeny (FIG. 2) suggests that mesophilic crenarchaeota are an ancient lineage. This leads us to propose that mesophilic crenarchaeota represent a third archaeal phylum that we suggest naming the Thaumarchaeota (from the Greek '*thaumas*', meaning wonder). This choice was made to avoid any name that referred to phenotypic properties, such as mesophily, that could be challenged by the future identification of non-mesophilic organisms that belong to this phylum or the discovery of mesophilic relatives of cultivated hyperthermophilic crenarchaeota.

We stress that the classification of archaeal group I and its relatives as crenarchaeota was dubious from the outset, because their sequences formed only a sister group of hyperthermophilic crenarchaeota in the first rRNA trees[23]. The acceptance of this classification was probably influenced by the fact that the proposal to split the archaeal domain between Crenarchaeota and Euryarchaeota had only recently been made[6]. Clearly, the current classification of mesophilic crenarchaeota as Crenarchaeota is misleading, just as it is misleading to call methanogens 'methanogenic bacteria' because all methanogens are archaea. The proposal to establish mesophilic crenarchaeota as a third archaeal phylum goes beyond purely taxonomic purposes, and will stimulate research on this group of organisms and, more generally, on the Archaea.

Further phylogenetic analyses that include new members of the Thaumarchaeota are required to confirm the position of this phylum in the archaeal phylogeny. In any case, even if the basal branching of mesophilic crenarchaeota is challenged in favour of a sister grouping with hyperthermophilic crenarchaeota, this should not, in our opinion, change their phylum status, as they would remain a highly diversified and ancient group that have peculiar genomic characteristics. If the emergence of Thaumarchaeota prior to the speciation of Crenarchaeota and Euryarchaeota (as supported by R-protein analysis) is confirmed, this will leave open the nature of the last archaeal ancestor, which might have been either a mesophilic or psychrophilic organism (such as Thaumarchaeota) or a hyperthermophilic or thermophilic organism (such as cultivated crenarchaeota and some euryarchaeota). Importantly, the nature of the archaeal ancestor provides a different meaning for the HGTs from mesophilic euryarchaeota and bacteria to Thaumarchaeota that were highlighted from environmental genomics studies[32]. If the ancestor of Archaea was a hyperthermophile, HGT might have enabled the adaptation of hyperthermophilic thaumarchaeal lineages towards mesophily, as has been previously suggested[32]. Conversely, if the archaeal ancestor was a mesophile, HGT might have occurred between organisms that were thriving in the same low-temperature environments. Further studies on Thaumarchaeota will be essential to gain fundamental insights into the origin and early evolution of Archaea.

Mesophile
This term is normally restricted to organisms that have optimal growth temperatures of between 20 and 50°C. Here, however, the term mesophilic crenarchaeota is given to all non-hyperthermophilic crenarchaeota, even though some of them (presently uncultivated) are psychrophiles (optimal growth temperature of between 0 and 20°C) or moderate thermophiles (optimal growth temperature of between 50 and 70°C).

1. Pace, N. R. A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–740 (1997).
2. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl Acad. Sci. USA* **74**, 5088–5090 (1977).
3. Fox, G. E. *et al.* The phylogeny of prokaryotes. *Science* **209**, 457–463 (1980).
4. Woese, C. R. Bacterial evolution. *Microbiol. Rev.* **51**, 221–271 (1987).
5. Woese, C. R. in *Archaea: Evolution, Physiology and Molecular Biology* (eds Garrett, R. A. & Klenk, H. P.) 1–15 (Blackwell publishing, Oxford, 2006).
An eloquent historical review that reported, for the first time, all the steps that led to the discovery of Archaea. Forms the introduction of an interesting book that covers the various aspects of archaeal physiology and molecular biology and focuses on the similarities between Archaea and Eukaryotes.
6. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl Acad. Sci. USA* **87**, 4576–4579 (1990).
7. Woese, C. R., Gupta, R., Hahn, C. M., Zillig, W. & Tu, J. The phylogenetic relationships of three sulfur dependent archaebacteria. *Syst. Appl. Microbiol.* **5**, 97–105 (1984).
8. Prangishvilli, D., Zillig, W., Gierl, A., Biesert, L. & Holz, I. DNA-dependent RNA polymerase of thermoacidophilic archaebacteria. *Eur. J. Biochem.* **122**, 471–477 (1982).
9. Makarova, K. S. & Koonin, E. V. Comparative genomics of archaea: how much have we learned in six years, and what's next? *Genome Biol.* **4**, 115 (2003).
10. Brochier, C., Forterre, P. & Gribaldo, S. An emerging phylogenetic core of Archaea: phylogenies of transcription and translation machineries converge following addition of new genome sequences. *BMC Evol. Biol.* **5**, 36 (2005).
Revealed a conserved core of vertically inherited genes in different cellular systems, which proved that reconstructing the phylogeny of species is a feasible task in Archaea.
11. Gribaldo, S. & Brochier-Armanet, C. The origin and evolution of Archaea: a state of the art. *Phil. Trans. R. Soc. Lond. B* **361**, 1007–1022 (2006).
12. Daubin, V., Gouy, M. & Perriere, G. A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome Res.* **12**, 1080–1090 (2002).
13. Wolf, Y. I., Rogozin, I. B., Grishin, N. V., Tatusov, R. L. & Koonin, E. V. Genome trees constructed using five different approaches suggest new major bacterial clades. *BMC Evol. Biol.* **1**, 8 (2001).
14. Forterre, P., Gribaldo, S. & Brochier-Armanet, C. in *Archaea: Evolution, Physiology and Molecular Biology* (eds Garrett, R. A. & Klenk, H. P.) 17–29 (Blackwell publishing, Oxford, 2006).
15. Bernander, R. Chromosome replication, nucleoid segregation and cell division in Archaea. *Trends Microbiol.* **8**, 278–283 (2000).
16. Myllykallio, H. *et al.* Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* **288**, 2212–2215 (2000).
17. Makarova, K. S. & Koonin, E. V. Evolutionary and functional genomics of the Archaea. *Curr. Opin. Microbiol.* **8**, 586–594 (2005).
18. Klenk, H. P. in *Archaea: Evolution, Physiology and Molecular Biology* (eds Garrity, G. M. & Klenk, H. P.) 75–95 (Blackwell publishing, Oxford, 2006).
19. Uemori, T., Sato, Y., Kato, I., Doi, H. & Ishino, Y. A novel DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus*: gene cloning, expression, and characterization. *Genes Cells* **2**, 499–512 (1997).
20. Margolin, W., Wang, R. & Kumar, M. Isolation of an *ftsZ* homolog from the archaebacterium *Halobacterium salinarium*: implications for the evolution of FtsZ and tubulin. *J. Bacteriol.* **178**, 1320–1327 (1996).
21. Allers, T. & Mevarech, M. Archaeal genetics — the third way. *Nature Rev. Genet.* **6**, 58–73 (2005).
22. Olsen, G. J., Lane, D. J., Giovannoni, S. J., Pace, N. R. & Stahl, D. A. Microbial ecology and evolution: a ribosomal RNA approach. *Annu. Rev. Microbiol.* **40**, 337–365 (1986).
23. DeLong, E. F. Archaea in coastal marine environments. *Proc. Natl Acad. Sci. USA* **89**, 5685–5689 (1992).
24. Fuhrman, J. A., McCallum, K. & Davis, A. A. Novel major archaebacterial group from marine plankton. *Nature* **356**, 148–149 (1992).
25. Schleper, C., Jurgens, G. & Jonuscheit, M. Genomic studies of uncultivated archaea. *Nature Rev. Microbiol.* **3**, 479–488 (2005).
A fascinating review that underlines the diversity of Archaea and the importance of using metagenomic approaches to gain insights into their biology and physiology.
26. Schleper, C. in *Archaea: Evolution, Physiology and Molecular Biology* (eds Garrett, R. A. & Klenk, H. P.) 39–50 (Blackwell publishing, Oxford, 2006).
27. Preston, C. M., Wu, K. Y., Molinski, T. F. & DeLong, E. F. A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov., sp. nov. *Proc. Natl Acad. Sci. USA* **93**, 6241–6246 (1996).
28. Schleper, C., Swanson, R. V., Mathur, E. J. & DeLong, E. F. Characterization of a DNA polymerase from the uncultivated psychrophilic archaeon *Cenarchaeum symbiosum*. *J. Bacteriol.* **179**, 7803–7811 (1997).
29. Garrett, R. A. & Klenk, H. P. (eds) *Archaea: Evolution, Physiology and Molecular Biology* (Blackwell publishing, Oxford, 2006).
30. Barns, S. M., Delwiche, C. F., Palmer, J. D. & Pace, N. R. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc. Natl Acad. Sci. USA* **93**, 9188–9193 (1996).
31. Hershberger, K. L., Barns, S. M., Reysenbach, A. L., Dawson, S. C. & Pace, N. R. Wide diversity of Crenarchaeota. *Nature* **384**, 420 (1996).
32. Lopez-Garcia, P., Brochier, C., Moreira, D. & Rodriguez-Valera, F. Comparative analysis of a genome fragment of an uncultivated mesopelagic crenarchaeote reveals multiple horizontal gene transfers. *Environ. Microbiol.* **6**, 19–34 (2004).
33. Robertson, C. E., Harris, J. K., Spear, J. R. & Pace, N. R. Phylogenetic diversity and ecology of environmental Archaea. *Curr. Opin. Microbiol.* **8**, 638–642 (2005).
A recent, exhaustive archaeal phylogeny that was based on an analysis of SSU rRNA sequences. Despite the use of a large number of sequences, this marker was unable to resolve the deepest nodes of the archaeal phylogeny.
34. Cubonova, L., Sandman, K., Hallam, S. J., Delong, E. F. & Reeve, J. N. Histones in *Crenarchaea*. *J. Bacteriol.* **187**, 5482–5485 (2005).
35. Ochsenreiter, T., Selezi, D., Quaiser, A., Bonch-Osmolovskaya, L. & Schleper, C. Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. *Environ. Microbiol.* **5**, 787–797 (2003).
36. Wuchter, C. *et al.* Archaeal nitrification in the ocean. *Proc. Natl Acad. Sci. USA* **103**, 12317–12322 (2006).
37. Leininger, S. *et al.* Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442**, 806–809 (2006).
38. Konneke, M. *et al.* Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**, 543–546 (2005).
The first report of the isolation of a member of the Thaumarchaeota, and the first demonstration of the ability of an isolated archaeon to oxidize ammonium.
39. Nunoura, T. *et al.* Genetic and functional properties of uncultivated thermophilic crenarchaeotes from a subsurface gold mine as revealed by analysis of genome fragments. *Environ. Microbiol.* **7**, 1967–1984 (2005).
40. Woese, C. R., Achenbach, L., Rouviere, P. & Mandelco, L. Archaeal phylogeny: reexamination of the phylogenetic position of *Archaeoglobus fulgidus* in light of certain composition-induced artifacts. *Syst. Appl. Microbiol.* **14**, 364–371 (1991).
41. Boussau, B. & Gouy, M. Efficient likelihood computations with nonreversible models of evolution. *Syst. Biol.* **55**, 756–768 (2006).
42. Hallam, S. J. *et al.* Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc. Natl Acad. Sci. USA* **103**, 18296–18301 (2006).
43. Matte-Tailliez, O., Brochier, C., Forterre, P. & Philippe, H. Archaeal phylogeny based on ribosomal proteins. *Mol. Biol. Evol.* **19**, 631–639 (2002).
44. Waters, E. *et al.* The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc. Natl Acad. Sci. USA* **100**, 12984–12988 (2003).
45. Slesarev, A. I. *et al.* The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. *Proc. Natl Acad. Sci. USA* **99**, 4644–4649 (2002).
46. Gribaldo, S. & Philippe, H. Ancient phylogenetic relationships. *Theor. Popul. Biol.* **61**, 391–408 (2002).
47. Huber, H. *et al.* A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**, 63–67 (2002).
48. Brochier, C., Gribaldo, S., Zivanovic, Y., Confalonieri, F. & Forterre, P. Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to Thermococcales? *Genome Biol.* **6**, R42 (2005).
49. Burggraf, S., Stetter, K. O., Rouviere, P. & Woese, C. R. *Methanopyrus kandleri*: an archaeal methanogen unrelated to all other known methanogens. *Syst. Appl. Microbiol.* **14**, 346–351 (1991).
50. Brochier, C., Forterre, P. & Gribaldo, S. Archaeal phylogeny based on proteins of the transcription and translation machineries: tackling the *Methanopyrus kandleri* paradox. *Genome Biol.* **5**, R17 (2004).
51. Bapteste, E., Brochier, C. & Boucher, Y. Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. *Archaea* **1**, 353–363 (2005).
52. Felsenstein, J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**, 401–410 (1978).
53. Tatusov, R. L. *et al.* The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29**, 22–28 (2001).
54. Forterre, P., Gribaldo, S., Gadelle, D. & Serre, M. C. Origin and evolution of DNA topoisomerases. *Biochimie* **89**, 427–446 (2007).
55. Makarova, K. S., Wolf, Y. I., Sorokin, A. V. & Koonin, E. V. Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. *Biol. Direct* **2**, 33 (2007).
A dedicated COG database for Archaea that highlights the important differences between Thaumarchaeota and hyperthermophilic crenarchaeota.
56. Nicol, G. W. & Schleper, C. Ammonia-oxidising Crenarchaeota: important players in the nitrogen cycle? *Trends Microbiol.* **14**, 207–212 (2006).
Together with reference 37, this review revealed the unsuspected role of archaea in the global nitrogen cycle, which was previously assumed to be carried out by bacteria.
57. Forterre, P., Brochier, C. & Philippe, H. Evolution of the Archaea. *Theor. Popul. Biol.* **6**, 409–422 (2002).
58. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).
59. Philippe, H. MUST, a computer package of Management Utilities for Sequences and Trees. *Nucleic Acids Res.* **21**, 5264–5272 (1993).
60. Altschul, S. F. & Koonin, E. V. Iterated profile searches with PSI-BLAST — a tool for discovery in protein databases. *Trends Biochem. Sci.* **23**, 444–447 (1998).
61. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).

Acknowledgments
The authors thank G. Sezonov for suggesting the name Thaumarchaeota, E. Koonin for unpublished communications and the referees for useful comments and suggestions.

DATABASES
Entrez Genome Project: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeprj
Aeropyrum pernix | *Caldivirga maquilingensis* | *Candidatus* Methanoregula boonei | *Cenarchaeum symbiosum* | *Halorubrum lacusprofundi* | *Haloquadratum walsbyi* | *Hyperthermus butylicus* | *Ignicoccus hospitalis* | *Metallosphaera sedula* | *Methanococcus aeolicus* | *Methanococcus vannielii* | *Methanocorpusculum labreanum* | *Methanoculleus marisnigri* | *Methanopyrus kandleri* | *Methanosaeta thermophila* | *Methanosphaera stadtmanae* | *Methanospirillum hungatei* | *Methanothermobacter thermautotrophicus* | *Nanoarchaeum equitans* | *Natronomonas pharaonis* | *Pyrobaculum aerophilum* | *Pyrobaculum arsenaticum* | *Pyrobaculum calidifontis* | *Pyrobaculum islandicum* | *Staphylothermus marinus* | *Thermofilum pendens*

FURTHER INFORMATION
Céline Brochier-Armanet's homepage: http://www.frangun.org
Céline Brochier-Armanet's laboratory website: http://lcb.cnrs-mrs.fr/
NCBI COGs database: http://www.ncbi.nlm.nih.gov/COG/

SUPPLEMENTARY INFORMATION
See online article: S1 (table) | S2 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF