Cell PRESS

Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota

Anja Spang[1], Roland Hatzenpichler[2], Céline Brochier-Armanet[3], Thomas Rattei[4,5], Patrick Tischler[5], Eva Spieck[6], Wolfgang Streit[6], David A. Stahl[7], Michael Wagner[2] and Christa Schleper[1]

[1] University of Vienna, Department of Genetics in Ecology, Althanstrasse 14, A-1090 Vienna, Austria
[2] University of Vienna, Department of Microbial Ecology, Althanstrasse 14, A-1090 Vienna, Austria
[3] Aix-Marseille Université, CNRS, Laboratoire de Chimie Bactérienne, Institut de Biologie Structurale et de Microbiologie, 13402 Marseille, France
[4] University of Vienna, Department of Computational Systems Biology, Althanstrasse 14, A-1090 Vienna, Austria
[5] Technische Universität München, Department of Genome-Oriented Bioinformatics, D-85350 Freising, Germany
[6] University of Hamburg, Biocenter Klein Flottbek, Microbiology and Biotechnology, Ohnhorststrasse 18, D-22609 Hamburg, Germany
[7] University of Washington, Department of Civil and Environmental Engineering, Seattle, WA 98195, USA

Globally distributed archaea comprising ammonia oxidizers of moderate terrestrial and marine environments are considered the most abundant archaeal organisms on Earth. Based on 16S rRNA phylogeny, initial assignment of these archaea was to the Crenarchaeota. By contrast, features of the first genome sequence from a member of this group suggested that they belong to a novel phylum, the Thaumarchaeota. Here, we re-investigate the Thaumarchaeota hypothesis by including two newly available genomes, that of the marine ammonia oxidizer *Nitrosopumilus maritimus* and that of *Nitrososphaera gargensis*, a representative of another evolutionary lineage within this group predominantly detected in terrestrial environments. Phylogenetic studies based on r-proteins and other core genes, as well as comparative genomics, confirm the assignment of these organisms to a separate phylum and reveal a Thaumarchaeota-specific set of core informational processing genes, as well as potentially ancestral features of the archaea.

How many archaeal phyla exist?

Since their recognition as a separate domain of life [1,2] from Eukarya and Bacteria, the Archaea have played an important part in models of the early evolution of cellular life forms [3–5]. In particular, the information-processing machineries of Archaea are considered ancestral forms of the more complex replication, transcription and translation machineries of the eukaryotic cell [6]. To be able to infer early traits (features of the last archaeal common ancestor, LACA) and to study the evolution of metabolism and information processing, it is essential to recognize the diversity within the domain and in particular deeply branching lineages.

Carl Woese and collaborators pointed out that 16S rRNA sequences supported a deep split within the Archaea forming two major lineages (kingdoms, *sensu* Woese [2]), the Crenarchaeota and Euryarchaeota. Crenarchaeota (from Greek for spring or fount [2]) were a potential ancestral branch exclusively represented by hyperthermophilic organisms and the term Euryarchaeota (from Greek for broad, wide) insinuated the global distribution and physiological diversity of its halophilic, methanogenic and thermophilic members. The bipartite separation was supported by several characteristic genetic features, (e.g. the cell division protein FtsZ and histones in Euryarchaeota and specific ribosomal proteins in Crenarchaeota). Interestingly, since 1990 all but two of the newly described species of archaea (~330 in total, www.dsmz.de) belong to one of the two archaeal kingdoms or phyla (*sensu* Bergey's manual [7]) based on their 16S rRNA phylogeny. In the past decade, two additional archaeal phyla, the Korarchaeota [8,9] and Nanoarchaeota [10], were postulated based on the 16S rRNA phylogenies of enrichments of Candidatus *Korarchaeum cryptophilum* and of the obligate symbiont Candidatus *Nanoarchaeum equitans*, respectively. However, analyses of different phylogenetic marker molecules were inconsistent and thus the phylogenetic placement of these organisms (in particular that of *Nanoarchaeum*) is still under debate [9,11] (Figure 1a,b).

With the discovery of ammonia-oxidizing archaea (AOA), accumulating evidence indicates that the archaeal tree indeed comprises more than two major lineages [12]. The discovery of these organisms was initially based on environmentally retrieved 16S rRNA sequences, which placed them as a sister group of the Crenarchaeota (Figure 1b) [13,14], suggesting that these archaea might have ancestors in hot environments and only later radiated into moderate environments. The AOA have since been referred to as Crenarchaeota in all subsequent 16S

Corresponding author: Schleper, C. (christa.schleper@univie.ac.at).

0966-842X/$ – see front matter doi:10.1016/j.tim.2010.06.003 Trends in Microbiology 18 (2010) 331–340

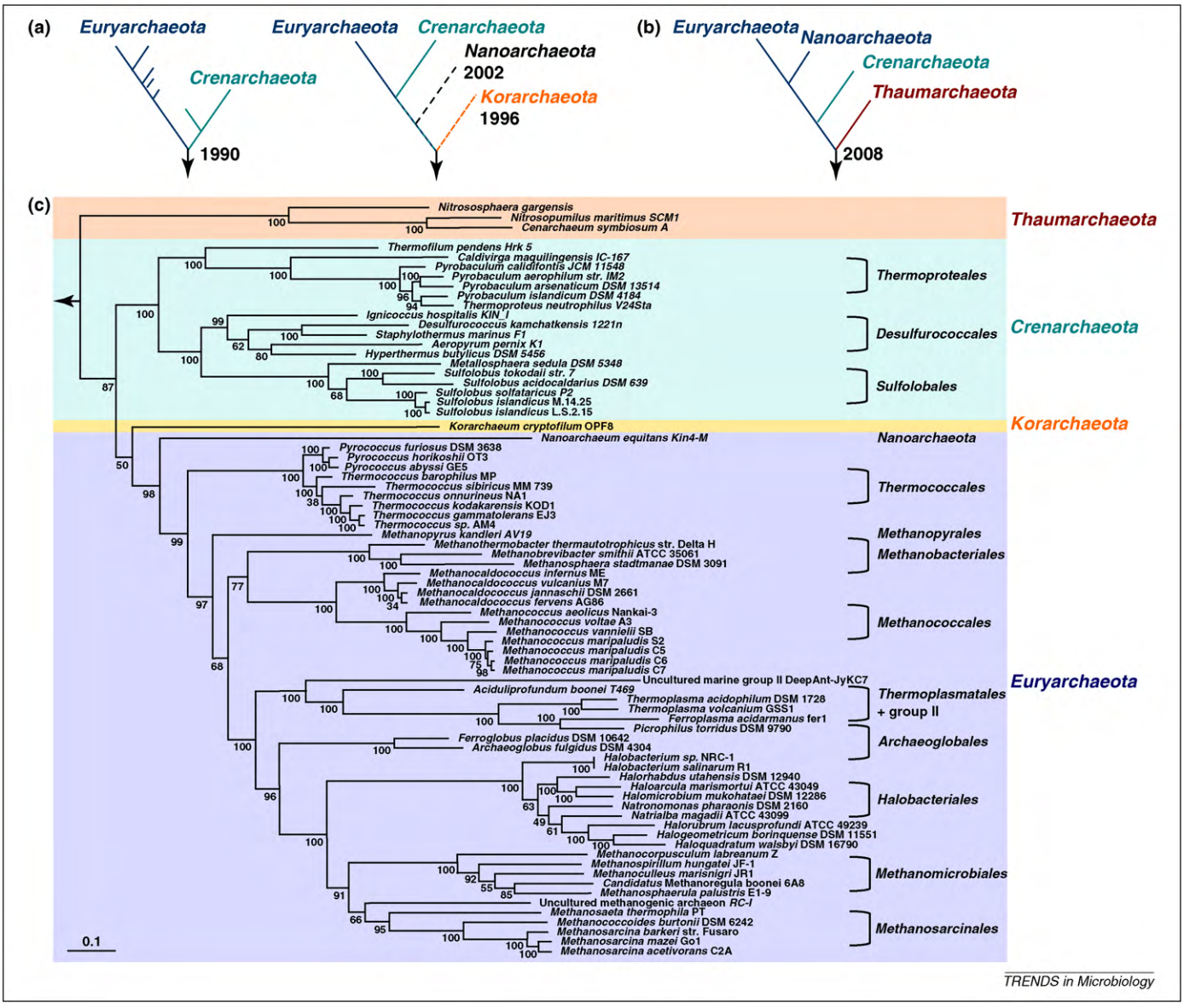

Figure 1. Phylogeny of Archaea. **(a)** Schematic 16S rRNA trees demonstrating the view of archaeal phyla over the past two decades. Since the original partition into Crenarchaeota and Euryarchaeota by Woese [2] (left), two additional archaeal phyla, Korarchaeota [8] and Nanoarchaeota [80], have been postulated (right). **(b)** In 2008, a concatenated data set of 53 ribosomal proteins placed *C. symbiosum* into a separate lineage [12]. **(c)** Rooted maximum likelihood tree of Archaea based on 53 concatenated ribosomal proteins (107 sequences, 4683 amino acid positions; eukaryotic sequences were used as an outgroup). The tree was built using PhyML [81], an LG model with γ correction (4 site categories and an estimated α parameter). The tree supports Thaumarchaeota as an independent and potentially deeply branching lineage. The tree also includes many novel genome sequences recently described for archaea. Numbers at nodes represent bootstrap values (100 replicates of the original data set). The scale bar represents 10% estimated sequence divergence. Proteins of the same strains (as identified in this tree behind each species name) were used in the other phylogenetic calculations.

rRNA-based diversity studies. Using the first available genome sequence of this group, that of *Cenarchaeum symbiosum*, a marine sponge symbiont [15,16], Brochier-Armanet *et al.* recently analysed a concatenated data set of 53 ribosomal proteins shared by Archaea and Eukarya [12]. The resulting tree, in contrast to rRNA phylogeny [17,18], not only resolved the ancient division between Crenarchaeota and Euryarchaeota, but also suggested that the 'moderate Crenarchaeota' including the AOA actually constitute a separate phylum of the Archaea that branched off before the separation of Crenarchaeota and Euryarchaeota. The name Thaumarchaeota (from the Greek word *thaumas* for wonder) was proposed for the new phylum [12]. However, the phylum-level status of the Thaumarchaeota was based on the analysis of a single genome of a marine archaeon (and a few environmental sequences) and thus the inclusion of additional AOA representing different phylogenetic lineages is important.

Novel ammonia oxidizers support the ribosomal protein trees

Gene signatures of Thaumarchaeota, including 16S rRNA genes and genes encoding key enzymes of ammonia oxidation, have been detected in moderate environments, such as soils [19–21], estuaries [22,23] and marine plankton [24,25], as well as in terrestrial hot springs and hydrothermal vent systems [26–29]. The shear numbers of these organisms in so many environments

and their phylogenetic breadth indicate that they constitute a large and ubiquitous group, with some marine planktonic lineages among the most abundant microorganisms on the planet [24]. With the capacity to oxidize ammonia to nitrite and to fix CO_2, at least some of these organisms (perhaps even most of them) are now thought to be significant for global nitrogen and carbon cycles. A pure culture of *Nitrosopumilus maritimus* from the marine group I.1a has allowed for detailed physiological analysis of an AOA [30] and for sequencing of another AOA genome [31]. Furthermore, the draft genome of *Nitrososphaera gargensis*, a moderately thermophilic ammonia oxidizer growing in stable laboratory enrichments [32], has recently been obtained (GenBank accession numbers GU797788–GU797828 and HM229997–HM230053). As revealed by phylogenetic analyses of its 16S rRNA and the A subunit of ammonia monooxygenase (Amo), which catalyses the first step in the oxidation of ammonia to nitrite, this organism is closely related to the major group of AOA detected in soils (group I.1b) [19]. It thus represents a distinct group that forms a sister lineage to the marine cluster (group I.1a), which includes *N. maritimus* and *C. symbiosum*. The availability of three AOA genome sequences (Table 1) representing two different lineages facilitated a more comprehensive test of the Thaumarchaeota hypothesis and a search for common patterns in the composition of their genomes.

An updated phylogenetic analysis of the data set used by Brochier *et al*. (53 concatenated ribosomal proteins) [12] confirms that Thaumarchaeota are a lineage that is distinct from Euryarchaeota and Crenarchaeota, with bootstrap values (BV) of 100% and 98% supporting the monophyly of Crenarchaeota and Euryarchaeota, respectively (Figure 1c). Within Euryarchaeota and Crenarchaeota, the monophyly of all archaeal orders is robustly supported (BV >95%) and their relationships are in agreement with previous work [33] suggesting that the tree is not affected by major artefacts. Concerning the phylogenetic position of Thaumarchaeota, a basal branching is observed as in previous analyses, but with slightly lower support (BV 87%) [12]. This indicates that additional data are required for better resolution of the position of Thaumarchaeota within the archaeal domain. However, this does not alter the conclusion that Thaumarchaeota represent a distinct major archaeal lineage because their emergence within either the Crenarchaeota or the Euryarchaeota was never observed in any phylogenetic analyses (see below).

A distinct set of information-processing genes in AOA

The presence and absence of genes involved in central informational processes (such as replication, transcription and translation) and genes involved in cell division were compared in all available archaeal genomes to identify phylum-specific distribution patterns. The recently published archaeal clusters of orthologous genes (arCOGs) [34] based on 41 archaeal genomes (13 crenarchaeotes including *C. symbiosum*, 27 euryarchaeotes and *Nanoarchaeum equitans*) were also considered. Several information-processing genes whose presence or absence is characteristic for Euryarchaeota and/or Crenarchaeota show a pattern in the three investigated thaumarchaeotal genomes that is distinctive from either of the two described phyla (Table 2), which points to fundamental differences in cellular processes. Thus, this gene content comparison strongly supports the Thaumarchaeota proposal, as earlier indicated on the basis of the *C. symbiosum* genome [12]. Similarly, *K. cryptophilum* exhibits a unique distribution pattern of information-processing genes. However, additional genome sequences of organisms related to *K. cryptophilum* are required before general conclusions on this provisional phylum can be made.

In the following sections we summarize and discuss the specific – and what we consider most relevant and informative – aspects of information-processing machineries in Thaumarchaeota that distinguish them from other archaeal phyla.

Translation: Thaumarchaeota have a specific set of ribosomal proteins

All domains of life share 34 [35] ribosomal (r-) protein families, whereas Archaea and Eukarya have another 33 families in common (see below) [35]. By contrast, neither Archaea nor Eukarya share an r-protein family exclusively with Bacteria. In addition to the 34 universal r-proteins, Bacteria contain 23 domain-specific families. Eukarya encode 11 domain-specific families, whereas only the r-protein family LXa is exclusively found in Archaea [35,36]. Based on these observations it was suggested that the archaeal ribosome represents a simplified version of the

Table 1. General features of the three AOA used for comparative genomic and phylogenetic analyses

	Nitrososphaera gargensis	Cenarchaeum symbiosum A	Nitrosopumilus maritimus SCM1
Habitat	Garga hot spring, enriched at 46 °C	Symbiont of the marine sponge Axinella mexicana	Tropical marine tank, 21–23 °C
Metabolism	Autotrophic ammonia oxidizer	Putative ammonia oxidizer	Autotrophic ammonia oxidizer
Genome size	>2.6 Mb (draft genome[a])	2.05 Mb	1.65 Mb
GC content	48%	57.7%	34.2%
Predicted genes	>3200[b]	2066[c]	1840[c]
Ribosomal RNA	1× 16S-23S rRNA operon; 1× 5S rRNA	1× 16S-23S rRNA operon; 1× 5S rRNA	1× 16S-23S rRNA; 1× 5S rRNA
Phylogenetic affiliation	Group I.1b AOA (soil group)	Group I.1a AOA (marine group)	Group I.1a AOA (marine group)
Reference	[32]	[15]	[31]

[a]Evidence of a non-redundant and almost complete draft genome sequence (two gaps left at the time of paper submission) of *N. gargensis* includes the presence of all tRNA synthetases, all r-proteins (found in the genomes of the closely related *C. symbiosum* and *N. maritimus*) and all RNA polymerase subunits in one gene copy. In addition, *N. gargensis* contains a comparable number of archaeal core-COGs as found in the genomes of other Archaea (e.g. 210 *C. symbiosum*, 211 *N. maritimus*, and 215 *N. gargensis*). GenBank accession numbers of the *N. gargensis* genes discussed are GU797786–GU797828 and HM22999–HM230053.
[b]According to the automated annotation number (ORFs >150 bp).
[c]According to NCBI (http://www.ncbi.nlm.nih.gov/.

Table 2. Phylum-specific genes involved in archaeal central information-processing machineries[a]

		Thaumarchaeota			*Euryarchaeota*	*Nanoarchaeota*	*Crenarchaeota*	*Korarchaeota*
Ribosomal proteins	**COG**	***C. sym.***	***N. mar.***	***N. gar.*[a]**				
r-protein S25e	4901	+	+	+	-	-	+	+
r-protein S26e	4830	+	+	+	-	+	+	+
r-protein S30e	4919	+	+	+	-	-	+	+
r-protein L13e	4352	-	-	(+)	-	-	+	+
r-protein L14e	2163	-	-	-	+ (some)	+	+	+
r-protein L34e	2174	-	-	-	+ (some)	+	+	+
r-protein L38e	-	-	-	-	-	-	+ /-	-
r-protein L29p	0255	+	+	+	+	+	+	-
r-protein LXa	2157	-	-	-	+ (most)	+	+	-
Translation								
RNA polG (=rpo8)	-	-	-	-	-	-	+	+
RNA pol A–1 ORF	0086	+	+	+	split	split	split	+
RNA pol B–1 ORF	0085	+	+	+	most split	split	+	+
Rpc34	5111	+	+	+	+ (many)	-	+	-
MBF1	1813	-	-	+	+	+	+	+
EIF1	4888	-	-	-	-	-	+	+
Replication								
DNA pol D (small)	1311	+	+	+	+	+	-	+
DNA pol D (large)	1933	+	+	+	+	+	-	+
RPA (Eury-like)	(1599)	+	+	+	+	+	- (not T. p.)	+
RPA / SSB (Cren-like, single OB-fold)	1599	+	+	+	+ (some)	-	+ (most)	+
> 1 PCNA homolog	0592	-	-	-	-	-	+ (not T. p)	-
Topoisomerases								
Topo IB	3569	+	+	+	-	-	-	-
Topo IA	0550	- CTD	- CTD	+	+	+	+	+
Reverse gyrase	1110	-	-	-	+ (HT)	+	+	+
Topo IIA (SU A+B)	0187/8	-	-	-	+ (17 org.)	-	-	-
Cell Division								
ESCRT-III (CdvB)	5491	+	+	+	-	-	+ CdvABC [b]	-
Vps4 (CdvC)	0464	+	+	+	-	-	+ CdvABC [b]	-
CdvA	-	+	+	+	-	-	+ CdvABC [b]	-
Fts Z	0206	+	+	+	+	+	-	+
Smc, Chromosome segregation ATPase	1196	+	+	+	+	-	-	+
ScpA + ScpB	1354+ 1386	+	+	+	+ (many)	-	-	+
Histones								
Histones (H3/H4)	2036	+	+	+	+	+	- (exc. two)	+
Repair								
Hef nuclease	1111+ 1948	-	-	-	+	+	-	-
ERCC4-type nuclease	1948	+	+	+	-	-	+	+
ERCC4-type helicase	1111	+	+	+	-	-	-	-
RadB	-	-	-	-	+	-	-	-
HSP70/DnaK, GrpE, Hsp40/DnaJ	0443, 0576, 0326	+	+	+	+	-	-	-
Hsp90-like domain	0326	+	-	+	-	-	-	-
UvrABC	0178, 0556, 0322	+	+	+	+ (mesophiles)	-	-	-

[a]Abbreviations: +, present; (+), potentially present; –, absent; CTD, C-terminal domain; HT, hyperthermophiles; exc, except for; Eury, Euryarchaeota; Cren, Crenarchaeota; org, organisms; *T. p., Thermophilum pendens; C. sym., Cenarchaeum symbiosum; N. mar., Nitrosopumilus maritimus; N. gar, Nitrososphaera gargensis.*

[b]GenBank accession numbers for the *N. gargensis* genes discussed are GU797788–GU797828.

[c]cdvABC operon not present in Thermoproteales.

eukaryotic one. In contrast to Bacteria and Eukarya, which acquired further r-proteins independently, it was suggested that the archaeal ribosome (apart from LXa) has undergone reductive evolution [35]. For example, the r-proteins L35ae and L41e are absent in almost all archaea. In addition, Euryarchaeota have lost at least five r-protein families, L38e, L13e, S25e, S26e and S30e. An even smaller number of r-proteins are encoded by later diverging euryarchaeotes such as Halobacteriales [35]. Thaumarchaeota show a lineage-specific pattern that is clearly distinct from those of Crenarchaeota and Euryarchaeota. For example, all three AOA genomes possess the r-proteins S26e, S25e and S30e, which are absent in all euryarchaeotes, but lack the r-proteins L14e and L34e, which are found in all crenarchaeotes, *K. cryptophilum* and *N. equitans*. *C. symbiosum* and *N. maritimus* also lack a homologue of L13e, which is found in all crenarchaeotes and *K. cryptophilum*. A putative L13e homologue was identified in the genome of *N. gargensis*, which might indicate that it was lost by reductive evolution only in the marine thaumarchaeotal lineage, as might be true for gene expression factor MBF1 [37], which is present in *N. gargensis* only. By contrast, putative homologues of the ribosomal protein S24e, which originally could not be identified in *C. symbiosum* [12], were detected in all Thaumarchaeot*a* when the *N. gargensis* homologue was used as a query sequence (GU797793 for *N. gargensis*, Nmar0535, CENSYa0426). Interestingly, the Archaea-specific r-protein LXa is missing in both Thaumarchaeota and Korarchaeota. This might indicate a rather late acquisition of this protein in the archaeal domain because it is only present in Euryarchaeota and Crenarchaeota.

Transcription: the RNA polymerase of Thaumarchaeota testifies to an ancient lineage

In all three domains of life, DNA-dependent RNA polymerases (RNAPs) are composed of several subunits. Whereas the bacterial enzymes contain only five subunits, each of the three distinct eukaryotic RNAPs (I–III) consists of at least 12 subunits [38]. Archaea contain a single polymerase with 10–14 subunits, of which most are homologues of subunits of the eukaryotic polymerases [39]. Even subunit 8, until recently considered to be specific for Eukarya, is present in Crenarchaeota and Korarchaeota [40,41]. In addition, a homologue of the loosely associated eukaryotic RNAPIII-specific subunit 34 (rpc34) was detected in Crenarchaeota, Thaumarchaeota and late-diverging euryarchaeotal orders such as Archaeoglobales, Methanomicrobiales and Halobacteriales [42]. This suggests the presence of a rather complex RNAP in the last common ancestor of Archaea and Eukarya and represents a well-known example of the close relationship of information-processing machineries in these two domains of life [43]. Interestingly, in contrast to Bacteria

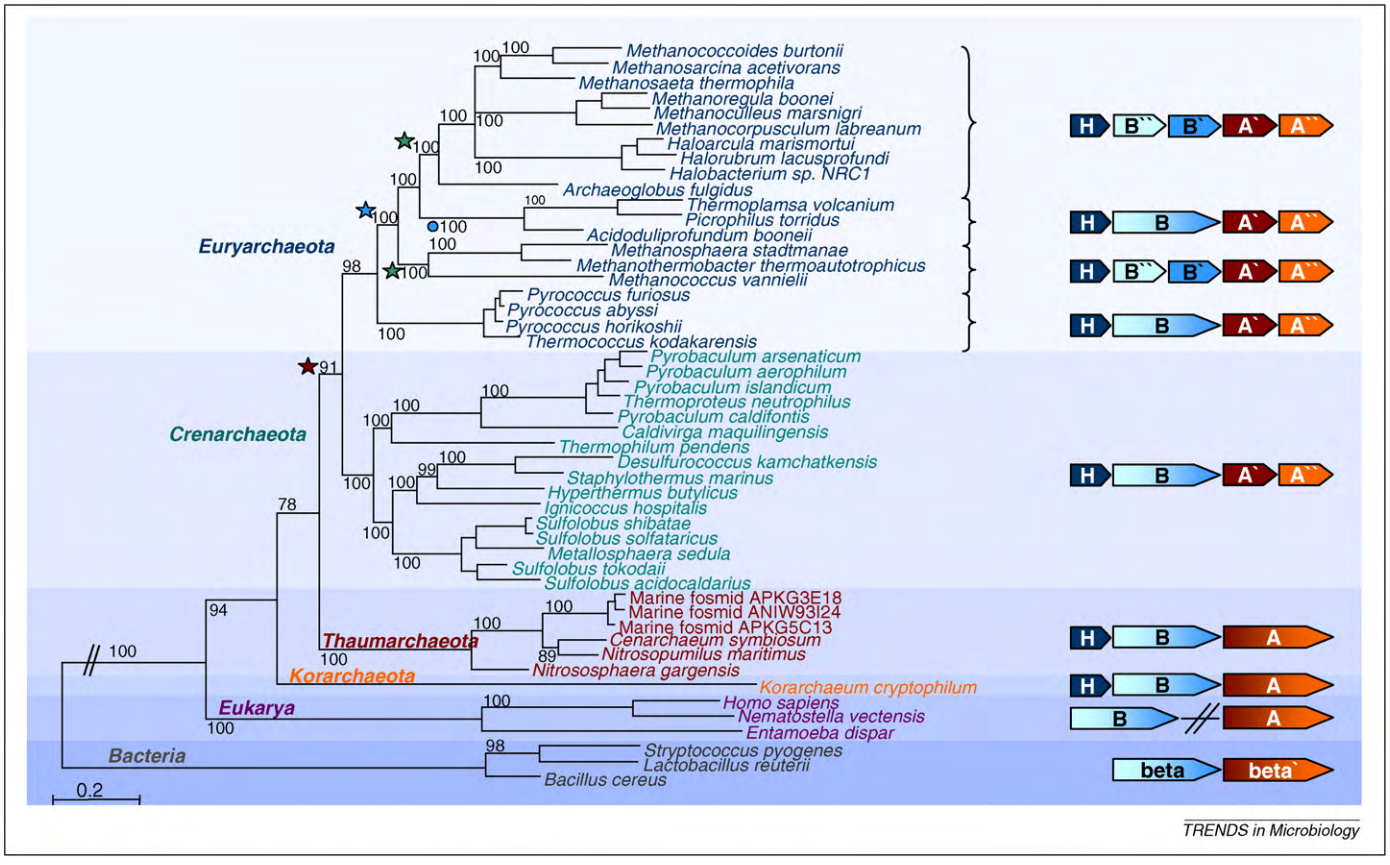

Figure 2. Phylogenetic tree based on archaeal RpoA proteins and schematic gene cluster encoding RNA polymerase subunits A, B and H. Bayesian interference phylogeny of RpoA proteins was performed after concatenation of crenarchaeotal and euryarchaeotal split archaeal RpoA subunits and alignment with the unsplit RpoA proteins of Thaumarchaeota and Korarchaeota and bacterial and eukaryotic homologues as outgroups (1051 amino acid positions, 30% conservation filter). In agreement with the unsplit rpoA genes, the phylogenetic analysis suggests early divergence of Korarchaeota and Thaumarchaeota, both of which are highly supported as distinct lineages. The red star indicates a potential split in the ancestral *rpoA* gene; green stars indicate potential splits in the *rpoB* gene; an alternative possibility is indicated by the blue star and blue circle, denoting potential split and fusion events for the *rpoB* gene, respectively. The scale bar represents 20% estimated sequence divergence.

and Eukaryota, which contain a single gene for RNAP subunit A (rpo/rpbA), in both Crenarchaeota and Euryarchaeota the A subunit is encoded by two separate genes that occur in a gene cluster encoding subunits H, B and A′ and A″ [44]. Protein sequence alignments have shown that the ends of A′ and the starts of A″ subunits coincide [43]. By contrast, investigation of this gene locus in *K. cryptophilum* [9] and Thaumarchaeota [15,38] revealed only a single *rpoA* gene.

Thus, the most parsimonious explanation for the split in subunit A would be that it occurred only once in evolution. This is supported by RpoA phylogeny, whereby Thaumarchaeota and Korarchaeota form two distinct lineages that branch off before Euryarchaeota and Crenarchaeota, possibly predating the rpoA split (Figure 2).

Replication and DNA-binding proteins: Thaumarchaeota are distinct from Crenarchaeota

Like Euryarchaeota and Korarchaeota, the three thaumarchaeotes contain two different archaeal DNA polymerase families (PolB and PolD). This is in contrast to Crenarchaeota, which only encode polymerases of the B family, of which all archaea contain at least one homologue (often more) [34,45]. Phylogenetic investigation of the large and small subunits of PolD revealed two monophyletic groups consisting of Euryarchaeota and Thaumarchaeota and a separated korarchaeotal branch (Figure 3), indicating early diversification of this gene in the archaeal domain and its potential loss in the lineage leading to Crenarchaeota.

Proliferating cell nuclear antigen (PCNA) is the archaeal sliding clamp involved in DNA replication [46]. Whereas only one homologue of this protein is found in Euryarchaeota, Korarchaeota and Thaumarchaeota, all Crenarchaeota (with one exception) contain two or three copies of this gene, which might be because of early gene duplication events in the respective lineages (Figure 4). Interestingly, PCNA proteins of Thaumarchaeota form a monophyletic and early-diverging archaeal group in phylogenetic analyses when eukaryotic sequences are used as an outgroup (Figure 4), whereas the korarchaeotal PCNA protein does not form a distinct lineage, indicating that it might have been acquired laterally.

Chromatin proteins with a histone fold related to eukaryotic histones H3 and H4 are encoded by most euryarchaeotes [47], except for Thermoplasmatales [48], and were long considered to be absent in Crenarchaeota. Their finding in the recently obtained genomes of *K. cryptophilum* and the crenarchaeotes *Thermophilum pendens* and *Caldivirga maquilensis* [47], as well as in environmental sequences and Thaumarchaeota [49], supports the previously proposed idea that histones were present in the last common ancestor of Archaea and Eukarya [47].

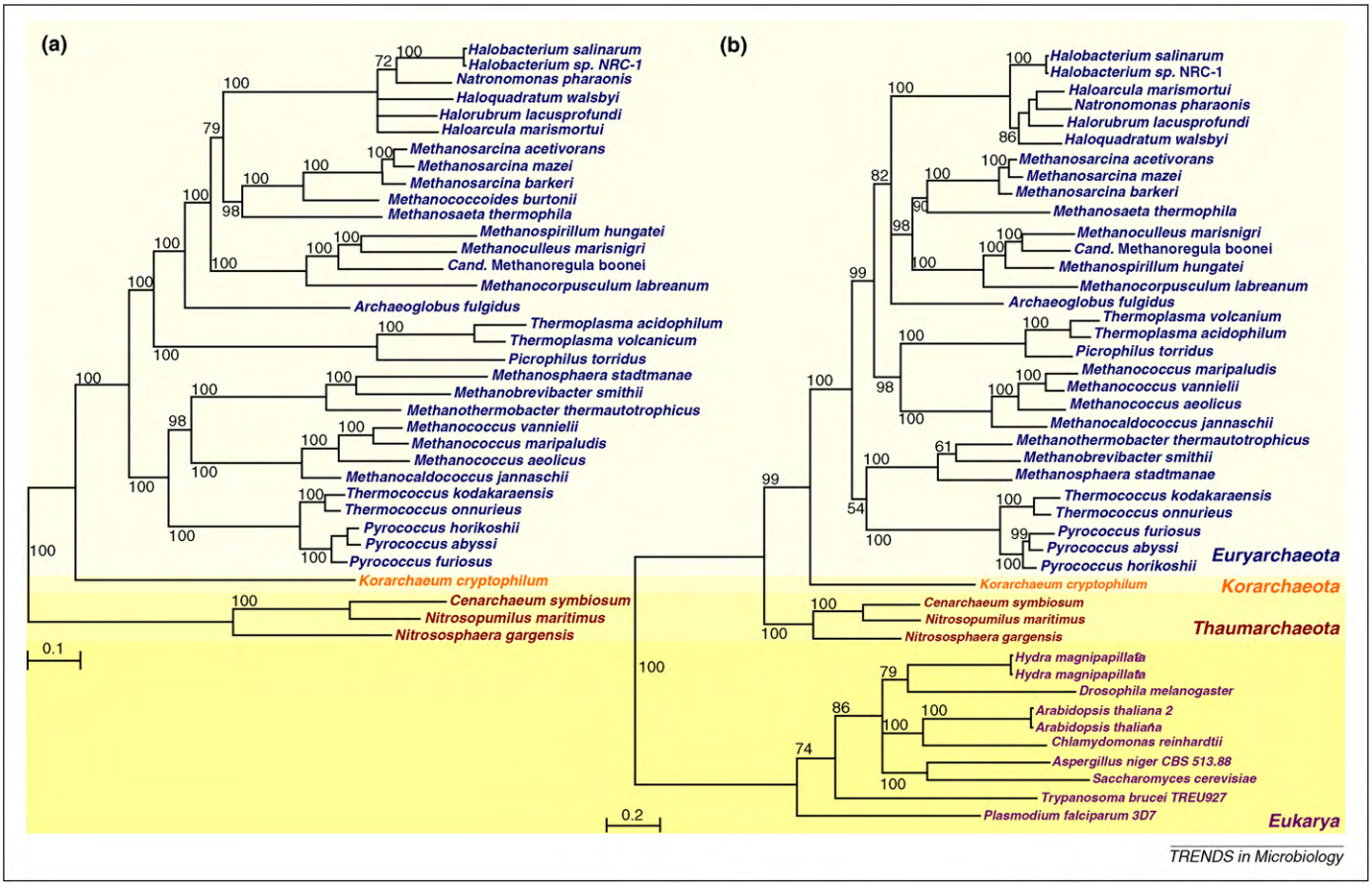

Figure 3. Phylogeny of archaeal polymerase D. Bayesian inference phylogeny of **(a)** the large subunit of DNA polymerase D (scale bar indicates 10% estimated sequence divergence) and **(b)** the small subunit of DNA polymerase D (scale bar indicates 20% estimated sequence divergence). Calculations were based on 944 and 418 aligned amino acid positions (30% conservation filter) for the large and small subunits, respectively. Numbers represent Bayesian likelihood values. As only archaea contain the complete homologue of the large polymerase D, no outgroup could be included in **(a)**.

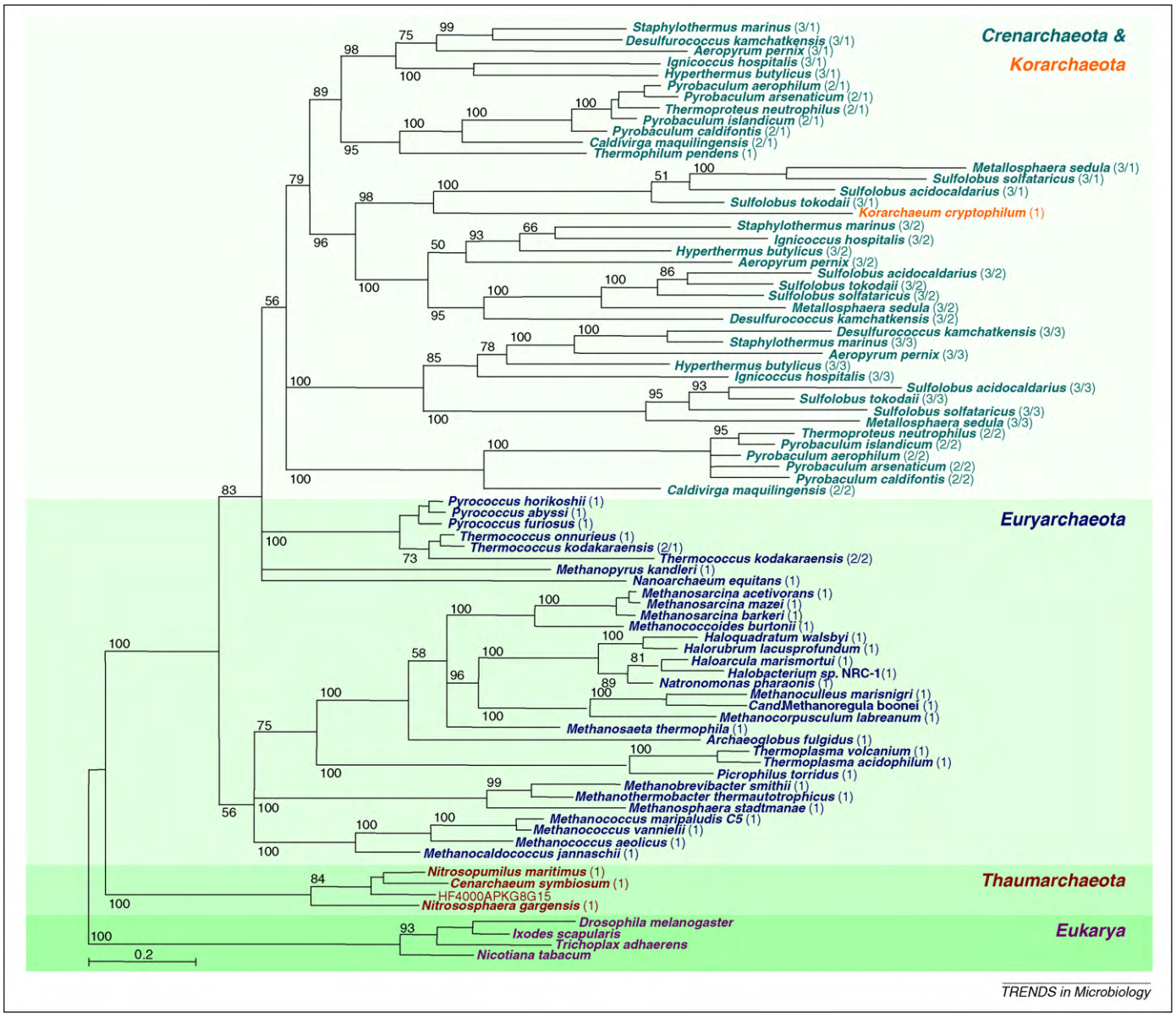

Figure 4. Phylogeny of PCNA homologues in Archaea. Bayesian inference phylogeny of PCNA based on 177 aligned amino acid positions (30% conservation filter). Numbers at nodes represent Bayesian likelihood values. The first number in parentheses shows the number of PCNA homologues per genome. Except for *T. pendens*, Crenarchaeota contain at least two or three copies, as shown by the second number in parentheses. The scale bar refers to 20% estimated sequence divergence.

Topoisomerases: more surprises from Thaumarchaeota

DNA topoisomerases play a crucial role in many cellular processes that involve DNA because they solve topological problems of circular and/or large linear double helices that arise during replication, transcription, recombination and chromosome segregation [50]. Even though topoisomerases of diverse families are known to have been transferred horizontally several times in evolution, they might still serve as important genomic marker genes defining specific phylogenetic lineages [51]. Interestingly, Thaumarchaeota contain genes for topoisomerase IB, which is absent from the genomes of all other archaea described but occurs in Bacteria, Eukarya and some eukaryotic DNA viruses (e.g. mimivirus) [52]. Phylogenetic analysis of topoisomerase IB (with *C. symbiosum* included) showed a clear separation of two deeply branching monophyletic lineages, namely a eukaryotic–archaeal and a viral–bacterial one [52]. Within the eukaryotic–archaeal branch, topoisomerase IB of Thaumarchaeota (including sequences derived from recently published archaeal soil fosmids [53]) forms a monophyletic lineage with high bootstrap support that represents a sister group to the eukaryotic homologues [52]. Thus, topoisomerase IB does not seem to have been acquired recently by the Thaumarchaeota via lateral gene transfer from Eukarya, but rather might have already been present in the LACA. According to this scenario, Thaumarchaeota would have retained topoisomerase IB during evolution, whereas all other archaea have lost this enzyme but retained the functionally related topoisomerase IA.

The genome of *N. gargensis* is particularly informative with respect to the potential ancestral distribution of topoisomerases. It is the only known archaeal genome that encodes both topoisomerase IA and topoisomerase IB. The topoisomerase IA gene from *N. gargensis* is most closely related to a homologue in the soil fosmid clone 54d9 [54].

This fosmid originated, according to a linked 16S rRNA gene, from the AOA soil group I.1b to which *N. gargensis* also belongs. Furthermore, the presence of a C-terminal domain of topoisomerase IA in both *C. symbiosum* and *N. maritimus* strengthens the hypothesis that the common ancestor of Thaumarchaeota contained topoisomerase IA as well as topoisomerase IB.

Reverse gyrase (TopoII) is a hallmark of hyperthermophilic and extremely thermophilic organisms [55], even though it has been detected in some thermophilic bacteria [56], but its absence in the moderate thaumarchaeota is not surprising and might represent an environmental rather than a phylogenetic signal. Environmental selection might also account for the presence of the UvrABC repair system and the chaperonin GroEL in Thaumarchaeota and their absence in most crenarchaeota and hyperthermophilic euryarchaeota [57]. In this respect, the genome of the first enriched thermophilic ammonia-oxidizer *Nitrosocaldus yellowstonii* [26] will be of special interest, because this might help to distinguish genes that were vertically inherited from those acquired by lateral gene transfer or lost in the course of adaptation to high or mesophilic temperatures, respectively.

Cell division: genes for two modes of cell division in Thaumarchaeota?

The tubulin-related GTPase FtsZ represents the major cell-division protein in most bacteria and in Euryarchaeota [58]. Crenarchaeota, however, do not encode FtsZ proteins, but seem to exploit a different mode of cell division [59–60] that involves three proteins, termed CdvA, CdvB and CdvC [59]. CdvA is a protein with a coiled-coil domain, CdvB is homologous to an endosomal sorting complex required for transport (ESCRT)-III like protein and CdvC is homologous to the AAA+type ATPase Vps4 [60]. The homologous proteins in Eukarya are involved in multivesicular body formation, cell division and virus budding [61–63]. The proposed crenarchaeotal cell division proteins (CdvA–C) occur in an operon in all sequenced members of Sulfolobales and Desulfurococcales, whereas they are absent in Korarchaeota and Euryarchaeota [59]. Interestingly, organisms of the crenarchaeotal order Thermoproteales encode neither these novel cell-division proteins nor FtsZ. Rather, the three thaumarchaeotes contain homologues of both systems [59–61] and might thus reflect the ancestral distribution of those genes in the LACA. Even though homologues of the three newly identified crenarchaeotal cell-division genes (*cdvA*, *cdvB*, *cdvC*) are present in Thaumarchaeota, they do not occur as clusters in the genomes but are rather randomly distributed. In a phylogenetic analysis of CdvA, Thaumarchaeota form a monophyletic and only distantly related sister group to Crenarchaeota (Figure 5), suggesting that the presence of *cdvA* in these two phyla does not result from a recent lateral gene transfer. Furthermore, the CdvB/ESCRT-III-encoding genes of Thaumarchaeota and Crenarchaeota differ. In Crenarchaeota, several CdvB/ESCRT-III-encoding genes are usually present and one of these orthologues is always associated with the CdvABC operon. This operon-associated *cdvB* gene has a conserved C-terminal extension containing a winged-helix domain (~50 aa in length), which is missing in the other CdvB homologues [60]. None of the three *cdvB* orthologues in *N. maritimus* and *C. symbiosum* or the five *cdvB* orthologues in *N. gargens*is contains this C-terminal domain.

Thaumarchaeota encode homologues of both archaeal cell-division systems (from Euryarchaeota and Crenarchaeota), so it remains to be shown experimentally which system is actually used during growth or whether proteins of both cell-division machineries interact. Such analyses might be particularly fruitful because they could shed light on the mode of cell division in LACA and potentially on the evolution of cell division mechanisms in Eukarya [64].

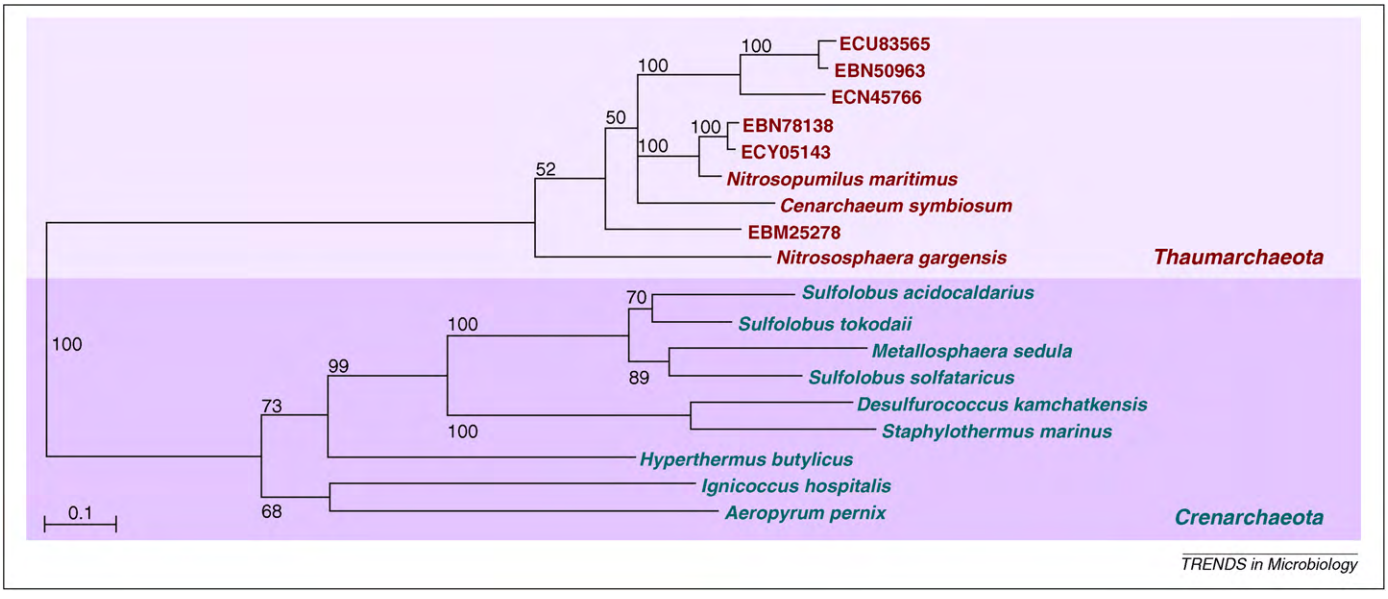

Figure 5. Bayesian inference phylogeny of the archaeal cell division protein CdvA. A 30% conservation filter was used. Calculation was stopped after 50 000 generations because a clear dichotomy between Crenarchaeota and Thaumarchaeota was apparent. The scale bar refers to 10% estimated sequence divergence. Numbers at nodes represent Bayesian likelihood values.

Conclusions

Phylogenetic analyses of several different marker genes and comparative gene content analysis of the genomes of three AOA suggests that they are indeed clearly distinct from all other archaea. From an evolutionary point of view, they form a separate and frequently deep-branching lineage within the domain Archaea. Although the exact placement of this lineage, in particular in relation to the Korarchaeota, cannot be unambiguously resolved before more genome sequences of both groups are available, there is no doubt that the AOA do not belong to Crenarchaeota, as initially suggested by comparative 16S rRNA sequence analysis. Consequently, these findings provide compelling support for the recently proposed assignment of the AOA to the new archaeal phylum Thaumarchaeota [12].

There are indications that Thaumarchaeota might also contain a specific membrane lipid: the tetraether lipid crenarchaeol was originally associated only with marine planktonic archaea [65], but was later found in many terrestrial hot springs [27,66–70] and was recently identified as a component of the lipid membranes of three AOA, *N. maritimus* [71], *N. yellowstonii* [26] and *N. gargensis* [72]. No crenarchaeote has been shown to harbour crenarchaeol; it might rather represent a Thaumarchaeota-specific membrane lipid and could possibly be called thaumarchaeol instead.

The recognition of a third archaeal phylum, the Thaumarchaeota, opens new perspectives regarding our view of the evolution of archaea and early cellular life forms. For example, it allows for suggestions on the occurrence of certain features in LACA. According to the gene content of the AOA in comparison with other archaeal phyla, it seems likely that LACA contained, for example, the complete set of r-proteins that are shared with Eukarya, a protein with a histone fold, a single gene for RNA polymerase A, an *ftsZ* homologue as well as *cdv* genes, topoisomerase IA as well as topoisomerase IB, and DNA polymerase D and DNA polymerase B. This indicates that LACA was probably more complex than anticipated, as suggested in other studies [34,35,73].

Investigation of Thaumarchaeota has already led to several unexpected discoveries, including the recognition that archaea are far more abundant in moderate aerobic habitats than originally anticipated [19,24] and that they contribute to nitrification [74–76] and exhibit extraordinary physiological properties [77]. Genomic and biochemical analyses indicate that these organisms almost certainly harbour even more wonders [78,79] that will be fully revealed only when their biochemical and cellular features are characterized in more detail.

Acknowledgements

This work was supported by a DOC-fForte fellowship to A.S. and a DOC fellowship to R.H. of the Austrian Academy of Sciences.

References

1 Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5088–5090
2 Woese, C.R. *et al.* (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U. S. A.* 87, 4576–4579
3 Cox, C.J. *et al.* (2008) The archaebacterial origin of eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 105, 20356–20361
4 Embley, T.M. and Martin, W. (2006) Eukaryotic evolution, changes and challenges. *Nature* 440, 623–630
5 Poole, A. and Penny, D. (2007) Eukaryote evolution: engulfed by speculation. *Nature* 447, 913
6 Garrett, R.A. and Klenk, H.P. (2007) *Archaea: Evolution, Physiology, and Molecular Biology*, Blackwell Publishing
7 Garrity, G.M. *et al.*, eds (2001) *Bergey's Manual of Systematic Bacteriology*, Springer
8 Barns, S.M. *et al.* (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc. Natl. Acad. Sci. U. S. A.* 93, 9188–9193
9 Elkins, J.G. *et al.* (2008) A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl. Acad. Sci. U. S. A.* 105, 8102–8107
10 Huber, H. *et al.* (2003) The phylum Nanoarchaeota: present knowledge and future perspectives of a unique form of life. *Res. Microbiol.* 154, 165–171
11 Brochier, C. *et al.* (2005) Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to Thermococcales? *Genome Biol.* 6, R42
12 Brochier-Armanet, C. *et al.* (2008) Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat. Rev. Microbiol.* 6, 245–252
13 DeLong, E.F. (1992) Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. U. S. A.* 89, 5685–5689
14 Fuhrman, J.A. *et al.* (1992) Novel major archaebacterial group from marine plankton. *Nature* 356, 148–149
15 Hallam, S.J. *et al.* (2006) Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc. Natl. Acad. Sci. U. S. A.* 103, 18296–18301
16 Hallam, S.J. *et al.* (2006) Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. *PLoS Biol.* 4, e95
17 Robertson, C.E. *et al.* (2005) Phylogenetic diversity and ecology of environmental Archaea. *Curr. Opin. Microbiol.* 8, 638–642
18 Schleper, C. *et al.* (2005) Genomic studies of uncultivated Archaea. *Nat. Rev. Microbiol.* 3, 479–488
19 Leininger, S. *et al.* (2006) Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* 442, 806–809
20 He, J. *et al.* (2007) Quantitative analyses of the abundance and composition of ammonia-oxidizing bacteria and ammonia-oxidizing Archaea of a Chinese upland red soil under long-term fertilization practices. *Environ. Microbiol.* 9, 2364–2374
21 Tourna, M. *et al.* (2008) Growth, activity and temperature responses of ammonia-oxidizing Archaea and Bacteria in soil microcosms. *Environ. Microbiol.* 10, 1357–1364
22 Mosier, A.C. and Francis, C.A. (2008) Relative abundance and diversity of ammonia-oxidizing Archaea and Bacteria in the San Francisco Bay estuary. *Environ. Microbiol.* 10, 3002–3016
23 Sahan, E. and Muyzer, G. (2008) Diversity and spatio-temporal distribution of ammonia-oxidizing Archaea and Bacteria in sediments of the Westerschelde estuary. *FEMS Microbiol. Ecol.* 64, 175–186
24 Karner, M.B. *et al.* (2001) Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* 409, 507–510
25 Wuchter, C. *et al.* (2006) Archaeal nitrification in the ocean. *Proc. Natl. Acad. Sci. U. S. A.* 103, 12317–12322
26 de la Torre, J.R. *et al.* (2008) Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. *Environ. Microbiol.* 10, 810–818
27 Reigstad, L.J. *et al.* (2008) Nitrification in terrestrial hot springs of Iceland and Kamchatka. *FEMS Microbiol. Ecol.* 64, 167–174
28 Wang, S. *et al.* (2009) Diversity and abundance of ammonia-oxidizing Archaea in hydrothermal vent chimneys of the Juan de Fuca ridge. *Appl. Environ. Microbiol.* 75, 4216–4220
29 Zhang, C.L. *et al.* (2008) Global occurrence of archaeal *amoA* genes in terrestrial hot springs. *Appl. Environ. Microbiol.* 74, 6417–6426
30 Könneke, M. *et al.* (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 437, 543–546
31 Walker, C.B. *et al.* (2010) *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine Crenarchaea. *Proc. Natl. Acad. Sci. U. S. A.* 107, 8818–8823

32 Hatzenpichler, R. *et al.* (2008) A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. *Proc. Natl. Acad. Sci. U. S. A.* 105, 2134–2139
33 Gribaldo, S. and Brochier-Armanet, C. (2006) The origin and evolution of Archaea: a state of the art. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 1007–1022
34 Makarova, K.S. *et al.* (2007) Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of Archaea. *Biol. Direct* 2, 33
35 Lecompte, O. *et al.* (2002) Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res.* 30, 5382–5390
36 Ramirez, C. *et al.* (1991) A small basic ribosomal protein from the extreme thermophilic archaebacterium *Sulfolobus solfataricus* that has no equivalent in *Escherichia coli*. *FEBS Lett.* 284, 39–41
37 de Koning, B. *et al.* (2009) Role of multiprotein bridging factor 1 in Archaea: bridging the domains? *Biochem. Soc. Trans.* 37, 52–57
38 Kwapisz, M. *et al.* (2008) Early evolution of eukaryotic DNA-dependent RNA polymerases. *Trends Genet.* 24, 211–215
39 Langer, D. *et al.* (1995) Transcription in Archaea: similarity to that in Eucarya. *Proc. Natl. Acad. Sci. U. S. A.* 92, 5768–5772
40 Koonin, E.V. *et al.* (2007) Orthologs of the small RPB8 subunit of the eukaryotic RNA polymerases are conserved in hyperthermophilic Crenarchaeota and "Korarchaeota". *Biol. Direct* 2, 38
41 Korkhin, Y. *et al.* (2009) Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure. *PLoS Biol.* 7, e102
42 Blombach, F. *et al.* (2009) Identification of an ortholog of the eukaryotic RNA polymerase III subunit RPC34 in Crenarchaeota and Thaumarchaeota suggests specialization of RNA polymerases for coding and non-coding RNAs in Archaea. *Biol. Direct* 4, 39
43 Pühler, G. *et al.* (1989) Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. *Proc. Natl. Acad. Sci. U. S. A.* 86, 4569–4573
44 Werner, F. (2008) Structural evolution of multisubunit RNA polymerases. *Trends Microbiol.* 16, 247–250
45 Cann, I.K. and Ishino, Y. (1999) Archaeal DNA replication: identifying the pieces to solve a puzzle. *Genetics* 152, 1249–1267
46 Barry, E.R. and Bell, S.D. (2006) DNA replication in the Archaea. *Microbiol. Mol. Biol. Rev.* 70, 876–887
47 Sandman, K. and Reeve, J.N. (2006) Archaeal histones and the origin of the histone fold. *Curr. Opin. Microbiol.* 9, 520–525
48 Daniels, J.P. *et al.* (2009) Identification of a crenarchaeal orthologue of Elf1: implications for chromatin and transcription in Archaea. *Biol. Direct* 4, 24
49 Cubonova, L. *et al.* (2005) Histones in Crenarchaea. *J. Bacteriol.* 187, 5482–5485
50 Forterre, P. and Gadelle, D. (2009) Phylogenomics of DNA topoisomerases: their origin and putative roles in the emergence of modern organisms. *Nucleic Acids Res.* 37, 679–692
51 Forterre, P. *et al.* (2007) Origin and evolution of DNA topoisomerases. *Biochimie* 89, 427–446
52 Brochier-Armanet, C. *et al.* (2008) A DNA topoisomerase IB in Thaumarchaeota testifies for the presence of this enzyme in the last common ancestor of Archaea and Eucarya. *Biol. Direct* 3, 54
53 Bartossek, R. *et al.* (2010) Homologues of nitrite reductases in ammonia-oxidizing Archaea: diversity and genomic context. *Environ. Microbiol.* 12, 1075–1088
54 Treusch, A.H. *et al.* (2005) Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic Crenarchaeota in nitrogen cycling. *Environ. Microbiol.* 7, 1985–1995
55 Forterre, P. (2002) A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. *Trends Genet.* 18, 236–237
56 Brochier-Armanet, C. and Forterre, P. (2007) Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene transfers. *Archaea* 2, 83–93
57 Kelman, Z. and White, M.F. (2005) Archaeal DNA replication and repair. *Curr. Opin. Microbiol.* 8, 669–676
58 Margolin, W. (2005) FtsZ and the division of prokaryotic cells and organelles. *Nat. Rev. Mol. Cell Biol.* 6, 862–871
59 Lindaas, A.C. *et al.* (2008) A unique cell division machinery in the Archaea. *Proc. Natl. Acad. Sci. U. S. A.* 105, 18942–18946
60 Samson, R.Y. *et al.* (2008) A role for the ESCRT system in cell division in Archaea. *Science* 322, 1710–1713
61 Samson, R.Y. and Bell, S.D. (2009) Ancient ESCRTs and the evolution of binary fission. *Trends Microbiol.* 17, 507–513
62 Williams, R.L. and Urbe, S. (2007) The emerging shape of the ESCRT machinery. *Nat. Rev. Mol. Cell Biol.* 8, 355–368
63 Winter, V. and Hauser, M.T. (2006) Exploring the ESCRTing machinery in eukaryotes. *Trends Plant Sci.* 11, 115–123
64 Bernander, R. (2003) The archaeal cell cycle: current issues. *Mol. Microbiol.* 48, 599–604
65 Damsté, J.S. *et al.* (2002) Crenarchaeol: the characteristic core glycerol dibiphytanyl glycerol tetraether membrane lipid of cosmopolitan pelagic Crenarchaeota. *J. Lipid Res.* 43, 1641–1651
66 Pearson, A. *et al.* (2004) Nonmarine crenarchaeol in Nevada hot springs. *Appl. Environ. Microbiol.* 70, 5229–5237
67 Pearson, A. *et al.* (2008) Factors controlling the distribution of archaeal tetraethers in terrestrial hot springs. *Appl. Environ. Microbiol.* 74, 3523–3532
68 Pitcher, A. *et al.* (2009) *In situ* production of crenarchaeol in two California hot springs. *Appl. Environ. Microbiol.* 75, 4443–4451
69 Schouten, S. *et al.* (2007) Archaeal and bacterial glycerol dialkyl glycerol tetraether lipids in hot springs of Yellowstone National Park (USA). *Appl. Environ. Microbiol.* 73, 6181–6191
70 Zhang, C.L. *et al.* (2006) Thermophilic temperature optimum for crenarchaeol synthesis and its implication for archaeal evolution. *Appl. Environ. Microbiol.* 72, 4419–4422
71 Schouten, S. *et al.* (2008) Intact membrane lipids of Candidatus "*Nitrosopumilus maritimus*", a cultivated representative of the cosmopolitan mesophilic Group I Crenarchaeota. *Appl. Environ. Microbiol.* 74, 2433–2440
72 Pitcher, A. *et al.* (2009) Crenarchaeol dominates the membrane lipids of Candidatus *Nitrososphaera gargensis*, a thermophilic Group I.1b archaeon. *ISME J.* 4, 542–552
73 Csuros, M. and Miklos, I. (2009) Streamlining and large ancestral genomes in Archaea inferred with a phylogenetic birth-and-death model. *Mol. Biol. Evol.* 26, 2087–2095
74 Erguder, T.H. *et al.* (2009) Environmental factors shaping the ecological niches of ammonia-oxidizing Archaea. *FEMS Microbiol. Rev.* 33, 855–869
75 Francis, C.A. *et al.* (2007) New processes and players in the nitrogen cycle: the microbial ecology of anaerobic and archaeal ammonia oxidation. *ISME J.* 1, 19–27
76 Nicol, G.W. and Schleper, C. (2006) Ammonia-oxidising Crenarchaeota: important players in the nitrogen cycle? *Trends Microbiol.* 14, 207–212
77 Martens-Habbena, W. *et al.* (2009) Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature* 461, 976–979
78 Ingalls, A.E. *et al.* (2006) Quantifying archaeal community autotrophy in the mesopelagic ocean using natural radiocarbon. *Proc. Natl. Acad. Sci. U. S. A.* 103, 6442–6447
79 Martin-Cuadrado, A.B. *et al.* (2008) Hindsight in the relative abundance, metabolic potential and genome dynamics of uncultivated marine Archaea from comparative metagenomic analyses of bathypelagic plankton of different oceanic regions. *ISME J.* 2, 865–886
80 Huber, H. *et al.* (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* 417, 63–67
81 Guindon, S. and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704