Cell PRESS

The archaeal 'TACK' superphylum and the origin of eukaryotes

Lionel Guy and Thijs J.G. Ettema

Department of Molecular Evolution, Biomedical Center, Uppsala University, Box 596, SE-751 24, Uppsala, Sweden

Although most hypotheses to explain the emergence of the eukaryotic lineage are conflicting, some consensus exists concerning the requirement of a genomic fusion between archaeal and bacterial components. Recent phylogenomic studies have provided support for eocyte-like scenarios in which the alleged 'archaeal parent' of the eukaryotic cell emerged from the Crenarchaeota/Thaumarchaeota. Here, we provide evidence for a scenario in which this archaeal parent emerged from within the 'TACK' superphylum that comprises the Thaumarchaeota, Crenarchaeota and Korarchaeota, as well as the recently proposed phylum 'Aigarchaeota'. In support of this view, functional and comparative genomics studies have unearthed an increasing number of features that are uniquely shared by the TACK superphylum and eukaryotes, including proteins involved in cytokinesis, membrane remodeling, cell shape determination and protein recycling.

The archaeal legacy of the eukaryotic cell

The origin of the eukaryotic cell remains a major enigma in modern biology. Many past and ongoing debates have thus far failed to reach a consensus about the events that have led to its emergence. During the past decades, several competing hypotheses have presented several conceptually different scenarios, many of which are based on circumstantial evidence. However, several key observations are generally accepted [1]. First, all eukaryotes emerged from a common ancestor (the proto-eukaryote, see Glossary), which most probably contained an alphaproteobacterial endosymbiont (the proto-mitochondrion) that evolved into the mitochondria and mitochondria-derived organelles such as hydrogenosomes and mitosomes. Second, notwithstanding eukaryote-specific genes, eukaryotic genomes can roughly be regarded as being chimeric: genes that are mostly involved in information storage and processing (replication, transcription and translation) display an archaeal affinity, whereas genes involved in metabolic processes tend to be bacterial in nature [2].

Anything beyond these points is subject to heated debate, mostly regarding the nature of the archaeal contribution to the proto-eukaryote. Two main schools of thought can be distinguished (Box 1). According to the first school, Archaea and Eukarya are sister clades that shared a common ancestor, following the classical 'Woesean' three-pronged classification of the domains of life [3]. Importantly, scenarios compatible with the three-domain topology require that the genes exclusively shared by Archaea and eukaryotes were vertically inherited from their common ancestor. In addition, such models generally assume that the proto-eukaryote evolved typical eukaryotic features such as a nucleus and an endomembrane system prior to the acquisition of the proto-mitochondrion (Box 1).

In stark contrast to the three-primary domain models stands a second school of thought that represents scenarios in which the eukaryotic lineage emerged from a fusion between an established archaeal lineage (the 'archaeal parent') [4] and one or multiple bacterial components, one of which evolved into the mitochondrion. In such fusion scenarios, the merger between the archaeon and the bacterial partner(s) is believed to have been established on the basis of an endosymbiotic relationship, and the cellular complexity that is characteristic of extant eukaryotes is generally assumed to have somehow been triggered by this endosymbiosis. A considerable variety of fusion models has been proposed over the years [5–9], and, whereas some of these models are supported by phylogenetic evidence, others are solely based on physiological and cytological considerations. Still, consensus has not been reached about the identity of the fusion partners, nor has convincing evidence been provided about the nature of the endosymbiosis. Regarding the identification of the archaeal parent, phylogenomic studies have proven extremely challenging, and conceptually different approaches have reached conflicting conclusions (Box 1).

Glossary

Archaeal parent: archaeal lineage that was involved in the fusion event that gave rise to the origin of the eukaryotic cell.
Archezoa: hypothetical eukaryotic clade that diverged prior to the origin of mitochondria.
Clade: a monophyletic group of organisms.
Eocyte: archaeal lineage from which certain eukaryotic features are thought to have emerged and which is equivalent to Crenarchaeota.
Eukaryogenesis: the process referring to the gradual emergence of the eukaryotic cell in all its complexity.
Ortholog: gene derived from a speciation event.
Paralog: gene derived from a duplication event.
Phylogenetic distribution: pattern of presence and absence of characters, in this case genes, across different taxa.
Proto-eukaryote: ancestral eukaryotic lineage that emerged either via cellular fusion of an archaeal and bacterial components or as a sister clade from the archaeal domain.
Proto-mitochondrion: ancestral alphaproteobacterial lineage that evolved into the mitochondrion.
Signature gene: gene that displays a phylogenetic distribution pattern that unites two clades, suggesting a common ancestry either via vertical or horizontal transmission.

Corresponding author: Ettema, T.J.G. (Thijs.Ettema@ebc.uu.se).

 0966-842X/$ – see front matter doi:10.1016/j.tim.2011.09.002 Trends in Microbiology, December 2011, Vol. 19, No. 12

Box 1. The evolutionary link between Archaea and Eukarya

Apart from identifying the Archaea as the third domain of life, Woese subtly pointed out that the Archaea and Eukarya shared a common branch in the universal tree, implying that they shared a common ancestry [3,46]. Yet, in the aftermath of this notion, phylogenomic and other types of evolutionary reconstructions that aimed at elucidating the nature of this ancestry have yielded conflicting results.

First, several phylogenetic studies are arguably compatible with Woese's three-primary domain scenario. However, this does by no means imply that the presumed common ancestor of Archaea and Eukarya was archaeal in nature because eukaryotic features that could have existed in this ancestor might have been lost in the branch leading to the Archaea [10]. In addition, Archaea-specific features, such as the unique chirality of archaeal lipids, might also have originated along this branch. Importantly, the proto-eukaryotic lineage that spawned from the archaeo-eukaryotic ancestor evolved some of the typical eukaryotic features, such as a nucleus and an endomembrane system, prior to the acquisition of the proto-mitochondrion [47]. The Archezoa hypothesis [48], which probably represents the best defined example of the three-primary domain models, has been in demise since mitochondria-derived organelles were identified in all presumed amitochondriate protists (archezoans) [49]. In addition, the established phylogenetic affiliation of archezoan lineages was later found to be the result of methodological artifacts [49].

Within the realm of fusion-invoking models, some studies point towards a link with the Euryarchaeota as the archaeal parent of the eukaryotes. For example, the hydrogen hypothesis [7] and the syntrophic hypothesis [8,50] are founded on physiological and cytological considerations and infer different forms of metabolic symbiosis between a methanogenic archaeon and one or more bacterial partners. Yet, a recent large-scale supertree analysis detected a phylogenetic signal that suggested a Thermoplasma-like ancestor for the eukaryotic lineage, albeit with limited support [51]. The latter observation is compatible with models that endorse a Thermoplasma-like ancestor of the eukaryotes on the basis of cytological considerations [52,53].

By contrast, several early phylogenetic studies have converged on solutions that involved a fusion between a crenarchaeal parent and a bacterial partner, such as the eocyte hypothesis devised by Lake and coworkers [54]. Evidence for eocyte-like scenarios has been based on phylogenetic analysis of ribosomal proteins [55] and translation factors [56] and on whole-genome-based conditioned reconstruction analysis [57]. More recently, phylogenomic analyses of universally conserved proteins, employing sophisticated evolutionary models, provided support for placing the stem of the eukaryotes as a sister clade to Crenarchaeota [11], or to a clade comprising Crenarchaeota and Thaumarchaeota [12]. Yet another study provided evidence for a eukaryotic origin either within or as a sister to Thaumarchaeota [58], a scenario that would be compatible with a recently proposed fusion hypothesis by Forterre [5], which involved the engulfment of a thaumarchaeon by a bacterium of the Planctomycetes-Verrucomicrobiae-Chlamydia superphylum, followed by a viral invasion.

Finally, a group of models have inferred that the archaeal parent should be found outside of the currently characterized archaeal lineages based on physiological and cytological considerations [59] or on phylogenomic analyses [1].

The lack of congruence between different studies has even prompted some to speak of a 'phylogenomic impasse' [10].

Recently however, employing sophisticated evolutionary models that accommodate across-data compositional heterogeneity, phylogenomic studies have provided robust support for the Eocyte-like scenarios, where eukaryotes form a sister clade of the Crenarchaeota [11] or a clade comprising Crenarchaeota and Thaumarchaeota [12]. Here, we elaborate on these results, and unite phylogenomic evidence with results from functional studies to propose a profound evolutionary link between eukaryotes and the proposed archaeal 'TACK' superphylum comprising Crenarchaeaota, Thaumarchaeota, Korarchaeota and the recently proposed phylum 'Aigarchaeota' (Box 2).

The archaeal 'TACK' superphylum and its phylogenomic affiliation with eukaryotes

In light of the issues outlined above, the phylogenetic position of the eukaryotic lineage relative to the currently available archaeal genome sequences was briefly re-evaluated using several recently developed phylogenetic algorithms and evolutionary models (Figure 1a–d). Essentially, these analyses revealed two important observations. First, they provided strong support for an archaeal clade

Box 2. The TACK superphylum in a nutshell

Thaumarchaeota: recently proposed phylum [60] of abundant chemolithoautotrophic ammonia-oxidizers that play an important role in biogeochemical cycles in both aquatic and terrestrial environments, such as the nitrogen cycle and the carbon cycle. Thaumarchaeota are also referred to as 'mesophilic or low-temperature Crenarchaeota' and were first discovered in marine environments [61,62], but are now also known to reside in terrestrial habitats [17]. Originally, the Thaumarchaeota was envisioned to be the deepest rooting archaeal phylum based on a phylogenetic analysis of concatenated protein datasets [60]. However, the deep-rooting position was found by rooting the tree with the Eukarya, and bacterial sequences were not included in this phylogenetic analysis [60]. In light of the current study, and those of others (for example, [11,12,58]), such rooting should be reconsidered.

'Aigarchaeota': recently proposed candidate phylum that currently comprises species of the Hot Water Crenarchaeotic Group I (HWCGI). A composite genome sequence is available for a single species, '*Ca.* Caldiarchaeum subterraneum', which was identified in a microbial mat at a geothermal stream of a subsurface goldmine [63]. The Caldiarchaeum genome contains some intriguing eukaryotic features, such as genes encoding a ubiquitin protein modifier system [13]. 'Aigarchaeota' might represent a deep-branching sister group to Thaumarchaeota [13] (Figure 1), suggesting a thermophilic origin for this phylum [13].

Crenarchaeota: well-characterized archaeal phylum first described by Woese [3], comprising mostly (acido)thermophilic anaerobes, although some aerobic (Sulfolobales) and micro-aerophilic (e.g. *Pyrobaculum aerophilum*) lineages exist. Crenarchaeal cells come in various shapes, ranging from coccoid to rod-shaped/filamentous, and they mostly rely on respiration for generation energy. Some crenarchaeal members have the ability to grow autotrophically using a recently identified 3-hydroxypropionate/4-hydroxybutyrate carbon dioxide assimilation pathway [64]. Crenarchaea of the orders Sulfolobales and Desulfurococcales utilize a unique Cdv cell division machinery that is related to eukaryotic membrane remodeling systems. Phylogenetically, Crenarchaeota represents a sister group to the clade comprising Thaumarchaeota (once known as 'low-temperature Crenarchaeota') and the candidate phylum 'Aigarchaeota'.

Korarchaeota: candidate phylum comprising a group of deep-branching Archaea with ultra-thin, needle-shaped cells measuring up to 100 μm in length, which are geographically restricted to terrestrial and marine thermal environments. Based on the only genome sequence that is currently available for this candidate phylum [65], that of '*Ca.* Korarchaeum cryptofilum', it relies on a simple mode of peptide fermentation for generation of energy and biomass.

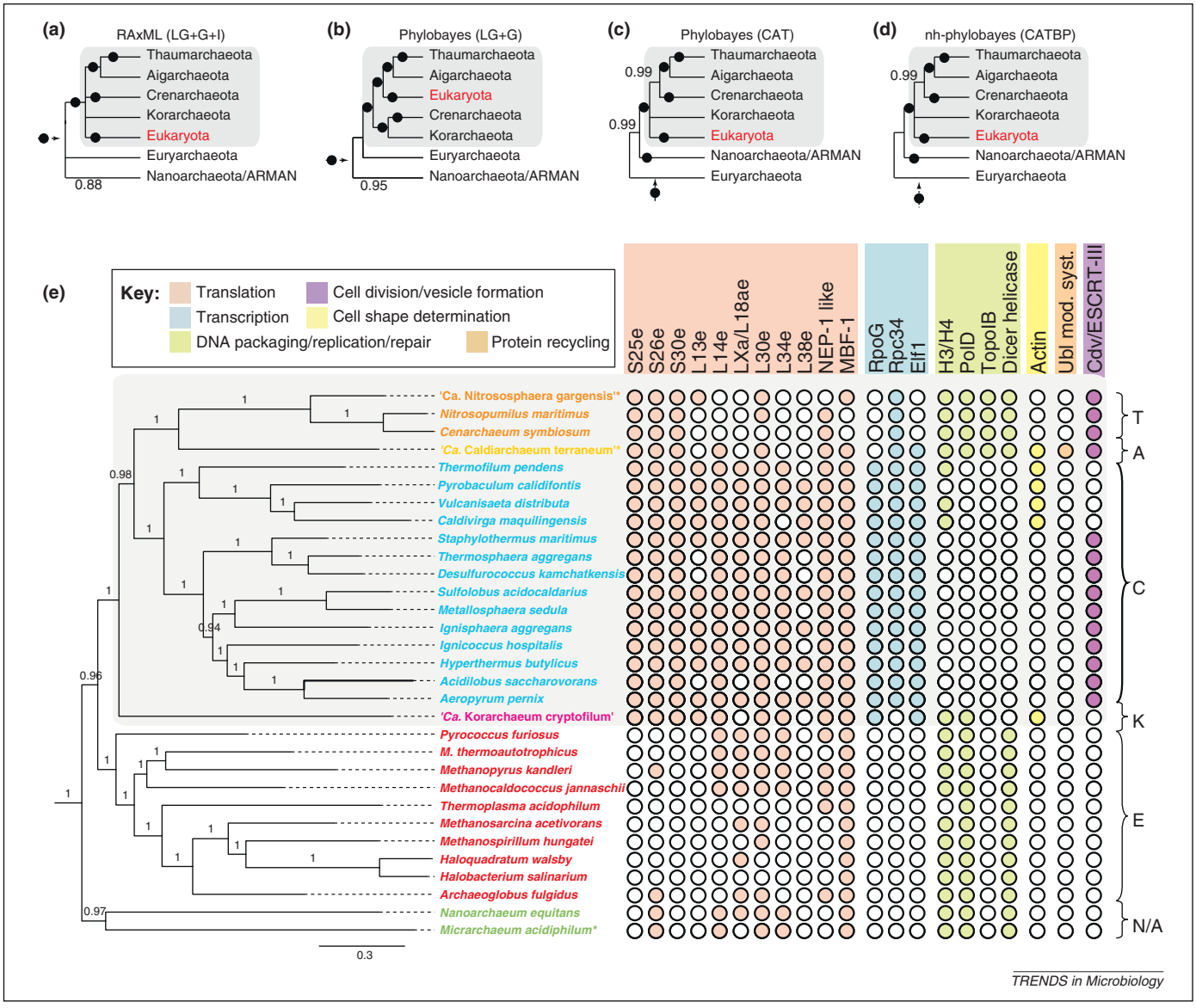

Figure 1. Evolutionary links between the TACK superphylum and eukaryotes. **(a–d)** Show a schematic overview of phylogenetic analyses that were performed with different methods to investigate the placement of Eukarya relative to the TACK superphylum. The phylogenetic trees are based on an aligned [39] dataset of 26 universally conserved proteins defined as the intersection between datasets used in the phylogenomic studies performed by Cox *et al.* [11] and Ciccarelli *et al.* [40] (COG0016, PheS; COG0024, Map; COG0048, RPS12; COG0049, RPS7; COG0052, RPS2; COG0080, RPL11; COG0081, RPL1; COG0085, RpoB; COG0086, RpoC; COG0087, RPL3; COG0091, RPL22; COG0092, RPS3; COG0093, RPL14; COG0094, RPL5; COG0096, RPS8; COG0097, RPL6; COG0098, RPS5; COG0099, RPS13; COG0100, RPS11; COG0102, RPL13; COG0103, RPS9; COG0186, RPS17; COG0197, RPL16; COG0201, SecY; COG0256, RPL18; COG0533, Gcp). Archaeal species included are those listed in panel **(e)**, and eukaryotic sequences were obtained from *Homo sapiens*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Dictyostelium discoideum*, *Leishmania infantum*, *Plasmodium falciparum* and *Giardia lamblia*. The following bacterial species were selected as outgroup: *Borrelia burgdorferi*, *Escherichia coli*, *Rickettsia prowazekii*, *Bacillus subtilis*, *Synechocystis* sp., *Chlamydia trachomatis*, *Rhodopirellula baltica*, *Bacteroidetes thetaiotaomicron*, *Campylobacter jejuni* and *Thermotoga maritima*. Unambiguously aligned regions with less than 50% gaps were kept for phylogenetic analyses either with RAxML 7.0.4 [41] (a, using LG+G+I model), Phylobayes 3.2f [42] (b and c, using LG + Gamma and CAT, respectively) and nh-phylobayes 0.2.3 [43] (d, using the CATBP model). Branches that are supported with a posterior probability (PP) of 1 (or 100% bootstrap support (BS) in the case of RAxML) are depicted with a filled circle (●) and branches with PP and BS support values lower than 0.95 and 70%, respectively, have been collapsed. The eukaryotic domain is highlighted in red, and the bacterial root is indicated with an arrow. Note that for each analysis, the eukaryotes are placed within the proposed TACK superphylum (shaded in grey) with highest support. Except for the Phylobayes analysis with the LG model, which placed eukaryotes as a sister to the clade comprising Thaumarchaeaota and 'Aigarchaeota' **(b)**, the position of the eukaryotes in the TACK superphylum was not fully resolved. For the analyses depicted in **(c)** and **(d)**, the bacterial root was placed within the euryarchaeal phylum, as was previously observed in studies by Cox *et al.* [11] and Foster *et al.* [12]. **(e)** Phylogenetic distribution of archaeal orthologs of eukaryotic core genes involved in translation (pink shading), transcription (blue shading), DNA packaging, replication and repair (green shading), cell shape determination (yellow shading), protein recycling (orange shading) and cell division and vesicle formation (purple shading). The phylogenetic tree depicted on the left is based on the dataset described above, except that eukaryotic data was omitted. Phylogenetic analysis was carried out using Phylobayes 3.2f [42] under the LG model with a continuous Gamma distribution. The tree was rooted with Bacteria (not shown for viewing clarity). For each branch, PP support values are indicated. Note that the proposed TACK superphylum (shaded in grey) is supported with a PP of 1.00. Abbreviations: T, Thaumarchaeota (orange font); A, 'Aigarchaeota' (yellow font); C, Crenarchaeota (blue font); K, Korarchaeota (magenta font); E, Euryarchaeota (red font); N/A, Nanoarchaeota and ARMAN (green font). An asterisk (*) denotes those species for which the genome sequence is incomplete: some proteins might thus be incorrectly scored as being absent. For example, the CdvA protein was not detected in the '*Ca.* Caldiarchaeum subterraneum' composite genome [13].

comprising Crenarchaeaota, Thaumarchaeota, Korarchaeota and the recently proposed phylum 'Aigarchaeota', which currently solely comprises the Hot Water Crenarchaeal Group I (HWCGI) archaeon '*Candidatus* Caldiarchaeum subterraneum' [13] (Figure 1a–d). Here, we propose to refer to this archaeal clade as the 'TACK' superphylum (Box 2), after the first letter of the included phyla, and in reference to the nautical term 'tacking', which means 'bringing (a vessel) into the wind in order to change course or direction'. To emphasize the anticipated evolutionary link between that TACK superphylum and eukaryotes (see below), the choice of the nautical lingo attempts to reflect the dramatic change in evolutionary fate that this archaeal lineage (the 'vessel') experienced in founding the eukaryotic lineage. Importantly, whereas the positions of the individual phyla were not always fully resolved, the proposed TACK superphylum was always separated from the branches leading to Euryarchaeota and/or Nanoarchaeota with robust support. In addition, Caldiarchaeum was unequivocally placed as a deep-branching sister lineage to the Thaumarchaeota in each analysis performed, as was previously suggested [13]. Hence, whether or not the 'Aigarchaeota' represent a separate phylum or a deeply branching sister lineage of the Thaumarchaota remains to be determined. The fact that Caldiarchaeum shares more genes with Korarchaeum than it does with Thaumarchaeota could be used to argue for the former scenario [13].

Second, the Eukaryota were invariably affiliated with the proposed TACK superphylum with robust support. Except for the Phylobayes analysis with the LG model, which placed eukaryotes as a sister to the clade comprising Thaumarchaeaota and 'Aigarchaeota' (Figure 1b), the position of eukaryotes within the TACK superphylum remains unresolved. Hence, at face value, the phylogenetic analysis presented in Figure 1 is compatible with recent phylogenomic analyses by Cox *et al*. [11] and Foster *et al*. [12], who also placed eukaryotes as a sister clade to Crenarchaeota, or to Crenarchaeota and Thaumarchaeota, respectively. The phylogenetic placement of the domain Eukarya inside the (primary) domain Archaea underlines the chimeric nature of the eukaryotic cell, and supports a scenario in which the archaeal gene set present in eukaryotes was vertically inherited from the archaeal parent that is affiliated with the TACK superphylum.

Genomic signatures that unite the TACK superphylum with eukaryotes

In the advent of the genomic era, comparative genomic studies have provided convincing evidence for a common evolutionary history between Archaea and Eukarya, identifying a common set of genes absent in Bacteria. Analyses of these archaeo-eukaryotic signature genes have revealed that several key eukaryotic molecular machineries clearly have an archaeal provenance, and that most of them are involved in information storage and processing [14–16]. For example, the basal components comprising the eukaryotic translational, transcriptional and replication systems, as well as the systems involved in RNA degradation and proteolysis are primarily archaeal in nature. Apart from these core machineries, the phylogenetic distribution of archaeo-eukaryotic signature genes displays a patchy distribution across archaeal genomes, with some genes seemingly being specific to the Euryarchaeota, whereas others appeared to be present only in members of the Crenarchaeota, Thaumarchaeota and/or Korarchaeota [4,17]. Here, by projecting these archaeo-eukaryotic signature genes against an archaeal species tree that contains an expanded set of sequenced genomes, a picture emerges that is compatible with a scenario where the archaeal parent of the eukaryotic cell emerges from within the TACK superphylum (Figure 1e).

On the one hand, for archaeo-eukaryotic signature genes that previously displayed a pattern compatible with a euryarchaeal origin of eukaryotes, orthologs could now also be identified in lineages that are part of the TACK superphylum. For example, genes encoding H3/H4-type histones, the small subunit of the PolD DNA polymerase and ERCC4/Dicer-type helicases are present in all available thaumarchaeal genomes and in Caldiarchaeum (Figure 1e). Moreover, H3/H4-type histones have also been identified in several crenarchaeal genomes and in the Korarchaeum genome (Figure 1e).

On the other hand, archaeo-eukaryotic signatures in support of an evolutionary affiliation between the TACK superphylum and eukaryotes remain intact, and new eukaryotic features are continually discovered following genomic exploration of new lineages of this clade. Examples of eukaryotic genes for which orthologs are exclusively found in members of the TACK phylum include several ribosomal proteins (S25e, S30e, L13e and L38e), the eukaryotic RNA polymerase III subunit Rpc34 [18], the eukaryotic transcription elongation factor Elf1 [19] and the archaeal counterpart of the small RPB8 subunit of the eukaryotic RNA polymerase (RpoG), although RpoG is absent in Thaumarchaeota and Caldiarchaeum [20] (Figure 1e).

Another interesting link between the TACK superphylum and eukaryotes is the Cdv cell division machinery recently discovered in hyperthermophilic Crenarchaeota [21,22] (Figure 1e). Cdv components CdvB and CdvC are homologous to Snf7 and Vps4, respectively, which are part of the eukaryotic ESCRT-III membrane remodeling machinery, mediating final stages of membrane scission in processes such as multivesicular body biogenesis, viral budding and, as was shown recently, cytokinesis [23]. Interestingly, crenarchaeal and thaumarchaeal genomes encode multiple CdvB paralogs, as does the genome of Caldiarchaeum (which seemingly lacks CdvA [13]). This opens up the possibility that these proteins have subfunctionalized in a similar fashion as in eukaryotes. This is further substantiated by the observation that CdvB homologs have been implicated in formation of membrane vesicles [24], suggesting a role in excretion, and by the detection of an archaeal CdvB paralog in virion particles [25], suggesting a role in viral budding [26]. As such, the eukaryotic Vps4- and Snf7-like proteins described here are thought to have originated from within the TACK superphylum.

A further link between the TACK superphylum and eukaryotes is represented by the recent identification of a gene cluster that encodes the components of a eukaryotic-type ubiquitin (Ubl) modifier system in the composite

Caldiarchaeum genome [13] (Figure 1e). The Ubl gene cluster comprises five genes, encoding eukaryotic-type Ubl, Ubl-activating enzyme E1l, Ubl-conjugating enzyme E2l, 26S proteasome regulatory subunit RPN11l and an adjacent gene encoding a small RING finger protein that might be the progenitor of the RING-type ubiquitin ligase E3 protein [13]. This finding strongly suggests that the eukaryotic-type Ubl modifier system was present in the archaeal parent of the eukaryotic cell, and significantly extends the set of archaeo-eukaryotic signature genes involved in proteolysis.

Finally, the recent discovery of an archaeal actin ortholog [27,28] in the crenarchaeal order Thermoproteales and in '*Ca.* Korarchaeum cryptofilum' represents yet another example of how the proposed TACK superphylum is linked to eukaryotes (Figure 1e). The archaeal actin ortholog, which was named Crenactin, was found to be involved in cell shape formation, forming helical bundles that traverse the length of rod-shaped *Pyrobaculum calidifontis* cells [29]. The proposed role in cell shape determination for Crenactin is strongly supported by the phylogenetic distribution of Crenactin: all rod-shaped (Thermoproteales) and filamentous (Korarchaeum) members of the TACK superphylum encode Crenactin in their genomes, and Crenactin is absent from genomes of coccoid members [29]. By inference, the presence of Crenactin in the Caldiarchaeum genome suggests that cells of this organism, for which no cytological information is available, are also rod-shaped or filamentous in nature. Based on these findings, it seems that the ancestral role of actin has primarily been directed towards cell shape determination, and that in the process of eukaryogenesis, actin adopted auxiliary cellular roles in vesicular transport, mechanical force generation and cytokinesis. Critically, the presence of an actin-based cytoskeleton in Archaea bears importance for the origin of the eukaryotic cell itself.

In summary, even though the distribution of the genes mentioned here is patchy, there is so far no gene that is exclusively represented in Euryarchaeota and the eukaryotes, whereas several genes are uniquely shared between TACK lineage and eukaryotes, thereby supporting the hypothesis that eukaryotes emerged from within the TACK superphylum.

Did eukaryotes emerge from a 'complex' archaeal lineage?

The observations described above are best compatible with a scenario in which the archaeal parent of eukaryotes, represented by a deeply-rooting TACK lineage, harbored a full complement of currently recognized archaeo-eukaryotic signature genes, including those universally present in Archaea and eukaryotes, and those present in only a subset of archaeal lineages (Figure 1e). Consequently, the archaeal parent was a relatively complex cellular entity that contained an actin cytoskeleton, ESCRT-like membrane remodeling systems and a ubiquitin protein modifier system. The anticipated complexity is perhaps best reflected by the multitude of potential cell division machineries that are inferred to be present in the ancestor of the eukaryotes [26,27]: did it utilize an ESCRT or FtsZ-mediated mechanism for cell scission, or did actin already play a role in this process, as it does in present-day eukaryotes? Or did these proteins perhaps coordinate cytokinesis in a concerted way? In any case, even if not in cytokinesis, we envision that actin has played a crucial role in the initial stages of eukaryogenesis. The actin-based cytoskeleton could have provided the proto-eukaryotic cell with primitive phagocytotic capabilities, sustaining the ability to form membrane protrusions. At some stage, this primitive phagocytotic machinery might have facilitated the fusion with the proto-mitochondrion, a key event for the origin of the eukaryotic cell [29].

In contrast to the assumed 'complex' archaeal parent of the eukaryotic cell stand the extant archaeal TACK lineages, among which eukaryotic features are patchily distributed. How can this be explained? It is of course possible that the complex ancestral TACK lineage from which eukaryotes emerged existed only transiently in time, although it cannot be excluded that 'complex' archaeal lineages remain to be discovered. In fact, '*Ca.* Caldiarchaeum subterraneum' might represent such relatively complex lineage, as it uniquely unites several eukaryotic features in one archaeon (actin, ESCRT and the Ubl-mediated protein modification system). Yet, Caldiarchaeum lacks some of the eukaryotic features conserved in other archaeal lineages (several ribosomal proteins and RpoG), and such patterns are even more apparent in other archaeal lineages in which eukaryotic features display an even patchier distribution, such as in the Crenarchaeota (Figure 1e). Altogether, the patchy distribution of eukaryotic features across extant archaeal lineages suggests that extant lineages have undergone differential reductive evolution [4,30].

Exploring uncharacterized TACK lineages will further unravel the origin of the eukaryotic cell

As outlined above, phylogenomic and functional evidence is accumulating in support of a scenario in which the archaeal parent of the eukaryotic cell might have emerged from within the proposed TACK superphylum. Given that genomic exploration of recently discovered archaeal (candidate) phyla, such as the Thaumarchaeota and 'Aigarchaeota', has added significant flavor to hypotheses entailing the evolutionary origins of the eukaryotic cell, the identification and characterization of other deeply-branching archaeal lineages of the TACK superphylum should be prioritized. Indeed, PCR-based surveys of a wide variety of environments have revealed numerous uncharacterized archaeal lineages that are anticipated to play pivotal roles in global energy cycles (reviewed in [31,32]). Interestingly, many of these lineages represent deeply rooting branches in the newly proposed TACK superphylum (Figure 2). For example, Thaumarchaeota seemingly represents a taxonomically broad phylum that includes several uncharacterized clades that were detected in a wide variety of habitats [33] (Figure 2, blue shading). Moreover, completely new, uncharacterized phyla might be part of the TACK superphylum, such as the putative candidate phylum comprising the deep archaeal lineages Ancient Archaeal Group (AAG) [34], Deep-Sea Archaeal Group (DSAG) [35] and the Marine Hydrothermal Vent Group (MHVG) [34] (Figure 2, pink shading).

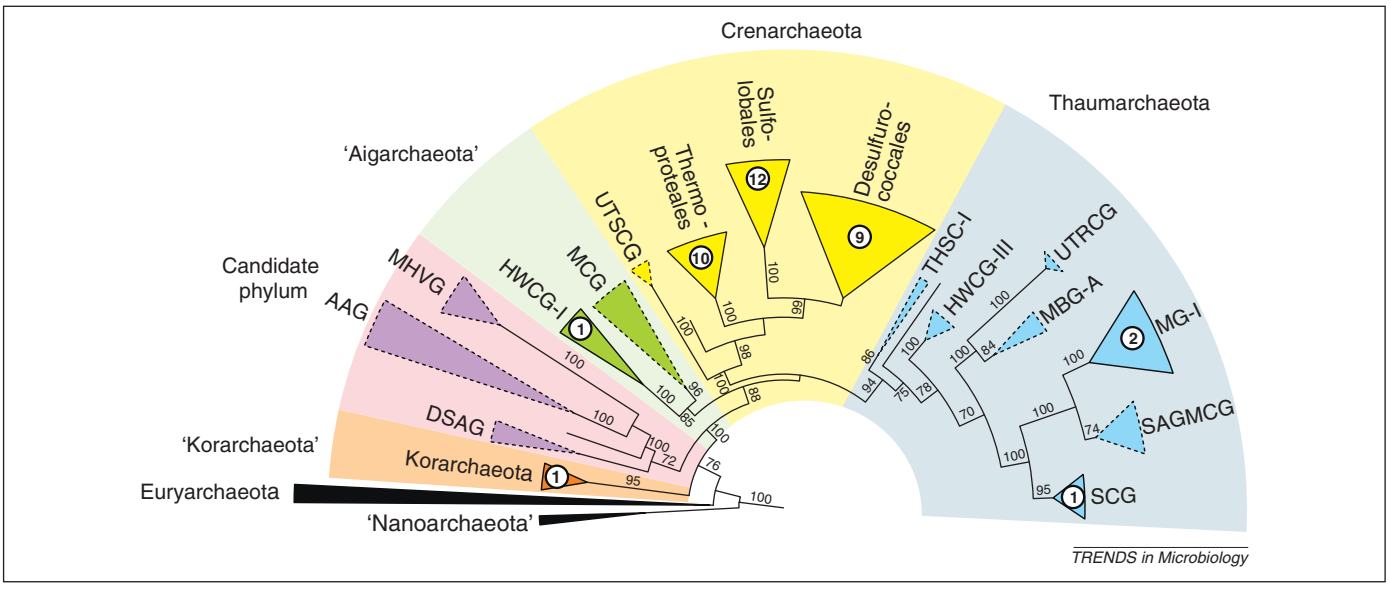

Figure 2. Phylogenetic diversity of cultivated and uncultivated members of the proposed TACK superphylum. A schematic phylogenetic tree is shown of 16S sequences of cultivated and uncultivated (environmental) members of the proposed TACK superphylum. Environmental sequences were retrieved from SILVA and an alignment was created using the SINA webaligner [44], which subsequently was used to generate a tree using RAxML (version 7.0.4) under the GTRGAMMA model and performing 100 non-parametric bootstraps replicates (BS) [41]. The tree was rooted with bacteria (not shown for viewing clarity), and sequences belonging to the phyla Euryarchaeota and Nanoarchaeota have been collapsed. For clarity, major clades are displayed as triangles, and (candidate) phyla are shaded in different colors as follows: Thaumarchaeaota, blue; 'Aigarchaeota', purple; Crenarchaeota, yellow; Korarchaeota, red. A putative new archaeal phylum is shaded in pink. Boundaries between (candidate) phyla were chosen so that the sequence identity between each sequence in the respective (candidate) phylum was less than 85% [45], and that each phylum was supported by a BS support value of at least 70%. Clades that contain species for which a genome sequence is available are drawn with a solid line, and the number of sequenced genome sequences is indicated in a white circle. Genomically uncharacterized clades are drawn with dashed lines. Abbreviations: AAG, Ancient Archaeal Group; DSAG, Deep-Sea Archaeal Group; HWCG-I, Hot Water Crenarchaeotic Group I; HWCG-III, Hot Water Crenarchaeotic Group III; MCG, Miscellaneous Crenarchaeotic Group; MG-I, Marine Group I; MBG-A, Marine Benthic Group A; MHVG, Marine Hydrothermal Vent Group; SCG, Soil Crenarchaeotic Group; SAGMCG, South Africa Gold Mine Crenarchaeotic Group; THSC-I, Terrestrial Hot Spring Crenarchaeota Group I; UTRCG, Uncultured Thaumarchaeota related clone group; UTSCG, Uncultured thermoacidic spring clone group.

Recently developed single cell-based methods [36] are now able to provide unique genomic insight in uncultured prokaryotes [37]. In the near future, such methods may represent viable alternatives to both traditional culture-based approaches, as well as to metagenomic and PCR-based surveys (which are known to have their limitations), to explore the uncharacterized diversity of the TACK superphylum at the genomic level. To this end, archaeal communities of interest residing in complex environmental samples can be specifically targeted or enriched using available cell sorting technologies, and, subsequently, the selected cells can be subjected to single cell genomics analyses.

Concluding remarks

The emergence of the eukaryotic cell remains a major enigma in modern biology, and because solid empirical data that might reveal profound insights in this process has been lacking thus far, hypotheses that have addressed this conundrum have predominantly been speculative in nature. Now that the amount of genomic data for a taxonomically diverse set of organisms is starting to grow, and that appropriate methods are becoming available to analyze this data in an evolutionary context, it is time to put past theories to the test and to establish better ones.

We propose that the archaeal parent of the proto-eukaryotic cell emerged from within a clade comprising Thaumarchaeota, 'Aigarchaeota', Crenarchaeota and Korarchaeota, and envision that the genomic exploration of these clades might teach us a great deal about the events that have led to the origin of the eukaryotic cell. The existing gap between extent archaeal and eukaryotic lineages in terms of cellular and organizational complexity is enormous and hard to reconcile from an evolutionary point of view. Yet, the recent discovery of proteins previously presumed to be typical eukaryotic signatures in archaeal lineages shows that this gap has certainly started to shrink. Furthermore, the existence of archaeal lineages that display certain levels of cellular complexity cannot be ruled out *a priori*. The recent discovery of spatial separation of energy generating and information processing processes in the crenarchaeote *Ignicoccus hospitalis* [38] has clearly demonstrated that cellular compartmentation is not restricted to eukaryotes (or bacteria belonging to the Planctomycetes). Certainly, the future exploration of the archaeal TACK superphylum (Box 3) will blow fresh wind

Box 3. Outstanding questions

- Can the phylogenetic position of eukaryotes be (further) resolved, e.g. by increased taxon sampling, or by better phylogenomic methods?
- Given that many uncharacterized archaeal lineages exist, such as those in the TACK superphylum, do extant lineages exist that bear a closer relation to eukaryotes than those studied thus far?
- Given the existence of Archaea such as *Ignicoccus hospitalis*, which contain intracellular membrane structures that govern spatial separation between energy generating and information processing processes, do archaeal lineages exist that show additional degrees of cellular compartmentation?
- What was the nature of the relation between the archaeal parent and the bacterial partner(s) from which the eukaryotic lineage supposedly emerged?

into our sails, allowing us to set direction into the unexplored territories of our murky past.

Acknowledgments

T.J.G.E. is supported by grants from the Swedish Research Council (621-2009-4813) and by a Marie Curie ERG grant (European Union). L.G. is supported by a fellowship from the Swiss National Science Foundation (PA00P3_131491). We are grateful to Eugene Koonin (NCBI, Bethesda, MD, USA) for stimulating discussions and valuable comments on the manuscript.

References

1 Yutin, N. *et al.* (2008) The deep archaeal roots of eukaryotes. *Mol. Biol. Evol.* 25, 1619–1630
2 Rivera, M.C. *et al.* (1998) Genomic evidence for two functionally distinct gene classes. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6239–6244
3 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
4 Koonin, E.V. (2010) The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biol.* 11, 209
5 Forterre, P. (2011) A new fusion hypothesis for the origin of Eukarya: better than previous ones, but probably also wrong. *Res. Microbiol.* 162, 77–91
6 Martin, W. *et al.* (2001) An overview of endosymbiotic models for the origins of eukaryotes, their ATP-producing organelles (mitochondria and hydrogenosomes), and their heterotrophic lifestyle. *Biol. Chem.* 382, 1521–1539
7 Martin, W. and Muller, M. (1998) The hydrogen hypothesis for the first eukaryote. *Nature* 392, 37–41
8 Moreira, D. and Lopez-Garcia, P. (1998) Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J. Mol. Evol.* 47, 517–530
9 Poole, A.M. and Penny, D. (2007) Evaluating hypotheses for the origin of eukaryotes. *Bioessays* 29, 74–84
10 Gribaldo, S. *et al.* (2010) The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? *Nat. Rev. Microbiol.* 8, 743–752
11 Cox, C.J. *et al.* (2008) The archaebacterial origin of eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20356–20361
12 Foster, P.G. *et al.* (2009) The primary divisions of life: a phylogenomic approach employing composition-heterogeneous methods. *Philos. Trans. R Soc. Lond. B: Biol. Sci.* 364, 2197–2207
13 Nunoura, T. *et al.* (2011) Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res.* 39, 3204–3223
14 Ettema, T.J. *et al.* (2005) Discovering novel biology by in silico archaeology. *Nat. Rev. Microbiol.* 3, 859–869
15 Makarova, K.S. *et al.* (1999) Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell. *Genome Res.* 9, 608–628
16 Makarova, K.S. and Koonin, E.V. (2003) Comparative genomics of Archaea: how much have we learned in six years, and what's next? *Genome Biol.* 4, 115
17 Spang, A. *et al.* (2010) Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. *Trends Microbiol.* 18, 331–340
18 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
19 Daniels, J.P. *et al.* (2009) Identification of a crenarchaeal orthologue of Elf1: implications for chromatin and transcription in Archaea. *Biol. Direct* 4, 24
20 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
21 Lindås, A.C. *et al.* (2008) A unique cell division machinery in the Archaea. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18942–18946
22 Samson, R.Y. *et al.* (2008) A role for the ESCRT system in cell division in archaea. *Science* 322, 1710–1713
23 Hurley, J.H. and Hanson, P.I. (2010) Membrane budding and scission by the ESCRT machinery: it's all in the neck. *Nat. Rev. Mol. Cell Biol.* 11, 556–566
24 Ellen, A.F. *et al.* (2009) Proteomic analysis of secreted membrane vesicles of archaeal *Sulfolobus* species reveals the presence of endosome sorting complex components. *Extremophiles* 13, 67–79
25 Maaty, W.S. *et al.* (2006) Characterization of the archaeal thermophile *Sulfolobus* turreted icosahedral virus validates an evolutionary link among double-stranded DNA viruses from all domains of life. *J. Virol.* 80, 7625–7635
26 Ettema, T.J. and Bernander, R. (2009) Cell division and the ESCRT complex: a surprise from the archaea. *Commun. Integr. Biol.* 2, 86–88
27 Makarova, K.S. *et al.* (2010) Evolution of diverse cell division and vesicle formation systems in Archaea. *Nat. Rev. Microbiol.* 8, 731–741
28 Yutin, N. *et al.* (2009) The origins of phagocytosis and eukaryogenesis. *Biol. Direct* 4, 9
29 Ettema, T.J. *et al.* (2011) An actin-based cytoskeleton in Archaea. *Mol. Microbiol.* 80, 1052–1061
30 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
31 Schleper, C. *et al.* (2005) Genomic studies of uncultivated archaea. *Nat. Rev. Microbiol.* 3, 479–488
32 Teske, A. and Sorensen, K.B. (2008) Uncultured archaea in deep marine subsurface sediments: have we caught them all? *ISME J.* 2, 3–18
33 Auguet, J.C. *et al.* (2010) Global ecological patterns in uncultured Archaea. *ISME J.* 4, 182–190
34 Takai, K. and Horikoshi, K. (1999) Genetic diversity of archaea in deep-sea hydrothermal vent environments. *Genetics* 152, 1285–1297
35 Vetriani, C. *et al.* (1999) Population structure and phylogenetic characterization of marine benthic Archaea in deep-sea sediments. *Appl. Environ. Microbiol.* 65, 4375–4384
36 Kalisky, T. and Quake, S.R. (2011) Single-cell genomics. *Nat. Methods* 8, 311–314
37 Woyke, T. *et al.* (2009) Assembling the marine metagenome, one cell at a time. *PLoS ONE* 4, e5299
38 Kuper, U. *et al.* (2010) Energized outer membrane and spatial separation of metabolic processes in the hyperthermophilic Archaeon *Ignicoccus hospitalis*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 3152–3156
39 Katoh, K. *et al.* (2009) Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* 537, 39–64
40 Ciccarelli, F.D. *et al.* (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science* 311, 1283–1287
41 Stamatakis, A. and Ott, M. (2008) Efficient computation of the phylogenetic likelihood function on multi-gene alignments and multi-core architectures. *Philos. Trans. R Soc. Lond. B: Biol. Sci.* 363, 3977–3984
42 Lartillot, N. *et al.* (2009) PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25, 2286–2288
43 Blanquart, S. and Lartillot, N. (2008) A site- and time-heterogeneous model of amino acid replacement. *Mol. Biol. Evol.* 25, 842–858
44 Pruesse, E. *et al.* (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196
45 Hugenholtz, P. *et al.* (1998) Novel division level bacterial diversity in a Yellowstone hot spring. *J. Bacteriol.* 180, 366–376
46 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
47 Embley, T.M. and Martin, W. (2006) Eukaryotic evolution, changes and challenges. *Nature* 440, 623–630
48 Cavalier-Smith, T. (1989) Molecular phylogeny. Archaebacteria and Archezoa. *Nature* 339, l00–l01
49 Embley, T.M. (2006) Multiple secondary origins of the anaerobic lifestyle in eukaryotes. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 361, 1055–1067
50 Lopez-Garcia, P. and Moreira, D. (1999) Metabolic symbiosis at the origin of eukaryotes. *Trends Biochem. Sci.* 24, 88–93
51 Pisani, D. *et al.* (2007) Supertrees disentangle the chimerical origin of eukaryotic genomes. *Mol. Biol. Evol.* 24, 1752–1760
52 Hixon, W.G. and Searcy, D.G. (1993) Cytoskeleton in the archaebacterium *Thermoplasma acidophilum*? Viscosity increase in soluble extracts. *Biosystems* 29, 151–160

53 Margulis, L. *et al.* (2000) The chimeric eukaryote: origin of the nucleus from the karyomastigont in amitochondriate protists. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6954–6959
54 Lake, J.A. *et al.* (1984) Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 81, 3786–3790
55 Vishwanath, P. *et al.* (2004) Ribosomal protein-sequence block structure suggests complex prokaryotic evolution with implications for the origin of eukaryotes. *Mol. Phylogenet. Evol.* 33, 615–625
56 Baldauf, S.L. *et al.* (1996) The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proc. Natl. Acad. Sci. U.S.A.* 93, 7749–7754
57 Rivera, M.C. and Lake, J.A. (2004) The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* 431, 152–155
58 Kelly, S. *et al.* (2011) Archaeal phylogenomics provides evidence in support of a methanogenic origin of the Archaea and a thaumarchaeal origin for the eukaryotes. *Proc. Biol. Sci.* 278, 1009–1018
59 Vellai, T. *et al.* (1998) A new aspect to the origin and evolution of eukaryotes. *J. Mol. Evol.* 46, 499–507
60 Brochier-Armanet, C. *et al.* (2008) Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat. Rev. Microbiol.* 6, 245–252
61 DeLong, E.F. (1992) Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. U.S.A.* 89, 5685–5689
62 Fuhrman, J.A. *et al.* (1992) Novel major archaebacterial group from marine plankton. *Nature* 356, 148–149
63 Nunoura, T. *et al.* (2005) 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
64 Berg, I.A. *et al.* (2007) A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science* 318, 1782–1786
65 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