The Dispersed Archaeal Eukaryome and the Complex Archaeal Ancestor of Eukaryotes

Eugene V. Koonin and Natalya Yutin

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894

Correspondence: koonin@ncbi.nlm.nih.gov

The ancestral set of eukaryotic genes is a chimera composed of genes of archaeal and bacterial origins thanks to the endosymbiosis event that gave rise to the mitochondria and apparently antedated the last common ancestor of the extant eukaryotes. The proto-mitochondrial endosymbiont is confidently identified as an α-proteobacterium. In contrast, the archaeal ancestor of eukaryotes remains elusive, although evidence is accumulating that it could have belonged to a deep lineage within the TACK (Thaumarchaeota, Aigarchaeota, Crenarchaeota, Korarchaeota) superphylum of the Archaea. Recent surveys of archaeal genomes show that the apparent ancestors of several key functional systems of eukaryotes, the components of the archaeal “eukaryome,” such as ubiquitin signaling, RNA interference, and actin-based and tubulin-based cytoskeleton structures, are identifiable in different archaeal groups. We suggest that the archaeal ancestor of eukaryotes was a complex form, rooted deeply within the TACK superphylum, that already possessed some quintessential eukaryotic features, in particular, a cytoskeleton, and perhaps was capable of a primitive form of phagocytosis that would facilitate the engulfment of potential symbionts. This putative group of Archaea could have existed for a relatively short time before going extinct or undergoing genome streamlining, resulting in the dispersion of the eukaryome. This scenario might explain the difficulty with the identification of the archaeal ancestor of eukaryotes despite the straightforward detection of apparent ancestors to many signature eukaryotic functional systems.

The origin of the eukaryotes obviously is one of the key problems in the entire study of the history of life on Earth. The extreme complexity of the eukaryotic cellular organization compared with that of archaeal and bacterial (collectively, prokaryotic) cells cries for an evolutionary explanation and even has brought the specter of “irreducible complexity” into the scientific debate (Kurland et al. 2006; Martin et al. 2007). The “standard model” derived primarily from the classic phylogenetic analysis of 16S RNA by Woese and coworkers has eukaryotes as the sister group of Archaea, to the exclusion of Bacteria (Woese et al. 1990; Pace 1997, 2006, 2009). However, advances of comparative genomics along with discoveries of cell biology precipitated a major shift in our understanding of the origin of eukaryotes by showing that eukary-
otes are Archaeo–Bacterial chimeras. First, because it has been shown that all extant eukaryotes possess mitochondria or their degraded relatives, hydrogenosomes and mitosomes, by inference the Last Common Eukaryotic Ancestor (LECA), already harbored the α-proteobacterial endosymbiont that gave rise to the mitochondria (Émbley and Martin 2006; van der Giezen 2009; Shiflett and Johnson 2010). Second, comparative genomic analysis indicates that eukaryotes possess two distinct sets of genes, one of which shows apparent phylogenetic affinity with homologs from Archaea, whereas the other one is more closely related to bacterial homologs (obviously, not all eukaryotic genes belong to these two classes because many are of uncertain origin, and many more appear to be unique to eukaryotes). The eukaryotic genes of apparent archaeal descent encode, primarily, proteins involved in information processing (translation, transcription, replication, repair), and the genes of inferred bacterial origin encode mostly proteins with “operational” functions (metabolic enzymes, components of membranes and other cellular structures, etc.) (Esser et al. 2004; Rivera and Lake 2004; Pisani et al. 2007; Thiergart et al. 2012). Some of the informational and operational systems of eukaryotes show qualitative separation between “archaeal” and “bacterial” genes: thus, the key proteins involved in DNA replication in Archaea and eukaryotes are not homologous to the functionally analogous proteins of Bacteria (Leipe et al. 1999), and conversely, some of the principal enzymes of membrane biogenesis are homologous in eukaryotes and Bacteria but not in Archaea (Pereto et al. 2004).

The parsimonious scenario of eukaryogenesis that takes into account the ancestral presence of mitochondria in eukaryotes and the hybrid composition of the eukaryotic gene complement seems to be that the first eukaryote emerged through the invasion of an archaeon by an α-proteobacterium (Martin and Muller 1998; Rivera and Lake 2004; Martin and Koonin 2006; Martin et al. 2007). Under this scenario, a plausible chain of events has been proposed leading from the endosymbiosis to the evolution of eukaryotic innovations such as the endomembrane system including the nucleus and the cytoskeleton (Koonin 2006; Martin and Koonin 2006). Furthermore, argument has been presented that the energy requirements for the functioning of a eukaryotic cell that is orders of magnitude bigger than typical prokaryotic cells cannot be met by means other than utilization of multiple “power stations” such as the mitochondria (Lane and Martin 2010; Lane 2011).

Despite the plausibility of the symbiogenetic scenario, alternatives are still seriously considered (Embley and Martin 2006; Poole and Penny 2007a,b). Conceptually, these alternatives converge to the archezoon hypothesis under which the host of the α-proteobacterial endosymbiont was not an archaeon but a primitive, amitochondrial proto-eukaryote that already possessed the hallmarks of the eukaryotic cellular architecture (including the nucleus) and was capable of phagocytosis, a function that is sometimes considered essential for the acquisition of endosymbionts (Cavalier-Smith 1991, 2009; Kurland et al. 2006; Poole and Penny 2007a,b). The two scenarios of eukaryogenesis substantially differ with respect to the level of complexity that is attributed to the host of the mitochondrial endosymbiont. Under the symbiotic hypothesis, the host was a “garden-variety” archaeon, and the dramatic complexification of the cellular organization was triggered by the symbiosis. In contrast, the archezoon hypothesis posits that at least some major features of the eukaryotic organizational complexity, such as the endomembrane system and the cytoskeleton, evolved before and independent of the symbiosis, and were already in place in the organism that hosted the mitochondrion. Under the archezoon scenario, the presence of genes of apparent archaeal origin in the ancestral Eukaryotic gene set is explained by postulating that the proto-eukaryotic lineage was a sister group of Archaea or possibly even a derivative of one of the archaeal lineages. The archezoon hypothesis was severely damaged by the demonstration that all unicellular eukaryotes previously thought to be primitively amitochondrial actually possess derived organelles of α-proteobacterial descent. Nevertheless, the
The archezoan scenario has not been abandoned but, rather, was amended with the additional proposition that the ancestral Archezoan lineage had gone extinct (Poole and Penny 2007a,b). A purported key argument in support of the archezoan hypothesis is that the complex organization of the eukaryotic cell, in particular, the cytoskeleton, which provides for the capability of phagocytosis, would enable engulfment of potential endosymbionts (Cavalier-Smith 2009).

Another aspect of eukaryogenesis that received considerable attention ever since the discovery of the Archaea is the exact nature of the evolutionary relationship between Archaea and eukaryotes. The “standard model” of Woese and colleagues clearly defines these two domains of cellular life as sister groups (Woese et al. 1990; Pace 1997, 2006). However, very shortly after the publication of the three-domain tree of life, an alternative relationship has been inferred from phylogenetic trees of the same 16S rRNA that were constructed using a different method. The topology of these alternative trees underlies the eocyte hypothesis according to which eukaryotes evolved from within the Archaea and are a sister group to the “eocytes,” that is, the branch that is currently known as Crenarchaeota (Lake et al. 1984; Lake 1988, 1998; Rivera and Lake 1992). Support for the eocyte hypothesis has been subsequently claimed from comparative analysis of ribosomal protein sequences (Vishwanath et al. 2004) and from a novel phylogenomic approach (Rivera and Lake 2004). Most convincingly, a subsequent phylogenetic analysis of multiple conserved genes that used the technique that eliminates fast-evolving alignment columns and is supposed to minimize the effect of several common artifacts of phylogenetic analysis (long-branch attraction and others) supported the affinity of eukaryotes with Crenarchaeota (Cox et al. 2008). The investigators of this study suggested that the alternative topologies including the standard three-domain phylogeny resulted from such undetected artifacts.

However, the eocytes are by no means the only proposed candidate for the role of the archaean ancestor of eukaryotes. The origin of the “archaeal” genes of eukaryotes from Euryarchaeota and, specifically, from methanogens was proposed on the basis of biological considerations, within the framework of the hydrogen hypothesis (Martin and Muller 1998) and one of the so-called syntrophic hypotheses (Lopez-Garcia and Moreira 2006), which postulate different forms of metabolic cooperation between the archaean and bacterial partners of the primary endosymbiosis. The origin of these genes from methanogens also is compatible with the results of some phylogenetic analyses (Moreira and Lopez-Garcia 1998; Horiike et al. 2004). Others have argued that the archaean ancestor of eukaryotes lies outside the known diversity of Archaea, in accord with the standard model, on the basis of biological considerations (Velai et al. 1998) or phylogenetic analysis results (Hedges et al. 2001; Tekai and Yeramian 2005; Ciccarelli et al. 2006; Fukami-Kobayashi et al. 2007).

Most of the aforementioned phylogenetic studies either used a relatively small number of concatenated, highly conserved protein sequences (e.g., those of ribosomal proteins) for phylogenetic tree construction or built trees on the basis of gene repertoire comparisons (phyletic patterns), or else used other features, such as domain architectures of multidomain proteins, as phylogenetic characters. An extensive phylogenomic study that analyzed nearly 6000 gene sets from 185 genomes using a supertree approach has suggested phylogenetic affinity between eukaryotes and Thermoplasma albeit with limited statistical support (Pisani et al. 2007). Thermoplasma or a related, wall-less archaean also has been proposed as the likely archaean ancestor of eukaryotes on the basis of biochemical and cytological considerations (Searcy et al. 1978; Margulis and Stolz 1984; Hixon and Searcy 1993; Margulis 1996; Margulis et al. 2000).

An exhaustive phylogenetic analysis of individual eukaryotic genes of apparent archaean origin suggested a deep (i.e., outside of the known archaean diversity) origin for most of these genes but also identified a considerable number of genes with a crenarchaeal affinity and a small number of genes with a euryarchaeal
affinity (Yutin et al. 2008). In this work, the possibility was introduced that these results did not necessarily result from artifacts of phylogenetic analysis but, rather, might reflect diverse sources of eukaryotic genes due to horizontal gene transfer (HGT) between different Archaea.

Such was the state of the art in the study of the archaeal roots of eukaryotes—far from certainty, even if leaning toward the crenarchaeal (eocyte) origin—when genome analysis of several poorly characterized Archaea suggested the existence of multiple new phyla and yielded unexpected discoveries that seem to call for a change of perspective on the origin of eukaryotes. In this article, we examine the scattered distribution of apparent ancestors of signature eukaryotic genes among Archaea and develop the “dispersed eukaryome” scenario for the origin of eukaryotes.

NEW ARCHAEAL PHYLA, THE TACK SUPERPHYLUM, AND THE COMPLEX ARCHAEAL ANCESTOR

The new outlook of archaeal diversity and evolution emerged in 2008, with the sequencing of the genome of Candidatus Korarchaeum cryptofilum, the first (and so far the only) representative of a poorly characterized group of uncultured Archaea known as Korarchaeota (Elkins et al. 2008) and the almost simultaneous demonstration that mesophilic marine Archaea, previously included in the Crenarchaeota, represented a distinct lineage that was roughly equidistant from Crenarchaeota and Euryarchaeota, and was denoted Thaumarchaeota (Brochier-Armanet et al. 2008). Subsequent studies have shown that Thaumarchaeota are a widespread microbial group of major geochemical importance that includes the key ammonia oxidizers in marine and soil habitats (Pester et al. 2011; Brochier-Armanet et al. 2012; Lloyd et al. 2013).

Phylogenomic analysis enabled by the sequencing of the first representative genome of the Korarchaeota also suggested a new phylum that either is distantly related to Crenarchaeota or branched off before the Euryarchaeota–Crenarchaeota split (Elkins et al. 2008). Subsequent studies have supported these observations and revealed a substantial diversity of Korarchaeota in various marine and terrestrial habitats (Reigstad et al. 2010; Miller-Coleman et al. 2012). The more recent sequencing of the uncultivated archaeon Candidatus Cardiarchaeum subterraneum pointed to the existence of yet another archaeal phylum, dubbed Aigarchaeota (Nunoura et al. 2011). Additional archaeal phyla are likely to lurk within the uncultured microbial diversity, as suggested by 16S RNA phylogenies (Robertson et al. 2005; Teske and Sorensen 2008; Pace 2009; Brochier-Armanet et al. 2011; Rajendhran and Gunasekaran 2011) and the emerging results of single-cell genomics (Lloyd et al. 2013), but, even with the currently available genomes that provide for a more satisfactory phylogenomic analysis, the complex picture of archaeal evolution is starting to take shape.

Several phylogenetic analyses of multiple conserved genes seem to consistently reveal the existence of a deeply rooted archaeal “TACK” superphylum that encompasses Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota (Guy and Ettema 2011; Williams et al. 2012; Wolf et al. 2012; Yutin et al. 2012; Martijn and Ettema 2013). The validity of the TACK superphylum is further supported by the reconstruction of the evolution of the archaeal gene repertoire, which assigned a massive gene gain to the base of the TACK (Fig. 1) (Wolf et al. 2012). This reconstruction, in an extension of previous analyses (Makarova et al. 2007; Csurós and Miklós 2009), reveals a striking trend in the evolution of Archaea that might reflect a general tendency of genome evolution, namely, the preponderance of genome reduction that is punctuated with episodes of explosive genome expansion (Wolf and Koonin 2013). According to these reconstructions, each of the major archaeal lineages underwent some degree of genome reduction, and the gene complement of the last archaeal common ancestor (LACA) was at least as complex as that of most of the extant Archaea. The reductive evolution associated with genome streamlining appears to be a general feature of successful lineages that reach large effective population sizes and are subject to strong selective pressure. However,
The evolution of the Archaea might have involved additional factors leading to genome reduction such as the adaptation to the thermophilic lifestyle (Forterre 1995; Boussau et al. 2008).

The delineation of new archaeal phyla and the putative TACK superphylum stimulated further phylogenomic effort aimed at the identification of the archaeal ancestor of eukaryotes. One such study has suggested an affinity of eukaryotes with Thaumarchaeota (Kelly et al. 2011). However, the two most detailed to date, state-of-the-art phylogenetic analyses of rRNA and universal protein-coding genes provided significant support for the monophyly of eukaryotes with the TACK superphylum but failed to recover a specific relationship between eukaryotes and any lineage within the TACK (Williams et al. 2012; Lasek-Nesselquist and Goeggelen 2013). Thus, the current last word on the evolutionary relationship between the cores of the information-processing systems of the Archaea and eukaryotes seems to be that the ancestors of eukaryotes are not a sister group to all Archaea but are more likely to comprise a deep archaeal branch sharing a common ancestry with the TACK superphylum (Fig. 1).

Figure 1. A reconstruction of the evolution of archaeal gene complements and the inferred origins of the eukaryome components. The boxes show the approximate maximum-likelihood estimates for the number of genes in extant and reconstructed ancestral genomes according to the color code shown on the right. The reconstruction was performed using the COUNT software (Csuros 2010). The tree topology is from the phylogenetic tree of concatenated ribosomal proteins (Wolf et al. 2012; Yutin et al. 2012). The inferred origins of some key eukaryotic genes and functional systems are indicated by arrows (see text); S30, L25, and L13 are ribosomal proteins. (From Wolf et al. 2012; modified, with permission, from the author.)
assignment is compatible with the identification of several ancestral genes that eukaryotes exclusively shared with the TACK superphylum, in contrast to a smaller number of such derived shared characters in Euryarchaeota (Fig. 1). Notably, the shared derived characters of eukaryotes and the TACK superphylum include components of the information system core such as three ribosomal proteins (Yutin et al. 2012), the RNA polymerase subunits RPB8 (Koonin et al. 2007) and RPC34 (Blombach et al. 2009), and the transcription factor Elf1 (Daniels et al. 2009).

THE DISPERSED ARCHAEAL EUKARYOME

Perhaps, the most remarkable discoveries coming from the recent comparative analysis of archaeal and eukaryotic genomes go beyond the identification of the specific eukaryotic ancestry within Archaea. The unexpected outcome of these comparisons is that, apart from the core of the information systems, the apparent archaeal ancestors of essential eukaryotic genes and entire functional systems—the set of genes that we here denote the “eukaryome”—are often found in different groups of Archaea.

Figure 2 shows the breakdown of the (putative) prokaryotic homologs of ancestral eukaryotic genes by the number of major groups of Archaea and Bacteria in which the respective genes are represented. Focusing on the archaeal homologs of eukaryotic genes, an enrichment in genes that are represented in only one archaeal lineage (and to a lesser extent, in two to four lineages) is notable. A further classification of the 77 putative ancestral genes that were identified in a single group of Archaea shows a wide spread across the entire archaeal domain (Fig. 2). Of special interest seems to be the enrichment of lineage-specific archaeal ancestral genes in Ca. Caldiarchaeum subterraneum, the only known representative of Aigarchaeota. Below we discuss some of the most striking, biologically relevant cases of the dispersal of eukaryome components among Archaea.

The Ubiquitin Signaling System

The ubiquitin (Ub) system is the extremely complex machinery that regulates protein degradation, topogenesis, and function in all eukaryotes through modification of proteins by conjugation with various forms of (poly)Ub and its paralogs (Burroughs et al. 2012a,b; van der Veen and Ploegh 2012). For many years, the Ub system was viewed as a eukaryotic innovation that seemed to have evolved from prokaryotic enzymes involved in coenzyme biosynthesis (Hochstrasser 2000). More recently, it has been shown that, with the participation of a homolog of the E1 subunit of eukaryotic Ub ligases, some of the archaeal homologs of Ub are conjugated with proteins and target them for degradation (Humbard et al. 2010; Miranda et al. 2011; Maupin-Furlow 2013). However, these are distant homologies, so the origin of the E2 and E3 subunits of the Ub ligases remained elusive.

Analysis of the genome of Ca. Caldiarchaeum subterraneum dramatically transformed the entire story of the origin of the Ub system by revealing a predicted operon encoding a Ub-like protein and homologs of all three Ub ligase subunits, along with a key deubiquitinating enzyme (Nunoura et al. 2011); operons with a similar organization of Ub-related genes are detectable also in several Bacteria, suggestive of horizontal dissemination of these operons among prokaryotes (Fig. 3A). Unlike the homologs of Ub system components detected previously in other Archaea, the homologs from Ca. Caldiarchaeum subterraneum show high sequence similarity to eukaryotic counterparts, and phylogenetic trees for the E1 subunit of Ub ligase (Fig. 3B) and the deubiquitinating enzyme MPN (Fig. 3C) place Ca. Caldiarchaeum subterraneum in the midst of eukaryotes. The possibility of acquisition of the Ub system by Ca. Caldiarchaeum subterraneum via HGT from a eukaryotic source can be effectively discarded given the operonic organization of the archaeal genes. Thus, these findings indicate that, to date, Ca. Caldiarchaeum subterraneum encodes the best candidate for the ancestor of the eukaryotic Ub system.

Notably, apparent archaeal orthologs of the distant Ub homolog Urm1, which is conserved in all eukaryotes and performs a dual function as a sulfur carrier in tRNA thiolation and in protein modification (Van der Veen et al. 2011; Wang...
et al. 2011), are detectable only among the members of the order Sulfolobales of the Crenarchaeota (Fig. 1) (Makarova and Koonin 2010).

Thus, within the broadly defined Ub system, two different archaeal ancestors of essential eukaryotic functional modules are detectable.

The Cytoskeleton

All eukaryotic cells possess an advanced, extremely elaborate cytoskeleton that consists of two major types of structural elements, actin-based filaments, and tubulin-based microtubules (Lowe and Amos 2009; Aylett et al. 2011). The cytoskeleton looms large in the discussion of the origin of eukaryotes, primarily because actin filaments play the central role in phagocytosis, the process that is thought to be critical for the engulfment of the proto-mitochondrial endosymbiont by its elusive future host (Cavalier-Smith 2009). For years, Bacteria and Archaea have been thought to encode only distant homologs of actin and tubulin, the proteins of the MreB/FtsA and FtsZ families, respectively, that perform essential functions in the septation of bacterial and some archaeal cells (Cabeen and Jacobs-Wagner 2010; Celler et al. 2013). However, analysis of the expanding archaeal genome collection overturned this perspective. Proteins with high sequence similarity...
E.V. Koonin and N. Yutin

Figure 3. The apparent archaeal ancestor of the ubiquitin system. The operon organization of the Ub system components in Ca. Caldarchaeum subterraneum and similar operon structures in Bacteria. (A) A maximum-likelihood tree for the E1 component of the Ub ligase. (B) A maximum-likelihood tree for the MPN deubiquitinating enzyme. The sequences for the tree construction were retrieved from GenBank nr and env_nr databases (NCBI, NIH) and aligned using the MUSCLE program (Edgar 2004). Positions that included gaps in more than one-third of the sequences and positions with low information content were removed before tree computation (Yutin et al. 2008), which left 133 unambiguous positions in the E1 alignment and 103 positions in the MPN alignment. Maximum-likelihood trees were constructed using the TreeFinder program (WAG matrix, G[Optimum]:4, 1000 replicates, Search Depth 2) (Jobb et al. 2004). The bootstrap values (shown for selected branches) represent expected-likelihood weights from 1000 local rearrangements. Branches with bootstrap support $<0.5$ were collapsed. For each sequence, the species name abbreviation and the gene identification numbers are indicated; (env) marine metagenome. Species: Acibo, Acidiliprodundum boonei T469; Acisa, Acidilobus saccharovorans 345-15; Aerpe, Aeropyrum pernix K1; Arath, Arabidopsis thaliana; Arcfu, Archaeoglobus fulgidus DSM 4304; Arcpr, Archaeoglobus profundus DSM 5631; Bacce, Bacillus cereus G9241; Barhe, Bartonella henselae strain. Houston-1; Caeb, Caenorhabditis briggsae AF16; Cafro, Methanocella arvoryzae MRE50; Can_Csub, Candidatus Caldarchaeum subterraneum; CanKo, Candidatus Korarchaeum cryptofilum OPF8; Censy, Cenarchaeum symbiosum A; Cloph, Clostridium phytofermentans ISDg; Dicdi, Dictyostelium discoideum AX4; Dicze, Dickeya zeae Ech1591; Drome, Drosophila melanogaster; Enthi, Entamoeba histolytica HM-1:IMSS; Ferac, Ferroplasma acidarmanus fer1; Ferpl, Ferroglobus placidus DSM 10642; Fraal, Frankia alni ACN14a; Guith, Guillardia theta; Halbo, Halogeogetrichum borinquense DSM 11551; Halla, Halorubrum lacusprofundi ATCC 49239; Halma, Haloarcula marismortui ATCC 43049; Halwa, Haloquadratum walsbyi DSM 16790; Helmo, Heliobacterium modesticaldum Ice1; Heman, Hemiselmis andersenii; Homsa, Homo sapiens; Isopa, Isosphaera pallida ATCC 43644; Metar, Methanocella arvoryzae MRE50; Methbo, Methanoregula boonei 6A8; Metbu, Methanosphaeides burtonii DSM 6242; Methu, Methanopirillum hungatei IF-1; Metma, Methanohalophilus mahii DSM 5219; Metpa, Methanospaeroida palustris E1-9c; Metru, Methanobrevibacter ruminantium M1; Metth, Methanoseta thermophila PT; Naegr, Naegleria gruberi strain NEG-M; Natph, Natronomonas pharaonis DSM 2160; Nemve, Nematostella vectensis; Nitma, Nitrosopumilus maritimus SCM1; Parte, Paramecium tetraurelia; Phatr, Phaeodactylum tricornutum CCAP 1055/1; Picto, Picrophilus torridus DSM 9790; Pirst, Pirellula staleyi DSM 6068; Playo, Plasmodium yoelii yoelii 17XNL; Proma, Prochlorococcus marinus str. MIT 9303; Pyrab, Pyrococcus abyssi GE5; Pyra, Pyrrobacteriarnunciaticum DSM 13514; Pyrica, Pyrobaculum calidifontis JCM 11548; Pyru, Pyrococcus furiosus DSM 3638; Rotde, Rothia dentocariosa M567; Sacce, Saccharomyces cerevisiae S288c; Stahe, Staphylococcus hominis DSM 12710; Strpu, Streptococcus pyogenes 17A; Stre, Streptomyces vescus ATCC 29083; Tetth, Tetrahymena thermophila; (Legend continues on following page.)
Figure 3. (Continued). Theko, Thermococcus kodakarensis KOD1; Thene, Pyrobaculum neutrophilum V24Sta; Theon, Thermococcus onnurineus NA1; Thesi, Thermococcus sibiricus MM 739; Thesp, Thermococcus sp. AM4; Thevo, Thermoplasma volcanium GSS1; Triva, Trichomonas vaginalis G3; uncar, uncultured archaeon GZfos18C8; uncCa, uncultured Candidatus Nitrosocaldus sp.; uncncr, uncultured crenarchaeote; unccr, uncultured marine crenarchaeote KM3-86-C1; Ustma, Ustilago maydis 521. Taxa: Ac, Crenarchaeota; Ae, Euryparchaeota; Ak, Korarchaeota; At, Tautomarchaeota; Ax, environmental samples; Ba, Actinobacteria; Bc, Cyanobacteria; Bf, Firmicutes; Bo, Planctomycetes; Bp, Proteobacteria; E8, Stramenopiles; E9, Viridiplantae; Ea, Amoebozoa; Ec, Alveolata; Eh, Cryptophyta; El, Opisthokonta; Eq, Heterolobosea; Ew, Parabasalia.
and phylogenetic affinity to eukaryotic actins were discovered in the order of Thermopro-
teales of the Crenarchaeota, Korarchaeum and Can. C. subterraneum, suggesting that actin was already present in the last common ancestor of the TACK superphylum of the Archaea (Yutin et al. 2009; Makarova et al. 2010). Following this comparative-genomic finding, it has been

C

Figure 3. Continued.
shown that these archaeal proteins, dubbed crenactins, indeed form helical filaments resembling typical eukaryotic actin filaments (Bernander et al. 2011; Ettema et al. 2011).

Highly conserved orthologs of tubulins so far have been discovered only in the genomes of two Thaumarchaeota. In this case, there is no operon organization to rule out HGT from eukaryotes to Archaea, but the results of phylogenetic analysis appear to be best compatible with these “artubulins” being the ancestors of the eukaryotic tubulins (Yutin and Koonin 2012). The functions of the artubulins remain to be characterized.

Thus, contrary to recent beliefs, both major forms of the eukaryotic cytoskeleton apparently constitute part of the archaeal heritage although the apparent ancestral forms were detected in different extant Archaea.

**Cell Division and Membrane-Remodeling Systems**

Obviously, cell division is an essential process in all cellular life forms. Nevertheless, there are fundamental mechanistic differences between the division mechanisms in Bacteria and at least some Archaea compared with eukaryotes. In Bacteria, cell division is coupled to DNA replication whereby the replicating DNA is immediately pumped into the daughter cells, and the division is completed through the formation of the septum, which starts with the Z-ring that consists of the FtsZ protein, a distant homolog of eukaryotic tubulins (Adams and Errington 2009; Margolin 2009). The FtsZ-based division system is present in most of the Euryarchaeota, Thaumarchaeota, and Korarchaeota, with the implication that these archaeal cells divide similarly to Bacteria (Makarova et al. 2010). Recently, however, a distinct division system homologous to the eukaryotic ESCRT-III membrane-remodeling complex has been experimentally characterized in the crenarchaeon *Sulfolobus* (Lindas et al. 2008; Samson et al. 2008; Ettema and Bernander 2009; Samson and Bell 2009) and subsequently identified with comparative-genomic methods in all Sulfolobales and Desulfurococcales (two of the three crenarchaeal orders) as well as some Euryarchaeota (Makarova et al. 2010). A follow-up study has shown that the ESCRT-III-like complex is the primary cell division system in the thaumarchaeon *Nitrosopumilus maritimus* (Pelve et al. 2011). The broad, even if patchy, distribution of the FtsZ-based and ESCRT-III-like division systems in Archaea implies that both were present in the LACA, with subsequent differential losses in multiple lineages. However, the Crenarchaeota of the order Thermoproteales lack both of these division systems, implying the existence of a third one, conceivably centered around the crenactin cytoskeleton (Makarova et al. 2010; Bernander et al. 2011). Given that the origin of crenactin can be mapped to the base of the TACK superphylum (Fig. 1; see above), one comes to the striking conclusion that the common ancestor of the TACK possessed all three cell division systems that are scattered among the extant Archaea.

**The RNA Interference System (RNAi)**

The RNA interference system (RNAi) is a major eukaryotic hallmark system that is involved in both antivirus defense and regulation of gene expression and apparently was already in place in the LECA (Cerutti and Casas-Mollano 2006; Shabalina and Koonin 2008; Carthew and Sontheimer 2009). The multiple RNAi mechanisms are centered around proteins of two families, the Dicers that are responsible for the processing of small interfering RNAs and microRNAs, and the Argonautes, some of which are nucleases directly attacking the RNA target (slicers), whereas others bind microRNA and guide it to the target, cleaving the latter (Tolia and Joshua-Tor 2007; Hock and Meister 2008; Jasakiewicz and Filipowicz 2008). The Argonautes have numerous homologs in Bacteria and Archaea, primarily Euryarchaeota, and phylogenetic analysis clearly points to the euryarchaeal origin of the eukaryotic Argonaute family (Makarova et al. 2009). The Dicers have no direct counterparts in Bacteria or Archaea. These proteins comprise a fusion of a helicase domain and two RNase III domains that is unique to eukary-
otes. The RNase III domain is of obvious bacterial provenance, whereas the helicase domain of Dicer has been traced to a euryarchaeal ancestor, the Hef helicase that is involved in DNA replication (Shabalina and Koonin 2008). Thus, the RNAi system, a hallmark eukaryotic machinery that has no direct counterpart in Archaia or Bacteria, has mixed archaeal and bacterial origins, with the archaeal components apparently derived from Euryarchaeota.

THE ARCHAEOAL ANCESTOR OF EUKARYOTES: A COMPLEX, ANCIENT LINEAGE IN THE TACK SUPERPHYLUM?

Combined with the results of phylogenomic analysis, the discovery of the dispersed archaeal eukaryome implies a complex archaeal ancestor of eukaryotes, which encoded the ancestors of various eukaryotic functional systems that are currently found in different lineages of extant Archaea (Fig. 1). Specifically, the hypothetical eukaryotic ancestor probably possessed a fairly advanced cytoskeleton that encompassed both actin filaments and tubulin microtubules and could provide for a primitive phagocytic capability (Yutin et al. 2009; Martijn and Ettema 2013). It stands to reason that the eukaryotic ancestor was a wall-less, mesophilic organism that coexisted with diverse Bacteria, so even a limited capacity for phagocytosis would greatly facilitate the capture of prospective endosymbionts. Extant mesophilic Archaea such as Methanosarcina or Halobacteria clearly have acquired numerous genes via HGT (Koonin et al. 2001; Deppenmeier et al. 2002; Nelson-Sathi et al. 2012; Wölf et al. 2012). When it comes to the ancestor of eukaryotes, this gene-gaining capacity might have been further enhanced by the primitive phagocytosis, conceivably through transient engulfment of other Archaea and Bacteria (Fig. 4). Such a lifestyle could result both in the accumulation of the eukaryome components that are currently found in diverse Archaea (see above) and in the acquisition of genes from diverse bacterial sources, possibly explaining the weakness of the α-proteobacterial signal among the eukaryotic genes of apparent bacterial descent (Esser et al. 2007).

As discussed above, phylogenomic analysis suggests that the ancestor of eukaryotes could have been a deep branch within the TACK superphylum that radiated from the other phyla shortly after the emergence of the TACK (Wil-
Thus, the hypothetical lineage of the eukaryotic ancestor would have the same evolutionary status as the currently known phyla, Crenarchaeota, Thaumarchaeota, Korarchaeota, and Aigarchaeota. Evolutionary reconstructions suggest that archaeal evolution was dominated by genome reduction on a path from complex ancestors to streamlined extant genomes (Wolf et al. 2012; Wolf and Koonin 2013). The ancestor of eukaryotes could have been an exceptionally complex ancient archaeon. Hence, a notable possibility: one of the extant phyla within the TACK, for example, Aigarchaeota or more likely another group lurking among the still uncultured Archaea, could be the ancestral lineage of the eukaryotes. We have difficulty ascertaining this relationship because of the contrasting trends of genome reduction in Archaea and complexification in eukaryotes, the likely compressed cladogenesis (Rokas and Carroll 2006; Puigbo et al. 2009) at the base of the TACK superphylum and acceleration of evolution during eukaryogenesis. It is interesting to point out the discovery of giant Thaumarchaeota associated with bacterial ectoparasites (Muller et al. 2010). Such forms might resemble the ancestor of the eukaryotes, and genome sequences of these and other “exotic” Archaea are eagerly anticipated.

The proposition that the archaeal ancestor of eukaryotes (the host of the mitochondrial endosymbiont) was a complex organism that possessed some signature features of eukaryotic cells (such as cytoskeleton) begs the question: Was the eukaryotic ancestor an archaeon or actually an Archeozoan? This question potentially could be considered semantic, but we are inclined to take it as a substantial one given the qualitative differences between the cellular organizations in eukaryotes versus Archaea and Bacteria (prokaryotes). Given the apparent unsustainability of the eukaryotic genome expansion and cell organization in the absence of multiple energy-converting organelles (such as mitochondria) (Lane and Martin 2010) and the lack of a plausible function for the nucleus in the context of the prokaryotic cell (Koonin 2006; Lopez-Garcia and Moreira 2006; Martin and Koonin 2006), we find it overwhelmingly more likely that the host of the mitochondrial endosymbiont was a bona fide archaeon.

CONCLUSIONS

Deep phylogenies are notoriously artifact-prone, and deriving definitive topology of the major branches of organisms is extremely hard; the results of independent analysis are often in conflict, leading to the notion of a “phylogenomic impasse” with regard to the origin of eukaryotes (Gribaldo et al. 2010). However, these legitimate concerns notwithstanding, there seems to be some emerging clarity with respect to the nature of the archaeal ancestor of eukaryotes. The two key observations are the apparent deep phylogenetic affinity of the core of the eukaryotic information-processing machinery with the archaeal TACK superphylum and the dispersal of the eukaryome components across Archaea. The combination of these findings implies a highly complex archaeal ancestor of eukaryotes that possessed certain signature eukaryotic features, such as the cytoskeleton and the Ub system, while remaining a typical archaeon in terms of overall cellular organization and genome structure. The presence of a well-developed cytoskeleton could facilitate the engulfment of Bacteria, creating the conditions for the evolution of endosymbiosis. The complexity of the archaeal ancestor was apparently fixed in the emerging eukaryotes thanks to endosymbiosis. In contrast, the proto-eukaryotic features were differentially lost in archaeal lineages in the course of reductive evolution, resulting in the currently observed dispersed eukaryome. Given the dispersed eukaryome, extensive sampling of the archaeal diversity by genome sequencing is essential to advance our understanding of eukaryogenesis.

ACKNOWLEDGMENTS

We thank Thijs Ettema, Kira Makarova, and Bill Martin for useful discussions. Our research is supported by intramural funds of the U.S. Department of Health and Human Services (to the National Library of Medicine, NIH).
REFERENCES


Cite this article as: Cold Spring Harb Perspect Biol 2014;6:a016188

Archaeal Ancestry of Eukaryotes

E.V. Koonin and N. Yutin


The Dispersed Archaeal Eukaryome and the Complex Archaeal Ancestor of Eukaryotes

Eugene V. Koonin and Natalya Yutin

Cold Spring Harb Perspect Biol 2014; doi: 10.1101/cshperspect.a016188

Subject Collection The Origin and Evolution of Eukaryotes

The Persistent Contributions of RNA to Eukaryotic Gen(om)e Architecture and Cellular Function
Jürgen Brosius

Green Algae and the Origins of Multicellularity in the Plant Kingdom
James G. Umen

The Archaeal Legacy of Eukaryotes: A Phylogenomic Perspective
Lionel Guy, Jimmy H. Saw and Thijs J.G. Ettema

Origin and Evolution of the Self-Organizing Cytoskeleton in the Network of Eukaryotic Organelles
Gáspár Jékely

On the Age of Eukaryotes: Evaluating Evidence from Fossils and Molecular Clocks
Laura Eme, Susan C. Sharpe, Matthew W. Brown, et al.

Origin of Spliceosomal Introns and Alternative Splicing
Manuel Irimia and Scott William Roy

Protein and DNA Modifications: Evolutionary Imprints of Bacterial Biochemical Diversification and Geochemistry on the Provenance of Eukaryotic Epigenetics
L. Aravind, A. Maxwell Burroughs, Dapeng Zhang, et al.

The Eukaryotic Tree of Life from a Global Phylogenomic Perspective
Fabien Burki

Eukaryotic Origins: How and When Was the Mitochondrion Acquired?
Anthony M. Poole and Simonetta Gribaldo

Bacterial Influences on Animal Origins
Rosanna A. Alegado and Nicole King

Missing Pieces of an Ancient Puzzle: Evolution of the Eukaryotic Membrane-Trafficking System
Alexander Schlacht, Emily K. Herman, Mary J. Klute, et al.

The Neomuran Revolution and Phagotrophic Origin of Eukaryotes and Cilia in the Light of Intracellular Coevolution and a Revised Tree of Life
Thomas Cavalier-Smith

Protein Targeting and Transport as a Necessary Consequence of Increased Cellular Complexity
Maik S. Sommer and Enrico Schleiff

How Natural a Kind Is "Eukaryote?"
W. Ford Doolittle

What Was the Real Contribution of Endosymbionts to the Eukaryotic Nucleus?
Insights from Photosynthetic Eukaryotes
David Moreira and Philippe Deschamps

Bioenergetic Constraints on the Evolution of Complex Life
Nick Lane

For additional articles in this collection, see http://cshperspectives.cshlp.org/cgi/collection/