The expanding Asgard archaea invoke novel insights into Tree of Life and Eukaryogenesis

Zhichao Zhou[1,†], Yang Liu[2,3,†], Karthik Anantharaman[1], Meng Li[2,3,*]

[1] Department of Bacteriology, University of Wisconsin–Madison, Madison, WI, 53706, USA

[2] Archaeal Biology Center, Institute for Advanced Study, Shenzhen University, Shenzhen, 518060, China.

[3] Shenzhen Key Laboratory of Marine Microbiome Engineering, Institute for Advanced Study, Shenzhen University, Shenzhen, 518060, China

[†]Zhichao Zhou and Yang Liu contributed equally to this review and Yang Liu is the co-first author.

[*]Correspondence: Meng Li, limeng848@szu.edu.cn

Abstract

The division of organisms on the tree of life into either a three-domain (3D) tree or a two-domain (2D) tree has been disputed for a long-time. Ever since the discovery of Archaea by Carl Woese in 1977 using 16S rRNA sequence as the evolutionary marker, there has been a great advance in our knowledge of not only the growing diversity of Archaea but also the evolutionary relationships between different lineages of living organisms. Here, we present this perspective to summarize the progress of archaeal diversity and changing notion of the Tree of Life. Meanwhile, we provide the latest progress in genomics/physiology-based discovery of Asgard archaeal lineages as the closest relative of Eukaryotes. Furthermore, we proposed three major directions for future research on exploring the "next one" closest Eukaryote relative, deciphering the function of archaeal eukaryotic signature proteins (ESPs) and eukaryogenesis from both genomic and physiological aspects, and understanding the roles of horizontal gene transfer (HGT), viruses, and mobile elements in eukaryogenesis.

The discovery of Archaea and two notions of the Tree of Life

The wide usage of 16S ribosomal RNA as the gold standard for determining the taxonomy of prokaryotes roots back to the late 1970s. In 1977, Woese and Fox discovered a new kind of microbial life that is not the same as typical bacteria based on 16S rRNA phylogeny (1). They initially named them "archaebacteria", representing a new group of life besides Bacteria and Eukarya. This is the emergence of the notion of a three-domain (3D) tree of life represented by – Bacteria, Archaea, and Eukarya (Figure 1A). They proposed this name because they believed "archaebacteria" were ancient, living in extreme environments, and had robust evolutionary connections to the organisms living on early earth (2). The short name "archaea" was then widely adopted by the scientific community after this finding was gradually accepted.

Nevertheless, in the late 1980s, there came two theories to deal with the placement of Archaea and Eukarya. One was the "Eocyte tree" model developed by James Lake (3). Eocytes, a group of extreme thermophiles without a nucleus, are believed to be the ancestors of Eukaryotes. They formed the "Karyotes" branch with Eukaryotes, comparable to the other branch composed of Prokaryotes (Figure 1B) (3, 4). On the contrary, Woese *et al.* claimed the division of Prokaryotes into two life domains – Bacteria and Archaea, and posited the Eukaryote branch as an independent lineage (Figure 1A) (1). Beyond this, other pieces of biochemical evidence were added; phylogenetic trees based on proteins encoding the information-processing machinery (RNA polymerases I, II, and III, DNA-dependent RNA polymerase, and ribosomal protein L11, L1, L10, and L12) were observed to be congruent with the 3D tree (5-7).

When it comes to the late 1990s and 2000s, the topology of 3D tree was depicted with the following features (8): 1) the root of three domains could not be determined from rRNA sequences; 2) Eukarya and Archaea had a common ancestral branch which was different from the Bacterial branch; 3) Eukarya and Archaea were close but were separated into two branches, in this manner, Archaea were distantly related to Bacteria, they were not monophyletic as previously thought (9). The establishment of the 3D tree since the late 1970s soon replaced the old taxonomic clustering system mainly based on morphological subjects (9). It challenged the simple division of Eukarya and Prokarya by the existence of a membrane-bounded nucleus. In terms of the evolutionary line of cell nucleus genetic material, the Eukarya line is as ancient as that of Archaea in the ribosomal

RNA-based 3D tree (Figure 1A), in this way, Eukarya was posited not directly evolving from either Archaea or Bacteria.

The Growing diversity of Archaea

Currently, based on NCBI Taxonomy data updated through October 2022, there are 39 phylum-level archaeal lineages: Euryarchaeota, Thermoplasmatota, Hydrothermarchaeota, TACK group (n=10), DPANN group (n=11), and Asgard group (n=15). While the classification of Archaea started from quite a few numbers of lineages. Ever since the establishment of Archaea, Euryarchaeota and Crenarchaeota were the first two major clades identified in Archaea (10). Euryarchaeota contains methanogens, anaerobic methanotrophs, and halophilic archaea (10). It also includes thermophilic *Thermococcales* which usually are acidophilic and do not contain a cell wall (10). Thermoplasmatota was first proposed by Rinke et al. (11) in 2019. It contains Poseidoniia (formerly Marine Group II and III), Thermoplasmata, and DHVE2 group. Hydrothermarchaeota was formerly the Marine Benthic Group E and was proposed as a new archaea phylum by Carr et al. (12) in 2019. It is named after its first discovered place of the continental slope and abyssal sediments (13).

Crenarchaeota was the most abundant component within the TACK group (including Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota). Crenarchaeota mostly comprises of hyperthermophiles that are discovered in geothermally-heated/acidic hot spring environments (10). Most of them are obligate anaerobes with either chemoorganotrophic or chemolithotrophic lifestyles (10). Many of them also live in hot springs, submarine volcanic, or hydrothermal vent habitats (10). In the middle 1990s, Korarchaeota was added to the tree of life (Figure 1C) (14). *Korarchaeum cryptofilum* is the first representative of this phylum (15). It is an obligate anaerobe with a chemoorganotrophic and hyperthermophilic lifestyle. From its genome, it is posited to live a fermentative lifestyle using peptides and amino acids (10). Additionally, it lacks the machinery to perform anaerobic respiration and likely depends on other community members in the form of a mutual dependence relationship to acquire co-factors (10). It was assumed that all archaeal lineages were obligate anaerobes and lived in extreme environments until the discovery of Thaumarchaeota in the 2000s. The first isolated strain of Thaumarchaeota, *Nitrosopumilus maritimus*, is a representative of their aerobic lifestyle dependent on oxidizing ammonia to nitrite (16). It is adapted to a life that can grow at ammonia concentrations that are one hundred times lower than those required by bacterial nitrifiers, which explains their widespread distribution in the open ocean (10). Aigarchaeota was formerly the HWCG I group and was proposed to be a new phylum by Nunoura et al. (17) in 2010. It contains two candidate genera – *Caldiarchaeum* and *Calditenuis*. The study of the Aigarchaeota genome reveals the first evidence of eukaryotic ubiquitin system homologues in archaea (17).

The DPANN superphylum of Archaea was first proposed in 2013 (18). The name is an acronym of the initials of the first five groups discovered – Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, and Nanohaloarchaeota. While, later on, many other groups were added – Altiarchaeota, Huberarchaea, Micrarchaeota, Pacearchaeota, Undinarchaeota, and Woesearchaeota. They are characterized as ultra-small archaea with nanometric cell size. They have small genome sizes with limited metabolic capabilities. Many members of DPANN depend on symbiotic/parasitic interactions with other organisms (19, 20). The specific placement of DPANN lineages on the archaeal tree is somewhat controversial (20). Studies have claimed DPANN to be a deep-branching archaeal superphylum in the basal position of the archaeal tree (21, 22). While, others have suggested that DPANN lineages are fast-evolving and have high mutational rates causing them to be grouped at the base of the phylogenetic tree caused by long-branch attraction artifacts without being actually related (23).

The emergence and development of Asgard group are described in detail in the following section. Many new lineages under the four above-mentioned archaeal groups have been discovered and characterized by the metagenomics method as well as the single-cell genomics, transcriptomics, proteomics, and enrichment culturing methods. For example, Marsarchaeota (25), Brockarchaeota (26), Bathyarchaeota (27), Geothermarchaeota (28), and Verstraetearchaeota (29) are newly discovered phyla within TACK superphylum. Hadesarchaea (30), Methanofastidiosa (31), Theionarchaea (32), and Nanohaloarchaea (33) are newly discovered classes within Euryarchaeota; Proteinoplasmatales (34) and Thermoprofundales (35) are newly discovered families within Euryarchaeota and Thermoplasmatota. The large volume of newly emerging archaea as well as bacteria has significantly increased the diversity and physiology knowledge while requiring improved classification and nomenclature. Consequently, a new classification platform, GTDB database, arises to achieve a better assessment of the diversity of archaea based on genomic materials (36). Based on the GTDB database (07-RS207), there are 18 phylum-level archaeal lineages. The archaeal reference phylogenomic unrooted tree containing 3412 representative species was inferred from 53 conserved marker proteins and decorated with GTDB taxonomy (Figure S1), showing a scene of flourishing branches and leaves of expanded archaeal lineages. Meanwhile, their corresponding NCBI taxonomy information was also provided (Supplementary Data 1). GTDB uses a relative evolutionary divergence (RED) method to delineate high-ranking microbial taxa, which normalizes the contents of each phylum by quantitative criteria based on genetic distance (36). Currently, a consensus statement for creating consistent rules for nomenclature of uncultivated taxa (37) and SeqCode to use genomic sequence of prokaryote as the nomenclatural type (38) are introduced to the scientific community to solve the issue. The diversity of archaea to discover, as well as the tools and platforms for classifying and naming microbial diversity, will continue in the foreseeable future.

Changing the Tree of Life (3D to 2D)

Since the very beginning of the discovery of Archaea, it has been a dispute to solve the phylogenetic position of Archaea and Eukarya. In the early stage before the 1990s, Woese's 3D tree and Lake's "Eocyte tree" model are the major competing notions. While both trees are quite limited by the small number of archaeal genomes available at that time. For instance, Korarchaeota was the third archaeal clade introduced to the 3D tree until the middle 1990s (14). Korarchaeota branched prior to the bifurcation of Crenarchaeota and Euryarchaeota and even formed a sister clade to Eukarya (14). On the other hand, Eocyte, which was believed to be a new biological kingdom in the "Eocyte tree" model, was later proved to be a member in Crenarchaeota (39). In addition, at that time, Archaea was limitedly represented by methanogens and halobacteria. It appears that the separation between Eocyte and Archaea at that time was caused by the absence of the intermediate archaea. Both examples indicate that the discovery of new archaeal clades will reform our knowledge on the Tree of Life, which leads to a better understanding of the emergence of eukaryotes.

Along with the increasing diversity of Archaea discovered, the general idea of the placement of Eukaryotes in the Tree of Life and the relationship between protoeukaryotes and the Archaea 'stem' in the tree has gradually changed. A key issue is where to place the protoeukaryotes or the last eukaryote common ancestor (LECA) in the phylogenetic tree. It was posited that the notion that LECA is parallel to the branch that leads to extant archaeal lineages brings out the 3D tree model (40). While, placing LECA as rooted from the Archaea 'stem' gives the 2D tree model (40). Ever since the 1990s, increasingly available genomic data of microorganisms and advanced evolutionary reconstruction methods have driven the revival of 2D tree – which is obviously different from the previous paradigm of dividing life forms simply into Eukaryotes and Prokaryotes (40) (based on the presence/absence of nucleus), or Karyotes and Prokaryotes (the "Eocyte tree" model) (3). By analyzing 26 universally conserved protein clusters, one of the first studies with genomic evidence from newly discovered archaea looked back to the 2D tree theory, and Eukarya was placed as the sister clade of the TACK superphylum (41) (Figure 1D). This study suggested that eukaryotes have an archaeal parent that is most likely affiliated with the TACK superphylum, and that archaeal gene sets found in eukaryotes were vertically inherited from the archaeal parent. It was once hypothesized that Eukarya was nested within the TACK superphylum (close to Korachaeota; see Figure 1D topology) or emerging as a sister group adjacent to the TACK superphylum (41, 42).

In 2015, the Asgard superphylum was discovered and characterized as containing the lineages bridging Eukarya and Archaea (43-50). Since the early time of placing Thorarchaeota and Lokiarchaeota as the closest lineages to Eukarya in a 2016 "Tree of Life" (Figure 1E), currently, there have been 17 Asgard lineages reported (Figure 1F). The reconstruction of the phylogeny of Asgard, other archaeal lineages, Bacteria, and Eukarya based on concatenated 29 universal markers supports either the origin of eukaryotes within Asgard superphylum or a deeper branch of eukaryote ancestor within archaea – both suggest a 2D tree topology for the Tree of Life (45). Simultaneously, Asgard archaea have been identified to contain a more expanded repertoire of eukaryotic signature proteins (ESPs) compared to the other archaeal lineages such as the TACK superphylum. The majority of Asgard ESPs belong to the 'intracellular trafficking, secretion and vesicular transport' and 'posttranslational modification, protein turnover and chaperones' functional categories (45). Many of them are from the roadblock superfamily and small GTPases family – two families that constitute or build the intra-cellular membrane-bound compartments of eukaryotes (45, 51). Both the phylogeny and ESP evidence support the archaeal origin of eukaryotes in that protoeukaryotes are either rooted within Asgard or derived from an even basal branch within the archaea domain, thus supporting the topology of a 2D tree of life.

Future directions

Unraveling the phylogenetic relationship between Archaea and eukaryotes and identifying the closest relative of eukaryotes

Phylogenetic advances are needed to determine the placement between archaea and eukaryotes. The key issue is the usage of the marker set of proteins used for these analyses. Normally, the disparity of different inter-phylum relationships is rooted in the choice of marker set. By ranking each marker based on the congruency of supporting the monophyly of well-established archaeal lineages, Dombrowski *et al.* provided two sets of markers with reasonable robustness (52). Meanwhile, the metrics for tree certainty have been developed for identifying the set of single-copy markers that perform best for predicting phylogeny (53). Collectively, a reasonable approach to reconciling the distinction of different tree topologies would involve choosing a reliable marker set. Inclusion of conserved core genes with few or no signal of duplications and horizontal gene transfers either within or among genomes/lineages would be a possible solution. It is also suggested to use a set of custom ranked markers that are suitable for the depth of phylogeny that is intended to be solved (54). Meanwhile, improved/balanced sampling of taxa for better representation of tree topology and lineage placement is required (53).

Asgard superphylum has been discovered and characterized as the lineages bridging between Eukarya and Archaea (43-50). Ever since the discovery of the first Asgard archaea lineage of Lokiarchaeota in 2015 (43), the knowledge of the closest relative of eukaryotes continues to push forward. In 2017, Zaremba-Niedzwiedzka et al. expanded Asgard superphylum with additional Thor-, Odin-, and Heimdallarchaeota and further suggested that Asgard archaea were affiliated with eukaryotes in phylogenomic analyses (44). Later on, many Asgard lineages were discovered in the recent five years from groups around the world (47, 48, 50, 55). In 2021, Wukongarchaeota – the Asgard archaea named after Sun Wukong (The Monkey King), a figure in Chinese mythology, was discovered and believed to be currently the closest relative (45, 48, 54). Based on different ways to reconstruct the phylogeny of eukaryotes and archaea, there is still some support for placing eukaryotes basal within archaea other than within the “expanded Heimdallarchaeota–Wukongarchaeota” branch (45). Discovery of more close relatives of eukaryotes from diverse environments continues to take place today. The search for the “next relative” is not only an attempt to solve or refine the phylogeny puzzle but also provides genomic context that is valuable to explore eukaryotic features of archaea in the context of the landmark of eukaryogenesis, deduce the metabolism of Asgard archaea ancestors, and the syntrophic relationship between protoeukaryotes and hosts (45).

The function of archaeal ESPs and eukaryogenesis

In 2008, it was reported that the cell division machinery (Cdv) of Crenarchaeota has sequence homology to core domains with eukaryotic endosomal sorting complexes required for transport (ESCRT) machinery, which suggested common evolutionary origins between Archaea and Eukarya (56). Later, more reports on ESPs and evidence of phylogenetic reconstruction indicated the close relationship between Archaea and Eukarya. When Asgard archaea were first uncovered with four phyla (Loki-, Thor-, Odin-, and Heimdallarchaeota), their genomes encoded the highest number of ESPs among prokaryotes (43, 44) (Figure 2). The functions of these ESPs mainly discovered in Asgard archaea can be categorized into trafficking machinery, cytoskeleton, ubiquitin system, etc. (43, 44, 57).

In 2021, more phylum-level lineages of Asgard archaea were proposed including *Ca.* Wukongarchaeota (45). Meanwhile, the increased number of Asgard genomes has resulted in a major expansion of the set of ESPs (45); more than half of them have not been described previously (44). The ESCRT machinery as components of the ubiquitin system has one of the most conserved genomic neighborhoods in the genomes of Asgard archaea, implying signatures of vertical inheritance of ESCRT machinery from the ancestor (45). The physiological functions of the ESCRT genes in Asgard archaea have been studied (58-61). Collectively, this posits that these cell structure and trafficking machinery involving ESPs are functionally active in Asgard archaea, resembling similar activities in eukaryotic cells. However, phyletic patterns of ESPs in Asgard archaea are extremely patchy and each largely follows a lineage-specific pattern, indicating that some ESPs might be acquired through horizontal gene transfer from a protoeukaryote to Asgard archaea (45).

On the other hand, understanding the functions and evolutionary mechanisms of these ESPs in Asgard archaea can further support the bridging role of Asgard archaea in prokaryotes and eukaryotes. By comparing the enzymatic activity and selectivity from different life domains, scientists can study the ESPs in the context of evolution, such as a recent study on adenylate kinase (67). Furthermore, using heterogeneous *in vivo* expression, recent studies have also demonstrated the functions of Asgard ESCRT-III and Vsp4, implying they have similar functional potential as their eukaryotic counterparts (58-61), yet the regulating mechanism of Asgard ESCRT and its role on eukaryogenesis need further investigations. Meanwhile, as Asgard archaea possess eukaryotic-like actin-regulating proteins but with much more primitive characters, many actin regulators might be annotated with unknown functions and have different protein architectures as eukaryotes (68). Since a sophisticated regulated actin cytoskeleton is the hallmark of eukaryotic cells, in the future, more efforts can be paid to determine and characterize other actin regulators and to link actin cytoskeleton to specific functions and structures in model Asgard archaea (68).

The roles of HGT, viruses, and mobile elements in eukaryogenesis

Based on inference from Asgard archaea genome collections, the deduced last Asgard archaea common ancestor (LACA) can use organic matter, i.e., fatty acids, hydrocarbons, and aromatics (69). Horizontal gene transfer (HGT) signals can be found in enzymes that are essential for these metabolisms beyond Asgard archaeal lineages and spanning across various groups in Bacteria and Archaea (69). Additionally, it is believed that HGT also contributes significantly to the symbiogenesis process between the archaeal progenitor and proto-mitochondria (69). Gene repertoire from free-living bacterial cells occupies a considerable size within the eukaryotic cells by HGTs during the evolution process (70). With more evidence from either cultured or uncultured Asgard archaea discovered in the future, the syntrophic patterns between archaea and its partner would be clearer. The next step could be to investigate explicit HGT processes by identifying the origin of HGT donors along with the evolutionary process. Possibly, Asgard archaea interacted with bacterial partners in either different partnerships or forms of energy exchange. Furthermore, indirect HGTs of bacterial genetic components to protoeukaryotes via Asgard archaea may be possible, resulting in a high functional volume and flexibility for protoeukaryotes to customize their functional capacity and adapt to the environment. The discovery of HGT content, origin, and timeline will help reconstruct the scenario of the partnership during the ancient eukaryogenesis process.

Three Asgard archaea virus families were discovered to be widely distributed in different marine sediments – Wyrdviruses, Verdandiviruses, and Skuldviruses (71). Verdandiviruses and Skuldviruses belong to Realm Duplodnaviria and Varidnaviria, respectively. Although members from both Duplodnaviria and Varidnaviria can also infect eukaryotes, and have been proposed present in LACA and the last universal cellular ancestor (71), current results indicate no direct evidence of these two Asgard archaea virus families infecting eukaryotes. Simultaneously, two other studies also reported Asgard archaea viruses (72, 73). They reported more eukaryotic virus features found in Asgard archaea viruses, i.e., ~1-5% genes associated with eukaryotic nucleocytoplasmic large DNA viruses and being able to hijack host ubiquitin systems. These newly discovered unique archaeal and eukaryotic virus hybridized features for Asgard archaea viruses consistently reflect the evolutionary position of their hosts. More research targeting viruses is encouraged to complement current findings due to little/no overlap of the currently discovered Asgard archaeal virome pools and their globally ubiquitous distribution. As Asgard archaea viruses are evolving along with the evolution of Asgard hosts, it remains unknown whether viruses have played promotive roles in the evolution of Asgard archaea or in eukaryogenesis. Furthermore, it is also intriguing to find whether there are any extant eukaryote viruses directly evolving from Asgard archaea viruses.

Viruses and mobile genetic elements (without viral feature proteins) constitute the mobilomes of Asgard archaea (71, 74). They can still reflect remnant HGT signals that descend from ancient Asgard archaea during eukaryogenesis to extant Asgard archaeal genomes (74). Expanded mobilomes in the future will help us better trace the evolutionary history of HGT events before or after eukaryogenesis. According to the conceptualized framework of “Heimdall nucleation–decentralized innovation–hierarchical import” (HDH) model for explaining eukaryogenesis (74), the step of forming fully-fledged protoeukaryotes involved domain-specific hierarchical HGTs indirectly through Asgard archaea lineages and/or other related lineages (such as TACK superphylum or other transitionary lineages between Asgard and TACK). The expanded genomic contents of Asgard and related archaea lineages will facilitate our understanding of the process of formation of eukaryotic complexity from its archaeal ancestor. Different Asgard archaeal lineages have considerably small overlaps on ESP pools, and their ESP pool volume seems to be not associated with their phylogenetic distance to eukaryotes. It is intriguing to examine why there is a high mobility of ESPs and what factors govern the distribution of ESPs among the Asgard archaeal lineages. Furthermore, since during evolutionary history, protoeukaryotes gained their functional complexity significantly through HGT and mobilomes from donors, it is unknown how the cells determine the fates of imported genes and orchestrate functional patches into essential, organized functions. Many aspects of this conceptualized framework require attention in the future. The newly discovered content and proposed hypotheses will help us better understand eukaryogenesis.

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (92251306, 32225003, 31970105, 92051102), the Innovation Team Project of Universities in Guangdong Province (No. 2020KCXTD023), and the Shenzhen Science and Technology Program (JCYJ20200109105010363). This work was partially funded by the US-National Science Foundation grants OCE-2049478 and DBI-2047598 to K.A.

Author Contributions

ML, ZZ, and YL designed the study; ZZ and YL wrote the paper with contributions from all authors; KA provided valuable suggestions. All authors approved the final version of the manuscript.

Competing Interests statement

The authors declare that they have no competing interests.

Figure caption

Figure 1. Progress in the "tree of life" as increasing archaeal lineages having been identified in recent decades. A) The unrooted Woese 3D tree first conceived in 1977 by Carl Woese group

(1) B) The unrooted Eocyte 2D tree drawn in 1988 by James Lake (4), C) The Korarchaeota "add-on" tree constructed in 1996 by Barnes, Pace and others (14), D) The TACK "add-on" tree constructed in 2011 by Guy and Ettema (41), E) The Tree of Life constructed in 2016 by Hug, Banfield and others (75), F) The Asgard "add-on" tree constructed in 2022 in this study. This tree includes all the 17 Asgard lineages to have been discovered since 2015. The shape of this tree is derived from Fig. 1 by Zhou et al., 2018 (40). All these trees are conceptualized trees based on the original topology. The branching order of Asgard lineages within the Asgard "add-on" tree was not represented for simplicity.

Figure 2. Eukaryotic signature proteins present in major archaeal groups. This figure indicates the emergence of homologues of eukaryotic signature proteins (ESPs) along the schematic tree of Archaea (referred to Eme et al., 2017 (57)). Note: the dash-lined illustration of nuclei and endoplasmic reticulum currently have not been proven to be present in the common ancestor of eukaryotes and Asgard archaea.

Figure S1. The archaeal reference phylogenomic unrooted tree. This tree contained 3412 representative species, was inferred from 53 conserved marker proteins and was decorated with GTDB taxonomy at the phylum level. The raw tree file was obtained from "https://data.ace.uq.edu.au/public/gtdb/data/releases/release207/207.0/ar53_r207.tree". The visualization of the tree is achieved using ChiPlot (https://www.chiplot.online/).

Supplementary Data 1. GTDB and NCBI taxonomy information (phylum-level) about 3412 archaeal representative species in GTDB r207. The missing NCBI taxonomy information of minor species in this table is due to that their entries have been suppressed by NCBI, but are still kept in GTDB database (07-RS207).

References:

1. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc Natl Acad Sci U S A. 1977;74(11):5088-90.

2. Kelly S, Wickstead B, Gull K. Archaeal phylogenomics provides evidence in support of a methanogenic origin of the Archaea and a thaumarchaeal origin for the eukaryotes. Proc R Soc Lond B Biol Sci. 2011;278(1708):1009-18.

3. Lake JA. Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. Nature. 1988;331(6152):184-6.

4. Lake JA, Henderson E, Oakes M, Clark MW. Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. Proc Natl Acad Sci U S A. 1984;81(12):3786-90.

5. Iwabe N, Kuma K-i, Kishino H, Hasegawa M, Miyata T. Evolution of RNA polymerases and branching patterns of the three major groups of archaebacteria. J Mol Evol. 1991;32(1):70-8.

6. Klenk H-P, Palm P, Zillig W. DNA-Dependent RNA Polymerases as Phylogenetic Marker Molecules. Syst Appl Microbiol. 1993;16(4):638-47.

7. Liao D, Dennis PP. Molecular phylogenies based on ribosomal protein L11, L1, L10, and L12 sequences. J Mol Evol. 1994;38(4):405-19.

8. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990;87(12):4576-9.

9. Pace NR. Time for a change. Nature. 2006;441(7091):289.

10. Madigan M, Bender K, Buckley D, Sattley W, Stahl D. Brock Biology of Microorganisms. 15th Global Edition. Boston: Pearson Education; 2018.

11. Rinke C, Rubino F, Messer LF, Youssef N, Parks DH, Chuvochina M, et al. A phylogenomic and ecological analysis of the globally abundant Marine Group II archaea (Ca. Poseidoniales ord. nov.). ISME J. 2019;13(3):663-75.

12. Carr SA, Jungbluth SP, Eloe-Fadrosh EA, Stepanauskas R, Woyke T, Rappé MS, et al. Carboxydotrophy potential of uncultivated Hydrothermarchaeota from the subseafloor crustal biosphere. ISME J. 2019; 13(6):1457-68.

13. Vetriani C, Jannasch HW, MacGregor BJ, Stahl DA, Reysenbach AL. Population structure and phylogenetic characterization of marine benthic archaea in deep-sea sediments. Appl Environ Microbiol. 1999;65(10):4375-84.

14. Barns SM, Delwiche CF, Palmer JD, Pace NR. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc Natl Acad Sci U S A. 1996;93(17):9188-93.

15. Elkins JG, Podar M, Graham DE, Makarova KS, Wolf Y, Randau L, et al. A korarchaeal genome reveals insights into the evolution of the Archaea. Proc Natl Acad Sci U S A. 2008;105(23):8102-7.

16. Konneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA. Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature. 2005;437(7058):543-6.

17. Nunoura T, Takaki Y, Kakuta J, Nishi S, Sugahara J, Kazama H, et al. Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. Nucleic Acids Res. 2011;39(8):3204-23.

18. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, et al. Insights into the phylogeny and coding potential of microbial dark matter. Nature. 2013;499(7459):431.

19. Dombrowski N, Lee JH, Williams TA, Offre P, Spang A. Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. FEMS Microbiol Lett. 2019;366(2).

20. He C, Keren R, Whittaker ML, Farag IF, Doudna JA, Cate JHD, et al. Genome-resolved metagenomics reveals site-specific diversity of episymbiotic CPR bacteria and DPANN archaea in groundwater ecosystems. Nat Microbiol. 2021;6(3):354-65.

21. Spang A, Caceres EF, Ettema TJG. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. Science. 2017;357(6351):eaaf3883.

22. Williams TA, Szöllősi GJ, Spang A, Foster PG, Heaps SE, Boussau B, et al. Integrative modeling of gene and genome evolution roots the archaeal tree of life. Proc Natl Acad Sci U S A. 2017;114(23):E4602-E11.

23. Aouad M, Taib N, Oudart A, Lecocq M, Gouy M, Brochier-Armanet C. Extreme halophilic archaea derive from two distinct methanogen Class II lineages. Mol Phylogen Evol. 2018;127:46-54.

24. Setubal JC. Metagenome-assembled genomes: concepts, analogies, and challenges. Biophysical Reviews. 2021;13(6):905-9.

25. Jay ZJ, Beam JP, Dlakić M, Rusch DB, Kozubal MA, Inskeep WP. Marsarchaeota are an aerobic archaeal lineage abundant in geothermal iron oxide microbial mats. Nat Microbiol. 2018;3(6):732-40.

26. De Anda V, Chen L-X, Dombrowski N, Hua Z-S, Jiang H-C, Banfield JF, et al. Brockarchaeota, a novel archaeal phylum with unique and versatile carbon cycling pathways. Nat Commun. 2021;12(1):2404.

27. Meng J, Xu J, Qin D, He Y, Xiao X, Wang F. Genetic and functional properties of uncultivated MCG archaea assessed by metagenome and gene expression analyses. ISME J. 2014;8(3):650-9.

28. Jungbluth SP, Amend JP, Rappé MS. Metagenome sequencing and 98 microbial genomes from Juan de Fuca Ridge flank subsurface fluids. Sci Data. 2017;4:170037.

29. Vanwonterghem I, Evans PN, Parks DH, Jensen PD, Woodcroft BJ, Hugenholtz P, et al. Methylotrophic methanogenesis discovered in the archaeal phylum *Verstraetearchaeota*. Nat Microbiol. 2016;1:16170.

30. Baker BJ, Saw JH, Lind AE, Lazar CS, Hinrichs K-U, Teske AP, et al. Genomic inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea. Nat Microbiol. 2016;1:16002.

31. Nobu MK, Narihiro T, Kuroda K, Mei R, Liu W-T. Chasing the elusive Euryarchaeota class WSA2: genomes reveal a uniquely fastidious methyl-reducing methanogen. ISME J. 2016;10(10):2478-87.

32. Lazar CS, Baker BJ, Seitz KW, Teske AP. Genomic reconstruction of multiple lineages of uncultured benthic archaea suggests distinct biogeochemical roles and ecological niches. ISME J. 2017;11(5):1118-29.

33. Narasingarao P, Podell S, Ugalde JA, Brochier-Armanet C, Emerson JB, Brocks JJ, et al. De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. ISME J. 2012;6(1):81-93.

34. Yin X, Zhou G, Cai M, Zhu Q-Z, Richter-Heitmann T, Aromokeye DA, et al. Catabolic protein degradation in marine sediments confined to distinct archaea. ISME J. 2022;16(6):1617-26.

35. Zhou Z, Liu Y, Lloyd KG, Pan J, Yang Y, Gu J-D, et al. Genomic and transcriptomic insights into the ecology and metabolism of benthic archaeal cosmopolitan, Thermoprofundales (MBG-D archaea). ISME J. 2018.

36. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res. 2022;50(D1):D785-D94.

37. Murray AE, Freudenstein J, Gribaldo S, Hatzenpichler R, Hugenholtz P, Kämpfer P, et al. Roadmap for naming uncultivated Archaea and Bacteria. Nat Microbiol. 2020;5(8):987-94.

38. Hedlund BP, Chuvochina M, Hugenholtz P, Konstantinidis KT, Murray AE, Palmer M, et al. SeqCode: a nomenclatural code for prokaryotes described from sequence data. Nat Microbiol. 2022;7(10):1702-8.

39. Koonin EV. Origin of eukaryotes from within archaea, archaeal eukaryome and bursts of gene gain: eukaryogenesis just made easier? Phil Trans R Soc B. 2015;370(1678):20140333.

40. Zhou Z, Liu Y, Li M, Gu J-D. Two or three domains: a new view of tree of life in the genomics era. Appl Microbiol Biotechnol. 2018;102(7):3049-58.

41. Guy L, Ettema TJ. The archaeal ‘TACK’ superphylum and the origin of eukaryotes. Trends Microbiol. 2011;19(12):580-7.

42. Williams TA, Foster PG, Nye TMW, Cox CJ, Embley TM. A congruent phylogenomic signal places eukaryotes within the Archaea. Proc R Soc Lond B Biol Sci. 2012;279(1749):4870-9.

43. Spang A, Saw JH, Jorgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature. 2015;521(7551):173-9.

44. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, Vancaester E, et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature. 2017;541(7637):353-8.

45. Liu Y, Makarova KS, Huang W-C, Wolf YI, Nikolskaya AN, Zhang X, et al. Expanded diversity of Asgard archaea and their relationships with eukaryotes. Nature. 2021;593(7860):553-7.

46. Liu Y, Zhou Z, Pan J, Baker BJ, Gu J-D, Li M. Comparative genomic inference suggests mixotrophic lifestyle for Thorarchaeota. ISME J. 2018;12(4):1021-31.

47. Cai M, Liu Y, Yin X, Zhou Z, Friedrich MW, Richter-Heitmann T, et al. Diverse Asgard archaea including the novel phylum Gerdarchaeota participate in organic matter degradation. Sci China Life Sci. 2020;63(6):886-97.

48. Xie R, Wang Y, Huang D, Hou J, Li L, Hu H, et al. Expanding Asgard members in the domain of Archaea sheds new light on the origin of eukaryotes. Sci China Life Sci. 2022;65(4):818-29.

49. Seitz KW, Lazar CS, Hinrichs K-U, Teske AP, Baker BJ. Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. ISME J. 2016;10(7):1696-705.

50. Sun J, Evans PN, Gagen EJ, Woodcroft BJ, Hedlund BP, Woyke T, et al. Recoding of stop codons expands the metabolic potential of two novel Asgardarchaeota lineages. ISME Commun. 2021;1(1):30.

51. Klinger CM, Spang A, Dacks JB, Ettema TJG. Tracing the Archaeal Origins of Eukaryotic Membrane-Trafficking System Building Blocks. Mol Biol Evol. 2016;33(6):1528-41.

52. Dombrowski N, Williams TA, Sun J, Woodcroft BJ, Lee J-H, Minh BQ, et al. Undinarchaeota illuminate DPANN phylogeny and the impact of gene transfer on archaeal evolution. Nat Commun. 2020;11(1):3939.

53. Martinez-Gutierrez CA, Aylward FO. Phylogenetic Signal, Congruence, and Uncertainty across Bacteria and Archaea. Mol Biol Evol. 2021;38(12):5514-27.

54. Liu Y, Li M. The unstable evolutionary position of Korarchaeota and its relationship with other TACK and Asgard archaea. mLife. 2022;1(2):218-22.

55. Seitz KW, Dombrowski N, Eme L, Spang A, Lombard J, Sieber JR, et al. Asgard archaea capable of anaerobic hydrocarbon cycling. Nat Commun. 2019;10(1):1822.

56. Lindås A-C, Karlsson EA, Lindgren MT, Ettema TJG, Bernander R. A unique cell division machinery in the Archaea. Proc Natl Acad Sci U S A. 2008;105(48):18942-6.

57. Eme L, Spang A, Lombard J, Stairs CW, Ettema TJ. Archaea and the origin of eukaryotes. Nat Rev Microbiol. 2017;15(12).

58. Lu Z, Fu T, Li T, Liu Y, Zhang S, Li J, et al. Coevolution of Eukaryote-like Vps4 and ESCRT-III Subunits in the Asgard Archaea. mBio. 2020;11(3):e00417-20.

59. Akıl C, Tran LT, Orhant-Prioux M, Baskaran Y, Manser E, Blanchoin L, et al. Insights into the evolution of regulated actin dynamics via characterization of primitive gelsolin/cofilin proteins from Asgard archaea. Proc Natl Acad Sci U S A. 2020;117(33):19904-13.

60. Higuchi A, Shihoya W, Konno M, Ikuta T, Kandori H, Inoue K, et al. Crystal structure of schizorhodopsin reveals mechanism of inward proton pumping. Proc Natl Acad Sci U S A. 2021;118(14):e2016328118.

61. Nachmias D, Melnikov N, Zorea A, Sharon M, Yemini R, De-picchoto Y, et al. Asgard ESCRT-III and VPS4 reveal conserved chromatin binding properties of the ESCRT machinery. ISME J. 2022. doi: 10.1038/s41396-022-01328-2.

62. Wang H-W, Wang J-W. How cryo-electron microscopy and X-ray crystallography complement each other. Protein Sci. 2017;26(1):32-9.

63. Hansson T, Oostenbrink C, van Gunsteren W. Molecular dynamics simulations. Curr Opin Struct Biol. 2002;12(2):190-6.

64. Velázquez-Campoy A, Ohtaka H, Nezami A, Muzammil S, Freire E. Isothermal Titration Calorimetry. Curr Protoc Cell Biol. 2004;23(1):17.8.1-.8.24.

65. Chan K-M, Delfert D, Junger KD. A direct colorimetric assay for Ca^{2+}-stimulated ATPase activity. Anal Biochem. 1986;157(2):375-80.

66. Kumar A, Balbach J. Real-time protein NMR spectroscopy and investigation of assisted protein folding. Biochim Biophys Acta. 2015;1850(10):1965-72.

67. Verma A, Åberg-Zingmark E, Sparrman T, Mushtaq AU, Rogne P, Grundström C, et al. Insights into the evolution of enzymatic specificity and catalysis: From Asgard archaea to human adenylate kinases. Science Advances. 2022;8(44):eabm4089.

68. Akıl C, Kitaoku Y, Tran LT, Liebl D, Choe H, Muengsaen D, et al. Mythical origins of the actin cytoskeleton. Curr Opin Cell Biol. 2021;68:55-63.

69. Spang A, Stairs CW, Dombrowski N, Eme L, Lombard J, Caceres EF, et al. Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. Nat Microbiol. 2019;4(7):1138-48.

70. Roger AJ, Muñoz-Gómez SA, Kamikawa R. The Origin and Diversification of Mitochondria. Curr Biol. 2017;27(21):R1177-R92.

71. Medvedeva S, Sun J, Yutin N, Koonin EV, Nunoura T, Rinke C, et al. Three families of Asgard archaeal viruses identified in metagenome-assembled genomes. Nat Microbiol. 2022;7(7):962-73.

72. Rambo IM, Langwig MV, Leão P, De Anda V, Baker BJ. Genomes of six viruses that infect Asgard archaea from deep-sea sediments. Nat Microbiol. 2022;7(7):953-61.

73. Tamarit D, Caceres EF, Krupovic M, Nijland R, Eme L, Robinson NP, et al. A closed Candidatus Odinarchaeum chromosome exposes Asgard archaeal viruses. Nat Microbiol. 2022;7(7):948-52.

74. Wu F, Speth DR, Philosof A, Crémière A, Narayanan A, Barco RA, et al. Unique mobile elements and scalable gene flow at the prokaryote–eukaryote boundary revealed by circularized Asgard archaea genomes. Nat Microbiol. 2022;7(2):200-12.

75. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new view of the tree of life. Nat Microbiol. 2016;1:16048.