



Review

Thermotogales origin scenario of eukaryogenesisTomohiko Kuwabara^{a,*}, Kensuke Igarashi^b^a Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba 305–8572, Ibaraki, Japan^b Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Sapporo 062–8517, Hokkaido, Japan

ARTICLE INFO

Article history:

Received 6 October 2019

Accepted 6 February 2020

Available online 7 February 2020

Keywords:

Cell fusion

Intra-outer-membrane multicellularity

Multicopy-integrated large genome

Non-homologous end-joining

Signal cleavage site

ABSTRACT

How eukaryotes were generated is an enigma of evolutionary biology. Widely accepted archaeal-origin eukaryogenesis scenarios, based on similarities of genes and related characteristics between archaea and eukaryotes, cannot explain several eukaryote-specific features of the last eukaryotic common ancestor, such as glycerol-3-phosphate-type membrane lipids, large cells and genomes, and endomembrane formation. *Thermotogales* spheroids, having multicopy-integrated large nucleoids and producing progeny in periplasm, may explain all of these features as well as endoplasmic reticulum-type signal cleavage sites, although they cannot divide. We hypothesize that the progeny chromosome is formed by random joining small DNAs in immature progeny, followed by reorganization by mechanisms including homologous recombination enabled with multicopy-integrated large genome (MILG). We propose that *Thermotogales* ancestor spheroids came to divide owing to the archaeal cell division genes horizontally transferred via virus-related particles, forming the first eukaryotic common ancestor (FECA). Referring to the hypothesis, the archaeal information-processing system would have been established in FECA by random joining DNAs excised from the MILG, which contained horizontally transferred archaeal and bacterial DNAs, followed by reorganization by the MILG-enabled homologous recombination. Thus, the large genome may have been a prerequisite, but not a consequence, of eukaryogenesis. The random joining of DNAs likely provided the basic mechanisms for eukaryotic evolution: producing the diversity by the formations of supergroups, novel genes, and introns that are involved in exon shuffling.

© 2020 The Authors. Published by Elsevier Ltd.

This is an open access article under the CC BY license. (<http://creativecommons.org/licenses/by/4.0/>)

1. Introduction

1.1. Eukaryogenesis studies from the prokaryotic side

Eukaryogenesis is considered a genomic innovation converting prokaryotes to eukaryotes, introducing organizational complexity, including endomembranes, cytoskeletons, and mitochondria, in addition to large cells and genomes to manage them (Koonin, 2010; Lane and Martin, 2010). The large genome appears the essential feature of eukaryotes, since prokaryotic genomes are necessarily small, while the other features can be seen in rudimentary level in prokaryotes (Lane and Martin, 2010). Aside from that, it is widely accepted that eukaryotes are chimeras of archaea and bacteria; eukaryotes contain informational genes similar to those of archaea and operational genes similar to those of bacteria (Rivera et al., 1998). Various types of genome fusion between archaea and bacteria and hypothetical scenarios have been proposed for eukaryogenesis, as reviewed previously (Embley and

Martin, 2006; Martin et al., 2015; Poole and Penny, 2006). The proposed archaeal species were changed from eocytes (hyperthermophilic Crenarchaea) (Rivera and Lake, 2004) through TACK superphylum (Guy and Ettema, 2011) to Asgard superphylum (Zaremba-Niedzwiedzka et al., 2017), including Lokiarchaeota (Spang et al., 2015), along with the spread of sampling areas, which eventually suggested that eukaryote-specific proteins could have originated from the deep sea (Spang et al., 2015; Zaremba-Niedzwiedzka et al., 2017). The specific bacterial species involved was proposed to be cyanobacteria or proteobacteria (Rivera and Lake, 2004), but this is not definite as these speculations are not supported by further studies (Gould et al., 2016; Guy and Ettema, 2011; Spang et al., 2015; Zaremba-Niedzwiedzka et al., 2017). Recently, Gould et al. (2016) proposed the hypothesis that the bacterial genes had been horizontally transferred to the α -proteobacterial proto-symbiont, and after endosymbiosis, the outer membrane (OM) vesicles of mitochondria transferred the genes to the host nucleus, accompanied by the endomembrane formation by the vesicles. This paper was a breakthrough for the understanding of the possible mechanism underlying endomembrane formation, but it appears to contain inconsistencies regarding the structural aspects of the cell. OM vesicle-producing bacte-

* Corresponding author.

E-mail address: kuwabara@biol.tsukuba.ac.jp (T. Kuwabara).

ria should have peptidoglycan (PG); no eukaryotes have been reported to have mitochondria with PG, although some chloroplasts contain PG (Aitken and Stanier, 1979; Pfanzagl et al., 1996), even with genes coding for PG formation obtained from bacteria other than cyanobacteria (Sato and Takano, 2017), the taxon believed to be the origin of chloroplasts. Moreover, lipopolysaccharide (LPS) (Schwechheimer and Kuehn, 2015) in the OM should have been immediately removed upon formation of endomembranes, unless the proto-symbiont is LPS-less α -proteobacteria, such as *Sphingomonas* (Kawasaki et al., 1994), which have never been considered as the origin of mitochondria. The lack of LPS is not discussed in the OM vesicle-derived endomembrane hypothesis (Gould et al., 2016). Thus, this hypothesis could be evaluated for the vesicle-fusion mechanism of endomembrane formation but not for the source of vesicles.

1.2. Eukaryogenesis studies from the eukaryotic side

According to the classification by Adl et al. (2012), eukaryotes can be divided into five supergroups: Stramenopiles, Alveolata, and Rhizaria (SAR); Archaeplastida; Excavata; Amoebozoa; and Opisthokonta. The mechanism of the diversification, which produced the supergroups from the last eukaryotic common ancestor (LECA), is unknown. Parasitic amitochondriate eukaryotes were previously thought to be closely related to the LECA (Cavalier-Smith, 1989; Pace, 1997). However, the phylogenetic proximity was found to be caused by a long branch attraction (Felsenstein, 1978) artifact with the discovery that their organelles, hydrogenosomes and mitosomes, are the retrogressed forms of mitochondria (Embley and Martin, 2006). A free-living excavate, *Naegleria gruberi*, was found to have a large genome (41 Mb), including 4133 genes that are shared with at least one other eukaryote supergroup, in addition to novel genes accounting for nearly 40% of the total protein-coding genes (Fritz-Laylin et al., 2010). *N. gruberi*, compared to other eukaryotes, appears evolutionarily closer to LECA; the supergroups may have evolved from the LECA by reducing the gene number (Fritz-Laylin et al., 2010). Protein-coding genes of *Saccharomyces cerevisiae* (Opisthokonta) with evolutionary affinities to the Crenarchaeota-Thaumarchaeota-Nanoarchaeota group and α -proteobacteria are only 3% and 4%, respectively, of the total, while those of *Cyanidioschyzon merolae* (Archaeplastida) are 3% and 6%, respectively, of the total, with novel genes accounting for more than 30% of the total in both microorganisms (Koonin, 2010). Thus, the possibility cannot be excluded that the LECA obtained more than 50% of its protein-coding genes from other bacteria or archaea. Pittis and Gabaldón (2016) reported that some bacterial proteins were obtained before acquisition of mitochondria, supporting the late mitochondria-acquisition (Koumandou et al., 2013), though this topic is still debated in the field (Degli Esposti, 2016; Martin et al., 2017). Such acquisition of bacterial proteins is possible by direct horizontal gene transfer (HGT) to the host or by more than one endosymbiotic event (Gabaldón, 2018). In addition to the transfer of bacterial genes, the production of novel genes appears an important feature of eukaryogenesis. However, how they were produced remains unanswered.

1.3. Are experimental studies on eukaryogenesis possible?

We thought experimental research on eukaryogenesis would be possible, if appropriate model microorganisms are present, by introducing and/or removing some genes, similar to induced pluripotent stem cells established by introducing four factors, Oct3/4, Sox2, c-Myc, and Klf4, into adult mouse fibroblasts (Takahashi and Yamanaka, 2006). However, such model microorganisms have not yet been obtained. Eukaryogenesis is likely to be a process

in which an anaerobic host prokaryote captures an aerobic α -proteobacterial symbiont, possibly under partially aerobic conditions. Environments around hydrothermal vents are one of the naturally occurring partially aerobic conditions, which is dynamically maintained by the mixing of anaerobic hydrothermal fluids and aerobic seawater (Jannasch and Mottl, 1985). We isolated a possible model prokaryote, *Thermosipho globiformans* (JCM 15059^T/DSM 19918^T/NBRC 109617^T) (Kuwabara et al., 2011), whose spheroids can produce progeny rods in periplasm (Kuwabara and Igarashi, 2017), from a hydrothermal vent of Suiyo Seamount (Japan), using an *in-situ* cultivation device that creates an interface of anaerobic and aerobic conditions (Kuwabara et al., 2006).

The genus *Thermosipho* belongs to the order *Thermotogales* (Huber and Stetter, 1992). *Thermotogales* represents (hyper)thermophilic and anaerobic heterotrophs having an enigmatic morphology, large periplasm at the ends of rods and the surrounding thick surface structure, toga, which consists of OM and amorphous layer (AL).¹ *Thermotogales* bacteria are microscopically distinguishable from the other prokaryotes by the large periplasm. *Thermotogales* is known to have been highly prone to HGT (Nelson et al., 1999; Nesbø et al., 2009; Zhaxybayeva et al., 2009), as they are suggested to have obtained as many as 80% of their protein-coding genes by HGT (Zhaxybayeva et al., 2009). Consistent with this are many clustered regularly interspaced short palindromic repeat (CRISPR) sequences (Horvath and Barrangou, 2010) in their genomes (Nesbø et al., 2009; Zhaxybayeva et al., 2009), indicative of previous infection by viruses and thereby HGT (Brodt et al., 2011). In this article, we hypothesize the mechanism of the chromosome formation in progeny. Using this hypothesis, we explain the establishment of archaeal information-processing system in *Thermotogales* ancestor spheroids as well as the generations of supergroups, novel genes, and introns, in eukaryogenesis.

2. Characterizations of *Thermotogales* spheroids as a proto-eukaryote

2.1. Intra-outer-membrane multicellularity of *Thermotogales*

Enrichment cultures for the model prokaryotes were performed with heterotrophic media at 55 °C. *T. globiformans*, which naturally forms spheroids containing multiple cells within an OM in the early growth phases, was selected as the model prokaryote (Kuwabara et al., 2011). This multicellular state is expressed in this article as intra-outer-membrane multicellularity (IOM), emphasizing the difference from ordinary multicellularity regarded as a

¹ **Cell structures of *Thermotogales*.** In *Thermotogales*, the cell is defined as the cytoplasmic membrane (CM) plus the entity surrounded by the CM, while the rod is defined as the total of the rod-shaped cell, periplasm, and the *Thermotogales*-specific surface structure, toga (Huber and Stetter, 1992), which is composed of an outer membrane (OM) and an amorphous layer (AL). The discrimination of cell and rod is necessary because cell division and rod fission do not occur simultaneously (Kuwabara and Igarashi, 2012), different from other bacteria (Errington, 2003; Margolin, 2005).

The AL is an approximately 40-nm-thick structure that lines the OM, which contains or is equivalent to peptidoglycan (PG) (Kuwabara et al., 2011), as described below. *Thermotogales* PG has not been definitively identified by TEM, but it has been extensively characterized biochemically (Boniface et al., 2006; Boniface et al., 2009). PG of *Thermotoga maritima* has been shown to have two different peptides (1:1 in the molar ratio). The third amino acid residues from the glycan chain, central to the three-dimensional (3D) structure of PG (Typas et al., 2012), are L-lysine and D-lysine. As a result of the heterogeneity, PG is expected to have a less compact 3D structure, which would be hard to form and easy to degrade, despite functioning in total as a cell wall. In TEM images, the thick AL is the only recognizable structure between the OM and the CM. Lysozyme treatment of rods eliminates the AL, causing the transformation of rods to spheroids (Kuwabara et al., 2011), a process that suggests that AL contains or is equivalent to PG. Although an OM protein, Omp α , is shown to be associated with AL (Engel et al., 1992), proteins are not affected by lysozyme, indicating that at least the main constituent of AL is PG.

eukaryote-specific feature. We considered how the IOM was involved in eukaryogenesis. We suggested that even if the genomic innovation to eukaryotes failed, the *Thermotogales* lineage would have been preserved, and speculated that the IOM was related to the genomic innovation and/or the lineage preservation.

2.2. Large cells and genomes, and an endomembrane of *Thermotogales* spheroids

Cell structure of *Thermotogales* is different from those of other bacteria.¹ Owing to these differences, *Thermotogales* can form rods with IOM (Kuwabara and Igarashi, 2012). Briefly, in *Thermotogales*, the septal PG is formed in the periplasm between previously-divided cells, unlike in other bacteria, in which the septal PG (plus OM in the case of gram-negative bacteria) forms concomitantly with the constriction of the CM (Errington, 2003; Margolin, 2005; Wissel et al., 2005). Through rounds of cell division without septation, rods with IOM are formed. The transformation of rods to spheroids by the alteration or degradation of surrounding PG is common in bacteria (Lam et al., 2009; Leaver et al., 2009; Young, 2008). The resulting spheroids lack the ability to accumulate PG. The transformation of *T. globiformans* occurs in the early growth phases (Kuwabara and Igarashi, 2017), in addition to the stationary phase as in other bacteria (Huber et al., 1989; Huber et al., 1990; Huber et al., 1986; Lam et al., 2009; Leaver et al., 2009; Podosokorskaya et al., 2011; Podosokorskaya et al., 2014; Young, 2008). Owing to the early transformation, spheroids with IOM, which are produced by the transformation from rods with IOM, enlarge in the exponential phase, accompanied by the growth of multiple cells. These cells fuse with each other and can form a single dish-shaped cell that continues to grow (Kuwabara and Igarashi, 2017).² The resulting large dish-shaped cells likely fulfill the size requirements of LECA (Koonin, 2010; Lane and Martin, 2010). Although the nucleoids of dish-shaped cells should only contain components of the genome of *T. globiformans*, we posit that their copies likely integrate into a larger entity. The nucleoid of large dish-shaped cells appears distinct from that of rods as it pertains to both size² and the transformation during growth; the nucleoid transitions from the dish shape to a ring shape with a hole in the center, finally to a lip shape with the vacancy pulled into opposite directions (Kuwabara and Igarashi, 2017), showing an appearance resembling dividing eukaryotic chromosome (Kuwabara et al., 2011). These macroscopic differences suggest that the large dish-shaped cells have multicopy-integrated large genome (MILG), not identical to the genome of rods, and that the IOM is likely necessary to create the MILG through cell fusion.

Spheroids having a large dish-shaped cell can have vesicles and/or rods in periplasm, as shown by transmission electron microscopy (TEM) images (Fig. 1). They are regarded as immature and mature progeny, respectively, based on comparison with optical microscopically observed various-sized moving progeny (MP), as

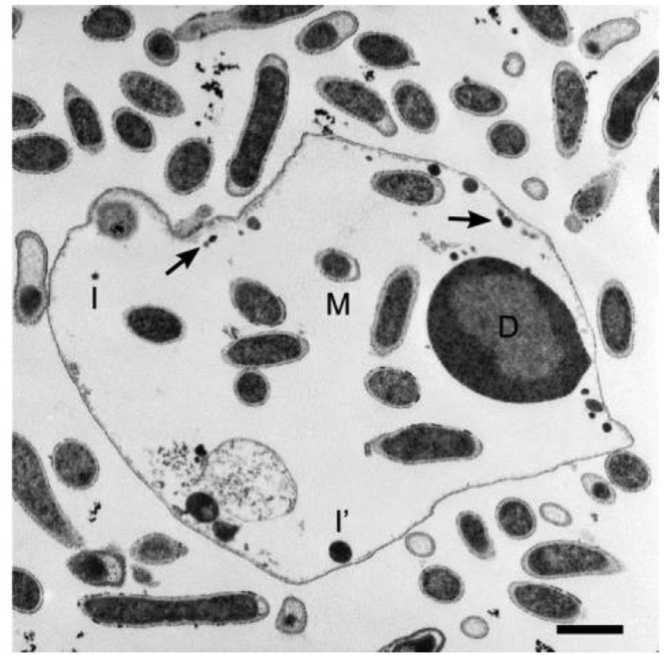


Fig. 1. Possible fusion of immature progeny for the formation of mature progeny. Source: The TEM image of spheroid containing various sizes of vesicles and rods with a large dish-shaped cell was reproduced from Kuwabara and Igarashi (2017). D, dish-shaped cell; i, small vesicle (immature progeny); i', large vesicle (immature progeny); and M, rod (mature progeny). Arrows indicate possible fusion of small vesicles. Bar: 1 μ m. A movie of epifluorescence microscopy results showing a spheroid having a rod-sized MP in periplasm is available as Online Resource 6 from Kuwabara and Igarashi (2017).

identified with DNA and membrane.³ The immature progeny likely fuse with each other (arrow) forming larger immature progeny (i'),

³ **Progeny production in spheroid periplasm.** Rapidly moving objects, which have various sizes depending on the cultivation time, can be produced in periplasm of spheroids having large dish-shaped cells (Kuwabara and Igarashi, 2017; Kuwabara et al., 2011). These objects are stainable with Live/Dead and FM1-43, indicating the presence of nucleic acids and membrane, respectively (Kuwabara et al., 2011). When fixed, they can be stained with 4',6-diamidino-2-phenylindole (DAPI), indicating the presence of DNA (Kuwabara and Igarashi, 2017). Given this, they are called moving progeny (MP), which should have a cytoplasmic membrane; they were called simply 'progeny' in a previous paper (Kuwabara and Igarashi, 2017). The size variety of MP suggests that MP grow in spheroid periplasm, although live-observations of the MP growth have been unsuccessful, because spheroids having small MP tend to sediment from just beneath the coverslip of observation chamber upon HTM (Kuwabara and Igarashi, 2017). In transmission electron microscopy (TEM), vesicles smaller than rods, as well as rods, which have togas, can be observed in periplasm of spheroids having a large dish-shaped cell (Fig. 1). Given their subcellular localization and size, the vesicles and rods are likely equivalent to optical-microscopically observed small and large MP and, thus, are called immature and mature progeny, respectively.

It is uncertain whether the movement of MP is derived from cell motility or Brownian movement. It should be noted that *Thermosiphon* genomes (Nesbø et al., 2009; Zhaxybayeva et al., 2009), including that of *T. globiformans* (GCA_003990895.1), contain flagella-related genes. Nevertheless, *Thermosiphon* is generally considered to be immotile, though with a few exceptional species (L'Haridon et al., 2001; Podosokorskaya et al., 2011; Podosokorskaya et al., 2014). The conditions under which the *Thermosiphon* flagellar genes are expressed remain to be clarified.

Other *Thermotogales* species have not been reported to produce MP. Nevertheless, *Thermotoga maritima* (Huber et al., 1986) does produce MP when the media of a stationary phase culture is exchanged via precipitation and resuspension, followed by cultivation (Kuwabara and Igarashi, 2017). Thus, the apparently special ability of *T. globiformans* to form in-periplasm progeny in the exponential phase likely resides in the high efficiency of spheroid formation in the early growth phases (Kuwabara and Igarashi, 2017). One of the possible reasons of the early spheroid formation could be related to its natural habitat, Suiyo Seamount, which is a caldera where fast bottom current flows twice a day (Kuwabara et al., 2006), likely bringing oxic seawater to the anoxic hydrothermal vents, and thus both anoxic and oxic environments alternate twice in a day. *Thermotogales* species isolated from other habitats are not reported to form spheroids in the early growth phases.

² **Large dish-shaped cells with a large nucleoid.** Spheroid growth can be studied by high-temperature microscopy (HTM), which was developed to live-observe the growth of anaerobic thermophiles (Kuwabara and Igarashi, 2012). *T. globiformans* spheroids with IOM having a diameter as small as 2 μ m in the early growth phases grow up to 12 μ m in diameter in the late exponential phase (Kuwabara and Igarashi, 2017). Along with the growth of spheroids, multiple cells fuse with each other to form a dish-shaped cell with a diameter of ~3 μ m, which further grows to ~12 μ m. Based on growth, the cell volume increased about 7500-fold in 367 min (Kuwabara and Igarashi, 2017). This increase is not surprising, considering the ability of *T. globiformans* rods to grow rapidly with the doubling time of 24 min (Kuwabara et al., 2011) and the inability of dish-shaped cells to divide (Kuwabara and Igarashi, 2017). Accompanying the cell fusion, the nucleoids from multiple cells fuse with each other and further enlarge to form a large nucleoid, occupying 1/7 to 1/6 area of spheroids (diameter, 7–9 μ m), as seen in epifluorescence microscopy images (Kuwabara and Igarashi, 2017; Kuwabara et al., 2011), which show very strong green epifluorescence of Live/Dead, indicating the viability of the cells.

which is not surprising since multiple cells translocate from rods into spheroids and fuse with each other to form dish-shaped cells (Kuwabara and Igarashi, 2017). The fusion of immature progeny is noteworthy given that the OM vesicles (Gould et al., 2016) are hypothesized to fuse with each other and form eukaryotic endomembranes. The immature progeny appears to be a better source of endomembranes than the OM vesicles, since the immature progeny membrane is likely derived from the CM³ and thus, may not be accompanied by either PG or LPS.

If the eukaryotic endomembranes are related to the *Thermotogales* CM, the similar N-terminal signal cleavage systems should be found between them. However, it is well known that the endoplasmic reticulum-type eukaryotic signal cleavage sites (Nielsen, 2017) are distinct from those of prokaryotes (Bagos et al., 2009; Nielsen, 2017) and have never been found in prokaryotic proteins. Surprisingly, the eukaryotic signal cleavage sites are found in *Thermotogales* periplasmic proteins if they are translated from non-ATG translation-initiable codons (Fig. 2). Briefly, FtsI, the D,D-transpeptidase catalyzing the final step of PG synthesis (Typas et al., 2012) is considered to be soluble in *Thermotogales*, since septal PG is formed far from the CMs of previously divided cells (Kuwabara and Igarashi, 2012). In fact, most *Thermosiphon* FtsIs are predicted by SignalP (Nielsen, 2017) to have eukaryotic signal cleavage sites. SignalP is considered the best predictor of N-terminal signal cleavage sites in bacteria and eukaryotes (Choo et al., 2009). Moreover, another tool, Signal-3L (Shen and Chou, 2007), produced similar but ambiguous results. In contrast to the *Thermosiphon* FtsIs, no signal cleavage sites were predicted in most FtsIs of other *Thermotogales* species. Since the cleavage site-predicted *Thermosiphon* FtsIs are all reported to be translated from non-ATG translation-initiable codons, the other *Thermotogales* FtsIs were also tested to determine if they were translated from non-ATG translation-initiable codons, as detailed in the legend of Fig. 2. Eukaryotic signal cleavage sites are now predicted in all *Thermosiphon* and *Thermotoga* FtsIs and in most *Thermotogales* FtsIs (Supplementary Table 1). Eukaryotic signal cleavage sites are also predicted in other proteins of *Thermotoga maritima* including xylanase (Liebl et al., 2008), glucose-binding protein (Nanavati et al., 2002; Palani et al., 2012), maltose-binding proteins (Nanavati et al., 2002), and arginine-binding protein (Ausili et al., 2013), for which the periplasmic localization has been biochemically suggested. This finding suggests that the endoplasmic reticulum membrane would have been derived from immature progeny, rather than from the OM vesicles, in which a comprehensive N-terminal signal cleavage system has not been reported. Unfortunately, no signal peptidases with appreciable homology between *Thermotogales* and eukaryotes have yet been found. This is likely because substrate specificity may not be evident based on the primary sequences of signal peptidases.

None of the *Thermotogales* FtsI sequences show bacterial signal cleavage sites (Supplementary Table 1), consistent with the aforementioned involvement of FtsI in the *Thermotogales*-specific morphogenesis, which suggests that FtsI is a *Thermotogales*-original protein. Further, no bacterial signal cleavage sites have been found in the arginine-binding protein. In contrast, some signal cleavage sites of the sugar-related proteins are predicted to be both eukaryotic and bacterial (Supplementary Table 2), suggesting that these proteins may have been obtained by HGT via signal cleavage sites.

It is currently unknown how the translation from non-ATG translation-initiable codons relates to eukaryogenesis. However, this is the first example of eukaryotic signal cleavage sites in prokaryotic proteins, which likely is significant given the importance of these sites in other prokaryotic proteins. Because of the large cell and genome sizes, the endomembrane formation, and the eukaryotic signal cleavage sites, *Thermotogales* ancestor spheroids,

which have the G3P-type membrane lipids, could be regarded as a proto-eukaryote (Poole and Penny, 2006) possessing some eukaryote-specific features, even if archaeal informational genes were absent.

2.3. Hypothesis of fragment-joining chromosome formation in progeny

The large dish-shaped cells can produce MP, which are detectable with optical microscopy, in spheroid periplasm.³ The size of the MP is negatively correlated to their number (Kuwabara and Igarashi, 2017; Kuwabara et al., 2011). This suggests that it is unlikely that an immature progeny matures by itself, which should yield a similar number of MP, irrespective of the size. In TEM images (Kuwabara and Igarashi, 2017), immature progeny can be as small as 80 nm in diameter (Fig. 1, label I). The small immature progeny may not have the whole genome of *T. globiformans* (1.9 Mb, GCA_003990895.1) (Igarashi, 2019), when compared with *Nanoarchaeum equitans* (Huber et al., 2002), which is one of the smallest prokaryotes with a cell size of 400 nm in diameter and a genome size of 0.49 Mb (Waters et al., 2003). Briefly, the amount of chromosomal DNA in a prokaryotic cell is represented by genome size multiplied by copy number (Taniguchi et al., 2010). Thus, even if the copy number of *N. equitans* were 55, the largest archaeal copy number of a euryarchaeon *Methanococcus maripaludis* (Hildenbrand et al., 2011; Soppa, 2011), the amount of DNA in the small immature progeny would be 0.22 Mb, approximately 11% of the genome, assuming the same concentration of chromosomal DNA. This amount is likely insufficient for the growth. Taken together with the apparent inconsistency between the numbers of immature and mature progeny, it is possible that immature progeny fuse with each other (arrow) to eventually form mature progeny with the former DNA fragments ligated to form the chromosome of the latter. Alternatively, mature progeny might be produced directly by the large dish-shaped cells, but if so, this should be accompanied by the removal of immature progeny. This removal is unlikely, considering that immature progeny has both DNA and a membrane.³ If this were the case, for what purpose are immature progeny produced? Furthermore, it is reasonable to think that the dish-shaped cells cannot produce PG, since they are generated through the results of the alteration or degradation of surrounding PG (Lam et al., 2009; Leaver et al., 2009; Young, 2008). If the dish-shaped cells can produce PG, why is PG absent from the spheroid OM (Fig. 1) (Kuwabara and Igarashi, 2017; Kuwabara et al., 2011)? Thus, the PG of mature progeny may not be formed by the large dish-shaped cells, but is likely formed under a different genetic control in spheroids, namely in large immature progeny (label I') when the PG-forming genes or the chromosome is completed. The presumed fragment joining appears to be consistent with the findings of Nesbø et al. (2006) who showed that some *Thermotoga* genes contain patches of four or more consecutive nucleotides termed "potentially recombinant fragments." Thus, we believe that the fragment joining theory is worth serious consideration, although such a mechanism of chromosome formation has not yet been identified.

If the DNA fragments in immature progeny have appropriate cohesive ends, the correct ligation of fragments reproducing the rod chromosome might be possible. However, such a cohesive end-generating mechanism is unknown. Thus, further research is required to understand the mechanism that ligates non-homologous ends. In *Mycobacterium* and eukaryotes, the non-homologous end-joining (NHEJ) system is reported, which repairs dsDNA-breaks with either blunt or cohesive ends (Brissett and Doherty, 2009; Glickman, 2014; Kegel et al., 2006). The NHEJ system employs Ku protein, which binds to the ends, and LigD or a homolog, which is

FtsI

Thermosipho globiformans (BAK08917) -----MYFIFVILIFVEIVKLFVFNKIE-
Thermosipho africanus (WP_012579298) -----MYFIFVILIFIEIVKLFVFNKIE-
Thermosipho melanesiensis (WP_041426274) -----MYSLLLLLLIETINILKINSSE-
Thermosipho sp.1063 (WP_075666653) -----MYIVLTVIIFIEMLSLFKINNTNSEEVLELS-
Thermosipho sp.1070 (WP_075666653) -----MKNSRVIFIYIVLTVIIFIEMLSLFKINNTNSEEVLELS-
Thermotoga maritima (NP_228679) -----MEKRLWLLLAIFAIFLTVSFNNLFFFPLAGERENWFVS-
Thermotoga petrophila (WP_011942760) -----MEKRLWLLLAIFALFLMVSLVNLFFFPLTERTNWFISI-
Thermotoga sp. RQ2 (WP_012310292) -----MEKRLWLLLAIFALFLMVSLVNLFFFPLTERTNWFISI-
Thermotoga naphthophila (WP_012895790) -----MEKRLWLLLAIFALFLMVSLVNLFFFPLTERTNWFISI-
Thermotoga sp. 2812B (WP_004080731) -----MEKRLWLLLAIFAIFLTVSFNNLFFFPLAGERENWFVS-
Thermotoga sp. RQ7 (AJG39912) -----MEKRLTLLLAVFALFLAVSFNINLFFFPPAGQKSHWYVS-
Thermotoga neapolitana (WP_038066249) -----MKNRLILLFMALSFVLI[▲]IKAFQVQILE-
Thermotoga sp. Cell2 (WP_038033741) -----MKNRLILILMALSFVLI[▲]IKAFQVQILE-
Pseudothermotoga lettingae (WP_012003466) -----MKLIFYVAAIFSAVMVVSFL[▲]LKGNAAYS[▲]PKHW-
Pseudothermotoga hypogea (AJC74350) -----MGNKLVSNRLKFLIFTFLFFSVGSFVLLSLRSSHFR-
Fervidobacterium nodosum (WP_011994909) -----MNKLSNKRYKIAFISFVLILSILIFKIWNNIYTNQYAINS-
Fervidobacterium pennivorans (AFG35501) -----MRRYIPFFIFFAMFFYVIFGLVKVQIIDRPK-
Fervidobacterium islandicum (WP_033191578) -----MSRSSHRYRLVLIITISLFAILLIKVYTN[▲]NIYTNQY-

***Thermotoga maritima* MSB8 proteins**

xylanase (AKE27878) -----MKILPSVLLLLGCVPFVSSQNVSLR-
 glucose-binding protein (NP_227930) -----MWNKTSLQGGVIMRKLLVFLSVLLVAGLSLALTIGVIG-
 maltose-binding protein (AKE28980) -----MKKFLVIALLVVSLVLAQPKLTIW-
 maltose-binding protein (AKE29588) -----MKRLLVMLVVVSALVLSQTKLTIW-
 arginine-binding protein (NP_228403) -----MKKLLAVSILLVSIVIFSGAIDEIKSR-
[▲]

Fig. 2. Eukaryotic signal cleavage sites predicted in *Thermotogales* periplasmic proteins.

Source: Gene and protein sequences were retrieved from National Center for Biotechnology Information (NCBI), as shown in Supplementary Tables 1 and 2. Protein sequences are aligned referring to the eukaryotic signal cleavage sites, which were predicted as follows and are indicated by triangles. The N-terminal hydrophobic stretches (von Heijne, 1983) formed by non-charged amino acid residues are highlighted. In the region from the N-terminus to the end of the hydrophobic stretch, amino acid residues encoded by translation-initiable codons other than ATG, such as GTG (for valine, V), ATA (isoleucine, I), and TTG (leucine, L), are underlined. The original sequence as well as the truncated sequences starting from the underlined positions (replaced by methionine) or internal methionine residues were analyzed for signal cleavage sites of eukaryotes and gram-negative and gram-positive bacteria using SignalP 4.1 (Nielsen, 2017) (<http://www.cbs.dtu.dk/services/SignalP/>) with default settings. For predicted eukaryotic signal cleavage sites in the sequence, the letter at the N-terminal position is written in red. Irrespective of the N-terminus, the same signal cleavage site was predicted in each FtsI, with one exception (Supplementary Table 1). Notably, FtsIs of *Pseudothermotoga thermarum* (AEH50221) and *Fervidobacterium penivorans* (WP_014452532) belonging to *Thermotogales*, in addition to *Petrotogales* and *Kosmotogales* bacteria, except for *Marinitoga piezophile* (AEX86383), were predicted to have no signal cleavage sites. This indicates that the order *Thermotogales*, especially *Thermosipho* and *Thermotoga* genera, is more closely related to eukaryotes than the other orders, *Petrotogales* and *Kosmotogales*, in the phylum *Thermotogae* (Bhandari and Gupta, 2014). In some *T. maritima* MSB8 proteins, signal cleavage sites of gram-negative and/or gram-positive bacteria were also predicted in conjunction with those of eukaryotes, but never alone (Supplementary Table 2).

recruited by Ku and ligates the ends. Unfortunately, no homologs of Ku and LigD of *Mycobacterium* were found in *Thermotogales* genomes during a Blastp search (Altschul et al., 1990). However, we found a LigD homolog (WP_121509529.1) with an identity as large as 44% in a mesophilic *Thermotogae* (Bhandari and Gupta, 2014), *Mesotoga* sp. H07pep.5.4 (Nesbø and Charchuk, 2019). Thus, the fragment-joining system in *Thermotogales*, which could be analogous to the NHEJ system of *Mycobacterium*, cannot be excluded.

In the presumed fragment joining system in *Thermotogales*, the ligation of DNA fragments would occur each time following fusion of immature progeny, which must result in a sequence different from that in the chromosome, since the order of the fusion of immature progeny is expected to be random. Thus, reorganization of random ligated DNA may be necessary to produce the rod chromosome. One of the possible mechanisms of reorganization would be homologous recombination with fragments ex-

cised from different loci of the MILG of large dish-shaped cells. For example, by using fragments including the border of DNAs neighboring in the MILG, the DNAs separated by unrelated fragments in the random ligated product could regain the neighboring locations. Nevertheless, this mechanism may not be sufficient for reorganization. Other mechanisms such as the knock-in activities of CRISPR systems (Auer et al., 2014), could also be involved in reorganization. Regardless, the chromosome formation by fragment joining followed by reorganization cannot be excluded.

3. HGT-centric eukaryogenesis originating from *Thermotogales* ancestor spheroids

3.1. Virus-related particles with large DNAs and broad host ranges

We postulate, in this article, a *Thermotogales* origin scenario of eukaryogenesis. The genetic vehicles transferring archaeal informational genes could have been virus-related particles, as described in the next paragraph. Thus, this scenario could be regarded as a version of viral eukaryogenesis (Livingstone Bell, 2001) with the specific host, *Thermotogales* ancestor spheroids, which were speculated to have large cells and genomes, and an endomembrane.

Accumulating evidence indicates that extracellular vesicles harboring viral genomes (Gaudin et al., 2014), also called vector particles (Chiura, 2004; Sugitate and Chiura, 2007) or virus-like particles (Geslin et al., 2007; Lossouarn et al., 2015a), and viruses (Gorlas et al., 2012; Krupovic et al., 2018; Lossouarn et al., 2015a, Lossouarn et al., 2015; Mercier et al., 2018) obtained from hydrothermal environments play crucial roles in the transfer of genetic elements among (hyper)thermophiles. The extracellular vesicles prepared from cell-free hydrothermal fluids have Aquificales-related dsDNA of 190–400 kb (Chiura, 2004; Chiura and Umitsu, 2004) or *Thermococcus*-related dsDNA of 310 kb (Sugitate and Chiura, 2007), both of which have the ability to alter the phenotypes of *E. coli*, suggesting broad host ranges. In contrast, vesicles (Gaudin et al., 2014; Geslin et al., 2007; Lossouarn et al., 2015a) or viruses (Gorlas et al., 2012; Krupovic et al., 2018; Lossouarn et al., 2015a, Lossouarn et al., 2015b; Mercier et al., 2018) isolated from particle-infected cells have dsDNA of 10–50 kb, with narrower host ranges. The DNA size and the host range of extracellular vesicles appear to be correlated, as observed with viruses (Krupovic et al., 2018). Regardless, the DNA sizes would be large enough to transfer the archaeal informational genes, in batches by more than single infection (Lowe and Paschke, 1968) when necessary, simultaneously. These virus-related particles would have been involved in the HGT of archaeal and bacterial genes into *Thermotogales* ancestor spheroids.

3.2. Establishment of FECA

If the fragment joining followed by the homologous recombination-mediated reorganization represents an aspect of chromosome formation in progeny, the establishment of FECA as well as the eukaryotic evolution could be, at least partly, explained. It would be reasonable to speculate that immature progeny fused with each other and formed eukaryotic endomembranes, including the nuclear membrane (Fig. 3), as suggested in the OM vesicle-derived endomembrane hypothesis (Gould et al., 2016). For the endomembrane formation, the mature rod formation from immature progeny should be prevented. We posit that the inability of dish-shaped cells to produce PG (Kuwabara and Igarashi, 2017) likely eased the permanent cessation of PG formation, for example, by inserting unrelated fragments into PG-forming genes. It is noteworthy that *Thermotogales* OM, which does not contain LPS (Sutcliffe, 2010), could be used for the CM of

the resulting eukaryote-like cells without significant modifications of lipids.

We assume that ancestors of *Thermotogales* had no CRISPR sequences. Therefore, their spheroids are expected to have been able to acquire a large amount of foreign DNA, taking advantage of the absence of the physical barrier, PG, without degrading them through CRISPR-related immunity (Horvath and Barrangou, 2010). Thus, it would be possible that virus-related particles with large DNA and broad host ranges transferred massive archaeal genes simultaneously, including those for cell division (Lindås et al., 2008) and cytoskeletons for phagocytosis (Yutin et al., 2009; Zaremba-Niedzwiedzka et al., 2017). The MILG of large dish-shaped cells would have contained DNA more foreign than *Thermotogales*' genetic material (for example 40 Mb in total, referring to the genome size of *N. gruberi*, vs 2 Mb, that of *Thermotogales*). The *Thermotogales* and foreign DNA fragments in immature progeny should have been ligated in a random order. The resulting DNAs had a linear morphology because the ends were unrelated with each other. Most DNAs were dysfunctional due to the inappropriateness of the joining order, but some could have had a chance to have functional genes owing to the aforementioned homologous recombination-mediated reorganization of DNA. The spheroids, in which the archaeal cell division genes (Lindås et al., 2008) became functional, were the FECA (Fig. 3). The absence of an active bacterial cell division system in the spheroids (Kuwabara and Igarashi, 2017) should have eased the recruitment of an archaeal cell division system. Even if massive HGT failed to transfer all archaeal informational genes, some bacterial genes could have substituted the missing archaeal ones. Such FECA members would have obtained the archaeal genes through subsequent HGT, which eventually replaced the bacterial genes. The selective force of replacement was the faster growth provided by the archaeal genes, which matched with the archaeal cell division system.

Archaeal gene donors are likely to have been the deep-sea members of Asgard archaea (Zaremba-Niedzwiedzka et al., 2017), consistent with one of the habitats of *Thermotogales*, deep-sea hydrothermal vents, but their pelagic members (Zaremba-Niedzwiedzka et al., 2017) could also have been involved in the HGT; pelagic SAR11 clade α -proteobacteria (relating to *Candidatus Pelagibacter ubique*) were observed in seawater near the habitat of *T. globiformans*³ (Kato et al., 2009). Since *Thermotogales* bacteria are suggested to have obtained as many as 80% of their protein-coding genes by HGT (Zhaxybayeva et al., 2009), it cannot be excluded that some genes of other bacteria were transferred before or upon the massive HGT, namely before acquisition of mitochondria, supporting the late-mitochondria acquisition (Koumandou et al., 2013). The origins of bacterial genes appear varied, as evaluated from the protein-coding genes of *S. cerevisiae* and *C. merolae* (Koonin, 2010), although these data would also reflect the HGT after the FECA formation.

3.3. Fragment joining suggests the mechanisms of eukaryotic evolution

The fragment joining followed by the reorganization could have enabled various chromosome compositions of FECA members, even from the spheroids having an identical gene repertoire (Fig. 3). These members are thought to have evolved into supergroups. With partial reorganization of random joined products, existing genes would be disrupted, while novel genes could be produced by chance, which should have been the genomic innovation process of eukaryogenesis. Genes not producing indispensable proteins are thought to have had a chance to be extinct during the reductive evolution from LECA (Fritz-Laylin et al., 2010; Koonin, 2010). Some gene-disrupting inserts became introns, when their corre-

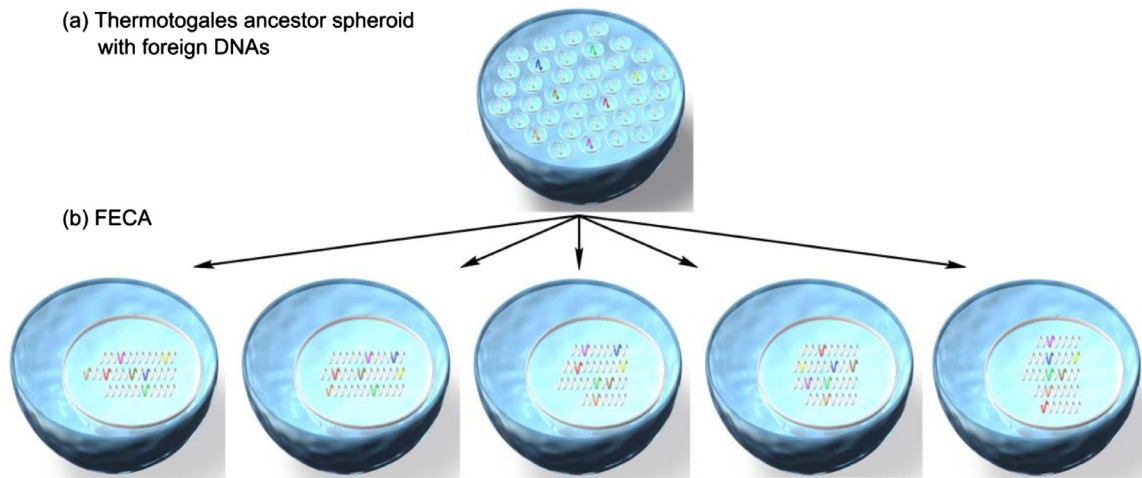


Fig. 3. A FECA formation model.

Source: Foreign DNAs were horizontally transferred into *Thermotogales* ancestor spheroids, whose growth yielded multicopy-integrated large genome (MILG). Immature progeny therefrom contained a fragment derived from *Thermotogales* (colored) or foreign (white) DNA (a). Immature progeny containing *Thermotogales* and foreign DNAs neighboring in the MILG would have also been produced, but are not shown for simplicity. The immature progeny fused with each other forming the nuclear membrane, with their DNAs random ligated, followed by homologous recombination-mediated reorganization. Among the resulting spheroids, those that became capable of division owing to the archaeal cell division genes are shown as FECA (b). Note that owing to the random joining of DNA fragments, the diversity is introduced in the FECA members; the lengths, numbers, and gene repertoires of chromosomes varied among them.

sponding parts in the primary transcripts happened to be spliced for end modification (Livingstone Bell, 2001), forming mRNA. All eukaryotes, including *N. gruberi* (Fritz-Laylin et al., 2010), have introns. Thus, it would not be surprising if introns were produced in the FECA. Introns are known to be involved in exon shuffling (Gilbert, 1978), a natural process of creating new combinations of exons by intronic recombination (Kolkman and Stemmer, 2001; Long et al., 2003; Patthy, 1999). One of the mechanisms of exon shuffling, illegitimate recombination (van Rijk and Bloemendal, 2003), is dependent on the NHEJ pathway for dsDNA-break repair (Kegel et al., 2006). Exon shuffling can produce new genes and is, thus, considered to be an evolutionary driving force in eukaryotes (Long et al., 2003; Patthy, 1999). Therefore, the presumed fragment joining in *Thermotogales* ancestor spheroids, which is thought to have generated supergroups, novel genes, and introns, likely provided basic mechanisms for eukaryotic evolution.

3.4. Establishment of LECA

Although our scenario covers up to the establishment of FECA, an example of LECA formation compatible with the FECA should be described. The members of the FECA community are thought to have acquired mitochondria by phagocytosis, supported by archaeal actin-related protein homologs (Yutin et al., 2009; Zaremba-Niedzwiedzka et al., 2017), and became the LECA. The FECA should have conserved energy for phagocytosis under anaerobic conditions, which was exerted under partially aerobic conditions, where they encountered the proto-symbiont α -proteobacteria. The geological site for the mitochondria acquisition may have been a partially aerobic hydrothermal vent area. Each member of the FECA community should have captured a distinct α -proteobacterium, implying that the mitochondrial origin was not unique among the members, although the selection after capture could have caused it to appear so. Members not acquiring mitochondria may not have conserved the energy for proliferation under aerobic conditions and, thus, are thought to have been extinct. However, such members could have survived under anaerobic conditions, sim-

ilar to an amitochondriate animal symbiont, *Monocercomonoides* sp. PA203, which entirely lacks mitochondrial proteins and is thought to have lost the mitochondria that were previously obtained (Karnkowska et al., 2016).

4. Experimental evaluation of the *Thermotogales* origin scenario is possible

The scenario can be experimentally evaluated by using *T. globiformans* spheroids. Whether or not the spheroids become capable of cell division, the most critical step of the scenario, would be to study the introduction of archaeal cell division genes. Nevertheless, there is a concern that the CRISPR-related immunities (Horvath and Barrangou, 2010) could prevent the acquisition of foreign genes. Thus, prior elimination of these activities may be necessary. The genome structure of large dish-shaped cells can be studied by single-cell sequencing (Eberwine et al., 2013). Doing so would suggest the mechanisms of large genome formation, progeny production, and prevention of cell division and PG accumulation. Fragment joining in progeny chromosome formation can be evaluated through purification of immature progeny. Metagenomic and proteomic analyses would reveal DNA sequences in immature progeny and the supposed DNA-joining enzymes. The eukaryotic signal cleavage sites in the *Thermotogales* periplasmic proteins can be studied by N-terminal sequencing and the signal cleave of the *in vitro* translated proteins.

5. Conclusions

A eukaryogenesis scenario is proposed that *Thermotogales* ancestor spheroids became capable of division owing to the horizontally transferred archaeal cell division genes, becoming the FECA. The genetic vehicles involved in the HGT were virus-related particles with large DNA and broad host ranges. The archaeal genes necessary for eukaryogenesis, including those for cell division and phagocytosis, were transferred practically at a time.

Spheroids of extant *Thermotogales* can show some eukaryote-specific features, such as large nucleoids, which are formed by

integration of multiple genome copies, and an endomembrane, which is the CM of immature progeny produced in periplasm, in addition to the G3P-type membrane lipids and the endoplasmic reticulum-type signal cleavage sites in periplasmic proteins. Thus, *Thermotogales* could be regarded as a proto-eukaryote. The former two features are acquired through *Thermotogales*-specific IOM and subsequent fusion of multiple cells, which are brought by retrogressing alterations of PG, the landmark of bacteria. Therefore, the *Thermotogales* cell structures may be related to eukaryogenesis. We hypothesize that the progeny chromosome is formed in immature progeny by random joining small DNAs excised from the MILG, followed by reorganization of the products by mechanisms including homologous recombination enabled with MILG.

In *Thermotogales* ancestor spheroids, the presumed absence of CRISPR-related immunity enabled the MILG to contain archaeal and bacterial DNAs horizontally transferred via virus-related particles. Fusion of immature progeny containing small DNAs from the MILG formed endomembranes including the nuclear membrane and random ligated DNAs. The spheroids, in which the archaeal cell division genes became functional through reorganization of random ligated DNAs by MILG-enabled homologous recombination, were the FECA. It means that the large genome, considered a eukaryote-specific feature, was not a consequence, but a prerequisite of eukaryogenesis. The random joining of DNAs likely provided the basic mechanisms for eukaryotic evolution, generating various FECA members leading to supergroups, novel genes, and introns that are involved in exon shuffling. Each FECA member phagocytosed a distinct α -proteobacterium as mitochondria, using the energy conserved under anaerobic conditions, which formed the LECA.

Author contributions

T.K. contributed to the sequence analyses and manuscript writing; K.I. contributed to sequence analyses and experiments.

Correspondence and requests for materials should be addressed to kuwabara@biol.tsukuba.ac.jp

Declaration of Competing Interest

There are no financial, general, or institutional competing interests.

Acknowledgements

We thank Emeritus Prof. Humitake Seki for providing the previous plant physiologist, T.K., the opportunity to visit Suiyo Seamount through the Special Coordination Fund from the Ministry of Education, Culture, Sports, Science and Technology of Japan for the Archaeal Park Project; an international research project on interaction between the sub-vent biosphere and the geo-environment. We also thank Prof. Hiroshi X. Chiura for the discussion on extracellular vesicles and Prof. Naoki Sato for discussion on endosymbiosis. We are grateful to Editage (<http://www.editage.jp>) for English language editing.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jtbi.2020.110192.

References

Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S., Brown, M.W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., Le Gall, L., Lynn, D.H., McManus, H., Mitchell, E.A.D., Mozley-Stanridge, S.E., Parfrey, L.W., Pawłowski, J., Rueckert, S., Shadwick, L., Schoch, C.L., Smirnov, A., Spiegel, F.W., 2012. The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* 59, 429–493.

Aitken, A., Stanier, R.Y., 1979. Characterization of peptidoglycan from the cyanelles of *Cyanophora paradoxa*. *J. Gen. Microbiol.* 112, 219–223.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi:10.1016/S0022-2836(05)80360-2.

Auer, T.O., Duroure, K., De Cian, A., Concordet, J.-P., Del Bene, F., 2014. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res.* 24, 142–153. doi:10.1101/gr.161638.113.

Ausili, A., Staiano, M., Dattelbaum, J., Varriale, A., Capo, A., D'Auria, S., 2013. Periplasmic binding proteins in thermophiles: characterization and potential application of an arginine-binding protein from *Thermotoga maritima*: a brief thermo-story. *Life* 3, 149–160.

Bagos, P.G., Tsigos, K.D., Plessas, S.K., Liakopoulos, T.D., Hamodrakas, S.J., 2009. Prediction of signal peptides in archaea. *Protein Eng. Des. Sel.* 22, 27–35. doi:10.1093/protein/gzn064.

Bhandari, V., Gupta, R.S., 2014. Molecular signatures for the phylum (class) *Thermotogae* and a proposal for its division into three orders (*Thermotogales*, *Kosmotogales* ord. nov. and *Petrotogales* ord. nov.) containing four families (*Thermotogaceae*, *Fervidobacteriaceae* fam. nov., *Kosmotogaceae* fam. nov. and *Petrotogaceae* fam. nov.) and a new genus *Pseudothermotoga* gen. nov. with five new combinations. *Antonie Van Leeuwenhoek* 105, 143–168. doi:10.1007/s10482-013-0062-7.

Boniface, A., Bouhss, A., Mengin-Lecreulx, D., Blanot, D., 2006. The MurE synthetase from *Thermotoga maritima* is endowed with an unusual D-lysine adding activity. *J. Biol. Chem.* 281, 15680–15686. doi:10.1074/jbc.M506311200.

Boniface, A., Parquet, C., Arthur, M., Mengin-Lecreulx, D., Blanot, D., 2009. The elucidation of the structure of *Thermotoga maritima* peptidoglycan reveals two novel types of cross-link. *J. Biol. Chem.* 281, 21856–21862. doi:10.1074/jbc.M109.034363.

Brissett, N.C., Doherty, A.J., 2009. Repairing DNA double-strand breaks by the prokaryotic non-homologous end-joining pathway. *Biochem. Soc. Trans.* 37, 539–545. doi:10.1042/bst0370539.

Brodth, A., Lurie-Weinberger, M.N., Gophna, U., 2011. CRISPR loci reveal networks of gene exchange in archaea. *Biol. Direct* 6, 65. doi:10.1186/1745-6150-6-65.

Cavalier-Smith, T., 1989. Archaeobacteria and archezoa. *Nature* 339, 100–101. doi:10.1038/339100a0.

Chiura, H.X., 2004. Novel broad-host range gene transfer particles in nature. *Microbes Environ.* 19, 249–264. doi:10.1264/jsme2.19.249.

Chiura, H.X., Umitsu, M., 2004. Isolation and characterisation of broad-host range gene transporter particles from geo-thermal vent of the Toyoha Mine. *Microbes Environ.* 19, 20–30. doi:10.1264/jsme2.19.20.

Choo, K.H., Tan, T.W., Ranganathan, S., 2009. A comprehensive assessment of N-terminal signal peptides prediction methods. *BMC Bioinformatics* 15 (10 Suppl). doi:10.1186/1471-2105-10-S15-S2, S2-S2.

Degli Esposti, M., 2016. Late mitochondrial acquisition, really? *Genome Biol. Evol.* 8, 2031–2035. doi:10.1093/gbe/evw130.

Eberwine, J., Sul, J.-Y., Bartfai, T., Kim, J., 2013. The promise of single-cell sequencing. *Nat. Methods* 11, 25–27. doi:10.1038/nmeth.2769.

Embley, T.M., Martin, W., 2006. Eukaryotic evolution, changes and challenges. *Nature* 440, 623–630.

Engel, Michael, A., Cejka, Z., Lupas, A., Lottspeich, F., Baumeister, W., 1992. Isolation and cloning of *OmpA*, a coiled-coil protein spanning the periplasmic space of the ancestral eubacterium *Thermotoga maritima*. *EMBO J.* 11, 4369–4378.

Errington, J., 2003. Regulation of endospore formation in *Bacillus subtilis*. *Nat. Rev. Micro.* 1, 117–126. doi:10.1038/nrmicro750.

Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27, 401–410. doi:10.2307/2412923.

Fritz-Laylin, L.K., Prochnik, S.E., Ginger, M.L., Dacks, J.B., Carpenter, M.L., Field, M.C., Kuo, A., Paredes, A., Chapman, J., Pham, J., Shu, S., Neupane, R., Cipriano, M., Mancuso, J., Tu, H., Salamov, A., Lindquist, E., Shapiro, H., Lucas, S., Grigoriev, I.V., Cande, W.Z., Fulton, C., Rokhsar, D.S., Dawson, S.C., 2010. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell* 140, 631–642.

Gabalón, T., 2018. Relative timing of mitochondrial endosymbiosis and the “pre-mitochondrial symbioses” hypothesis. *IUBMB Life* 70, 1188–1196. doi:10.1002/iub.1950.

Gaudin, M., Krupovic, M., Marguet, E., Gauliard, E., Cvirkaitė-Krupovic, V., Le Cam, E., Oberio, J., Forterre, P., 2014. Extracellular membrane vesicles harbouring viral genomes. *Environ. Microbiol.* 16, 1167–1175. doi:10.1111/1462-2920.12235.

Geslin, C., Gaillard, M., Flament, D., Rouault, K., Le Romancer, M., Prieur, D., Erauso, G., 2007. Analysis of the first genome of a hyperthermophilic marine virus-like particle, PAV1, isolated from *Pyrococcus abyssi*. *J. Bacteriol.* 189, 4510–4519.

Gilbert, W., 1978. Why genes in pieces? *Nature* 271, 501. doi:10.1038/271501a0.

Glickman, M.S., 2014. Double-strand DNA break repair in Mycobacteria. *Microbiol. Spectr.* 2. doi:10.1128/microbiolspec.MGM2-0024-2013.

Gorlas, A., Koonin, E.V., Bienvenu, N., Prieur, D., Geslin, C., 2012. TPV1, the first virus isolated from the hyperthermophilic genus *Thermococcus*. *Environ. Microbiol.* 14, 503–516. doi:10.1111/j.1462-2920.2011.02662.x.

Gould, S.B., Garg, S.G., Martin, W.F., 2016. Bacterial vesicle secretion and the evolutionary origin of the eukaryotic endomembrane system. *Trends Microbiol.* 24, 525–534.

Guy, L., Ettema, T.J.G., 2011. The archaeal ‘TACK’ superphylum and the origin of eukaryotes. *Trends Microbiol.* 19, 580–587.

Hildenbrand, C., Stock, T., Lange, C., Rother, M., Soppa, J., 2011. Genome copy numbers and gene conversion in methanogenic archaea. *J. Bacteriol.* 193, 734–743. doi:10.1128/JB.01016-10.

- Horvath, P., Barrangou, R., 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327, 167–170.
- Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C., Stetter, K.O., 2002. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* 417, 63–67.
- Huber, R., Langworthy, T.A., König, H., Thomm, M., Woese, C.R., Sleytr, U.B., Stetter, K.O., 1986. *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Arch. Microbiol.* 144, 324–333.
- Huber, R., Stetter, K.O., et al., 1992. The order *Thermotogales*. In: Balows, A., et al. (Eds.), *The Prokaryotes*. Springer, New York, pp. 3809–3815.
- Huber, R., Woese, C.R., Langworthy, T.A., Fricke, H., Stetter, K.O., 1989. *Thermosiphon africanus* gen. nov., represents a new genus of thermophilic eubacteria within the “*Thermotogales*”. *Syst. Appl. Microbiol.* 12, 32–37.
- Huber, R., Woese, C.R., Langworthy, T.A., Kristjansson, J.K., Stetter, K.O., 1990. *Fervidobacterium islandicum* sp. nov., a new extremely thermophilic eubacterium belonging to the “*Thermotogales*”. *Arch. Microbiol.* 154, 105–111.
- Igarashi, K., 2019. Draft genome sequence of *Thermosiphon globiformans* strain MN14. *Microbiol. Resour. Announc.* 8 (e01728–18). doi:10.1128/MRA.01728-18.
- Jannasch, W., H., Mottl, J., M., 1985. Geomicrobiology of Deep-Sea Hydrothermal Vents. *Science* 229, 717–725. doi:10.1126/science.229.4715.717.
- Karnkowska, A., Vacek, V., Zubáčová, Z., Treitli, S.C., Petrželková, R., Eme, L., Novák, S., Moriguchi, R., Barlow, L.D., Herman, E.K., Soukal, P., Hroudová, M., Doležal, P., Stairs, C.W., Roger, A.J., Eliáš, M., Dacks, J.B., Vlček, Č., Hampl, V., 2016. A eukaryote without a mitochondrial organelle. *Curr. Biol.* 26, 1274–1284. doi:10.1016/j.cub.2016.03.053.
- Kato, S., Hara, K., Kasai, H., Teramura, T., Sunamura, M., Ishibashi, J.-i., Kakegawa, T., Yamanaka, T., Kimura, H., Marumo, K., Urabe, T., Yamagishi, A., 2009. Spatial distribution, diversity and composition of bacterial communities in sub-seafloor fluids at a deep-sea hydrothermal field of the Suiyo Seamount. *Deep-Sea Res.* 1844–1855. doi:10.1016/j.dsr.2009.05.004, PT 1 56.
- Kawasaki, S., Moriguchi, R., Sekiya, K., Nakai, T., Ono, E., Kume, K., Kawahara, K., 1994. The cell envelope structure of the lipopolysaccharide-lacking gram-negative bacterium *Sphingomonas paucimobilis*. *J. Bacteriol.* 176, 284–290. doi:10.1128/jb.176.2.284-290.1994.
- Kegel, A., Martinez, P., Carter, S.D., Aström, S.U., 2006. Genome wide distribution of illegitimate recombination events in *Kluyveromyces fragilis*. *Nucleic Acids Res.* 34, 1633–1645. doi:10.1093/nar/gkl064.
- Kolkman, J.A., Stemmer, W.P.C., 2001. Directed evolution of proteins by exon shuffling. *Nat. Biotechnol.* 19, 423–428.
- Koonin, E.V., 2010. The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biol.* 11, 209. doi:10.1186/gb-2010-11-5-209.
- Koumandou, V.L., Wickstead, B., Ginger, M.L., van der Giesen, M., Dacks, J.B., Field, M.C., 2013. Molecular paleontology and complexity in the last eukaryotic common ancestor. *Crit. Rev. Biochem. Mol. Biol.* 48, 373–396. doi:10.3109/10409238.2013.821444.
- Krupovic, M., Cvirkaite-Krupovic, V., Iranzo, J., Prangishvili, D., Koonin, E.V., 2018. Viruses of archaea: structural, functional, environmental and evolutionary genomics. *Virus Res.* 244, 181–193. doi:10.1016/j.virusres.2017.11.025.
- Kuwabara, T., Igarashi, K., 2012. Microscopic studies on *Thermosiphon globiformans* implicate a role of the large periplasm of *Thermotogales*. *Extremophiles* 16, 863–870.
- Kuwabara, T., Igarashi, K., 2017. Progeny production in the periplasm of *Thermosiphon globiformans*. *Extremophiles* 21, 805–815.
- Kuwabara, T., Kawasaki, A., Uda, I., Sugai, A., 2011. *Thermosiphon globiformans* sp. nov., an anaerobic thermophilic bacterium that transforms into multicellular spheroids with a defect in peptidoglycan formation. *Int. J. Syst. Evol. Microbiol.* 61, 1622–1627.
- Kuwabara, T., Minaba, M., Saida, H., Urabe, T., Kamekura, M., 2006. A novel in situ cultivation device for isolation of thermophilic anaerobes of deep-sea hydrothermal vents. *Biosystem Stud.* 9, 218–227.
- Lam, H., Oh, D.-C., Cava, F., Takacs, C.N., Clardy, J., de Pedro, M.A., Waldor, M.K., 2009. D-amino acids govern stationary phase cell wall remodeling in bacteria. *Science* 325, 1552–1555.
- Lane, N., Martin, W., 2010. The energetics of genome complexity. *Nature* 467, 929–934. doi:10.1038/nature09486.
- Leaver, M., Dominguez-Cuevas, P., Coxhead, J.M., Daniel, R.A., Errington, J., 2009. Life without a wall or division machine in *Bacillus subtilis*. *Nature* 457, 849–853.
- L’Haridon, S., Miroshnichenko, L., M., Hippe, H., Fardeau, M.-L., Bonch-Osmolovskaya, E., Stackebrandt, E., Jeanthon, C., 2001. *Thermosiphon geolei* sp. nov., a thermophilic bacterium isolated from a continental petroleum reservoir in Western Siberia. *Int. J. Syst. Evol. Microbiol.* 51, 1327–1334. doi:10.1099/00207713-51-4-1327.
- Liebl, W., Winterhalter, C., Baumeister, W., Armbricht, M., Valdez, M., 2008. Xylanase attachment to the cell wall of the hyperthermophilic bacterium *Thermotoga maritima*. *J. Bacteriol.* 190, 1350–1358. doi:10.1128/jb.01149-07.
- Lindås, A.C., Karlsson, E.A., Lindgren, M.T., Ettema, T.J., Bernander, R., 2008. A unique cell division machinery in the Archaea. *Proc. Natl. Acad. Sci. USA* 105, 18942–18946.
- Livingstone Bell, P.J., 2001. Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? *J. Mol. Evol.* 53, 251–256. doi:10.1007/s002390010215.
- Long, M., Betrán, E., Thornton, K., Wang, W., 2003. The origin of new genes: glimpses from the young and old. *Nat. Rev. Genet.* 4, 865–875. doi:10.1038/nrg1204.
- Lossouarn, J., Dupont, S., Gorlas, A., Mercier, C., Biennu, N., Marguet, E., Forterre, P., Geslin, C., 2015a. An abyssal mobilome: viruses, plasmids and vesicles from deep-sea hydrothermal vents. *Res. Microbiol.* 166, 742–752. doi:10.1016/j.resmic.2015.04.001.
- Lossouarn, J., Nesbø, C.L., Mercier, C., Zhaxybayeva, O., Johnson, M.S., Charchuk, R., Farasin, J., Biennu, N., Baudoux, A.-C., Michoud, G., Jebbar, M., Geslin, C., 2015b. ‘Ménage à trois’: a selfish genetic element uses a virus to propagate within *Thermotogales*. *Environ. Microbiol.* 17, 3278–3288. doi:10.1111/1462-2920.12783.
- Lowe, R.E., Paschke, J.D., 1968. Pathology of a double viral infection of *Trichoplusia ni*. *J. Invertebr. Pathol.* 12, 438–443. doi:10.1016/0022-2011(68)90351-0.
- Margolin, W., 2005. FtsZ and the division of prokaryotic cells and organelles. *Nat. Rev. Mol. Cell Biol.* 6, 862–871.
- Martin, W.F., Garg, S., Zimorski, V., 2015. Endosymbiotic theories for eukaryote origin. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* 370. doi:10.1098/rstb.2014.0330, 20140330–20140330.
- Martin, W.F., Roettger, M., Ku, C., Garg, S.G., Nelson-Sathi, S., Landan, G., 2017. Late mitochondrial origin is an artifact. *Genome Biol. Evol.* 9, 373–379. doi:10.1093/gbe/evx027.
- Mercier, C., Lossouarn, J., Nesbø, C.L., Haverkamp, T.H.A., Baudoux, A.C., Jebbar, M., Biennu, N., Thiroux, S., Dupont, S., Geslin, C., 2018. Two viruses, MCV1 and MCV2, which infect *Marinotoga* bacteria isolated from deep-sea hydrothermal vents: functional and genomic analysis. *Environ. Microbiol.* 20, 577–587. doi:10.1111/1462-2920.13967.
- Nanavati, D., Noll, K.M., Romano, A.H., 2002. Periplasmic maltose- and glucose-binding protein activities in cell-free extracts of *Thermotoga maritima*. *Microbiology* 148, 3531–3537. doi:10.1099/00221287-148-11-3531.
- Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., Eisen, J.A., White, O., Salzberg, S.L., Smith, H.O., Venter, J.C., Fraser, C.M., 1999. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399, 323–329.
- [dataset] Nesbø, C.L., Charchuk, R., 2019. Mesotoga genome sequencing. REFSEQ, JFHM01000001. <https://www.ncbi.nlm.nih.gov/nucleotide/JFHM01000001.1?from=111591&to=112193>.
- Nesbø, C.L., Bapteste, E., Curtis, B., Dahle, H., Lopez, P., Macleod, D., Dlutek, M., Bowman, S., Zhaxybayeva, O., Birkeland, N.-K., Doolittle, W.F., 2009. The genome of *Thermosiphon africanus* TCF52B: lateral genetic connections to the Firmicutes and Archaea. *J. Bacteriol.* 191, 1974–1978.
- Nesbø, C.L., Dlutek, M., Doolittle, W.F., 2006. Recombination in *Thermotoga*: implications for species concepts and biogeography. *Genetics* 172, 759–769. doi:10.1534/genetics.105.049312.
- Nielsen, H., 2017. Predicting secretory proteins with SignalP. In: Kihara, D. (Ed.), *Protein Function Prediction: Methods and Protocols*. Springer, New York, New York, NY, pp. 59–73.
- Pace, N.R., 1997. A molecular view of microbial diversity and the biosphere. *Science* 276, 734–740. doi:10.1126/science.276.5313.734.
- Palani, K., Kumaran, D., Burley, S.K., Swaminathan, S., 2012. Structure of a periplasmic glucose-binding protein from *Thermotoga maritima*. *Acta Cryst F68*, 1460–1464. doi:10.1107/S1744309112045241.
- Patthy, L., 1999. Genome evolution and the evolution of exon-shuffling — a review. *Gene* 238, 103–114. doi:10.1016/S0378-1119(99)00228-0.
- Pfanzagl, B., Zenker, A., Pittenauer, E., Allmaier, G., Martinez-Torrecuadrada, J., Schmid, E.R., De Pedro, M.A., Löffelhardt, W., 1996. Primary structure of cyanell peptidoglycan of *Cyanophora paradoxa*: a prokaryotic cell wall as part of an organelle envelope. *J. Bacteriol.* 178, 332. doi:10.1128/jb.178.2.332-339.1996.
- Pittis, A.A., Gabaldón, T., 2016. Late acquisition of mitochondria by a host with chimeric prokaryotic ancestry. *Nature* 531, 101–104. doi:10.1038/nature16941.
- Podosokorskaya, O.A., Bonch-Osmolovskaya, E.A., Godfrey, A., Gavrilov, S.N., Beskovaynaya, D.A., Sokolova, T.G., Kolganova, T.V., Toshchakov, S.V., Kublanov, I.V., 2014. *Thermosiphon activus* sp. nov., a thermophilic, anaerobic, hydrolytic bacterium isolated from a deep-sea sample. *Int. J. Syst. Evol. Microbiol.* 64, 3307–3313.
- Podosokorskaya, O.A., Kublanov, I.V., Reysenbach, A.-L., Kolganova, T.V., Bonch-Osmolovskaya, E.A., 2011. *Thermosiphon affectus* sp. nov., a thermophilic, anaerobic, cellulolytic bacterium isolated from a Mid-Atlantic Ridge hydrothermal vent. *Int. J. Syst. Evol. Microbiol.* 61, 1160–1164.
- Poole, A.M., Penny, D., 2006. Evaluating hypotheses for the origin of eukaryotes. *BioEssays* 29, 74–84. doi:10.1002/bies.20516.
- Rivera, M.C., Lake, J.A., 2004. The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* 431, 152–155.
- Rivera, M.C., Jain, R., Moore, J.E., Lake, J.A., 1998. Genomic evidence for two functionally distinct gene classes. *Proc. Natl. Acad. Sci. USA* 95, 6239–6244.
- Sato, N., Takano, H., 2017. Diverse origins of enzymes involved in the biosynthesis of chloroplast peptidoglycan. *J. Plant Res.* 130, 635–645. doi:10.1007/s10265-017-0935-3.
- Schwechheimer, C., Kuehn, M.J., 2015. Outer-membrane vesicles from Gram-negative bacteria: biogenesis and functions. *Nat. Rev. Micro.* 13, 605–619. doi:10.1038/nrmicro3525.
- Shen, H.-B., Chou, K.-C., 2007. Signal-3L: a 3-layer approach for predicting signal peptides. *Biochem. Biophys. Res. Commun.* 363, 297–303. doi:10.1016/j.bbrc.2007.08.140.

- Soppa, J., 2011. Ploidy and gene conversion in Archaea. *Biochem. Soc. Trans.* 39, 150–154. doi:[10.1042/BST0390150](https://doi.org/10.1042/BST0390150).
- Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., van Eijk, R., Schleper, C., Guy, L., Ettema, T.J.G., 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521, 173–179. doi:[10.1038/nature14447](https://doi.org/10.1038/nature14447).
- Sugitate, T., Chiura, H.X., 2007. Functional gene transfer towards a broad range of recipients with the aid of vector particles originating from thermophiles. In: *Proceedings of International Symposium on Extremophiles and Their Applications, Tokyo. J-stage*, pp. 141–147.
- Sutcliffe, I.C., 2010. A phylum level perspective on bacterial cell envelope architecture. *Trends Microbiol.* 18, 464–470. doi:[10.1016/j.tim.2010.06.005](https://doi.org/10.1016/j.tim.2010.06.005).
- Takahashi, K., Yamanaka, S., 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676. doi:[10.1016/j.cell.2006.07.024](https://doi.org/10.1016/j.cell.2006.07.024).
- Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., Xie, X.S., 2010. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329, 533–538. doi:[10.1126/science.1188308](https://doi.org/10.1126/science.1188308).
- Typas, A., Banzhaf, M., Gross, C.A., Vollmer, W., 2012. From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nat. Rev. Microbiol.* 10, 123–136.
- van Rijk, A., Bloemendal, H., 2003. Molecular mechanisms of exon shuffling: illegitimate recombination. *Genetica* 118, 245–249.
- von Heijne, G., 1983. Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem.* 133, 17–21. doi:[10.1111/j.1432-1033.1983.tb07424.x](https://doi.org/10.1111/j.1432-1033.1983.tb07424.x).
- Waters, E., Hohn, M.J., Ahel, I., Graham, D.E., Adams, M.D., Barnstead, M., Beeson, K.Y., Bibbs, L., Bolanos, R., Keller, M., Kretz, K., Lin, X., Mathur, E., Ni, J., Podar, M., Richardson, T., Sutton, G.G., Simon, M., Söll, D., Stetter, K.O., Short, J.M., Noordewier, M., 2003. The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc. Natl. Acad. Sci. USA* 100, 12984–12988.
- Wissel, M.C., Wendt, J.L., Mitchell, C.J., Weiss, D.S., 2005. The transmembrane helix of the *Escherichia coli* division protein FtsI localizes to the septal ring. *J. Bacteriol.* 187, 320–328.
- Young, K.D., 2008. Why spherical *Escherichia coli* dies: the inside story. *J. Bacteriol.* 190, 1497–1498. doi:[10.1128/JB.01975-07](https://doi.org/10.1128/JB.01975-07).
- Yutin, N., Wolf, M.Y., Wolf, Y.I., Koonin, E.V., 2009. The origins of phagocytosis and eukaryogenesis. *Biol. Direct* 4, 9.
- Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U., Stott, M.B., Nunoura, T., Banfield, J.F., Schramm, A., Baker, B.J., Spang, A., Ettema, T.J.G., 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541, 353–358. doi:[10.1038/nature21031](https://doi.org/10.1038/nature21031).
- Zhaxybayeva, O., Swithers, K.S., Lapierre, P., Fournier, G.P., Bickhart, D.M., DeBoy, R.T., Nelson, K.E., Nesbø, C.L., Doolittle, W.F., Gogarten, J.P., Noll, K.M., 2009. On the chimeric nature, thermophilic origin, and phylogenetic placement of the *Thermotogales*. *Proc. Natl. Acad. Sci. USA* 106, 5865–5870.