

## Research Article

# Ribosomal Protein Cluster Organization in Asgard Archaea

Madhan R. Tirumalai <sup>1</sup>, Raghavan V. Sivaraman Jr. <sup>2</sup>, Layla A. Kutty <sup>3</sup>, Eric L. Song <sup>3</sup>,  
and George E. Fox <sup>1</sup>

<sup>1</sup>Department of Biology and Biochemistry, University of Houston, Houston, TX 77204-5001, USA

<sup>2</sup>Silver Lab, Wyss Institute for Biologically Inspired Engineering, Boston, MA, USA

<sup>3</sup>University of Texas at Austin, Austin, TX, USA

Correspondence should be addressed to Madhan R. Tirumalai; [mrtirum2@central.uh.edu](mailto:mrtirum2@central.uh.edu)

Received 20 July 2023; Revised 31 August 2023; Accepted 8 September 2023; Published 29 September 2023

Academic Editor: Stefan Spring

Copyright © 2023 Madhan R. Tirumalai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

It has been proposed that the superphylum of Asgard Archaea may represent a historical link between the Archaea and Eukarya. Following the discovery of the Archaea, it was soon appreciated that archaeal ribosomes were more similar to those of Eukarya rather than Bacteria. Coupled with other eukaryotic-like features, it has been suggested that the Asgard Archaea may be directly linked to eukaryotes. However, the genomes of Bacteria and non-Asgard Archaea generally organize ribosome-related genes into clusters that likely function as operons. In contrast, eukaryotes typically do not employ an operon strategy. To gain further insight into conservation of the r-protein genes, the genome order of conserved ribosomal protein (r-protein) coding genes was identified in 17 Asgard genomes (thirteen complete genomes and four genomes with less than 20 contigs) and compared with those found previously in non-Asgard archaeal and bacterial genomes. A universal core of two clusters of 14 and 4 cooccurring r-proteins, respectively, was identified in both the Asgard and non-Asgard Archaea. The equivalent genes in the *E. coli* version of the cluster are found in the *S10* and *spc* operons. The large cluster of 14 r-protein genes (*uS19-uL22-uS3-uL29-uS17* from the *S10* operon and *uL14-uL24-uL5-uS14-uS8-uL6-uL18-uS5-uL30-uL15* from the *spc* operon) occurs as a complete set in the genomes of thirteen Asgard genomes (five Lokiarchaeotes, three Heimdallarchaeotes, one Odinararchaeote, and four Thorarchaeotes). Four less conserved clusters with partial bacterial equivalents were found in the Asgard. These were the *L30e* (*str* operon in Bacteria) cluster, the *L18e* (*alpha* operon in Bacteria) cluster, the *S24e-S27ae-rpoE1* cluster, and the *L31e*, *L12..L1* cluster. Finally, a new cluster referred to as *L7ae* was identified. In many cases, r-protein gene clusters/operons are less conserved in their organization in the Asgard group than in other Archaea. If this is generally true for nonribosomal gene clusters, the results may have implications for the history of genome organization. In particular, there may have been an early transition to or from the operon approach to genome organization. Other nonribosomal cellular features may support different relationships. For this reason, it may be important to consider ribosome features separately.

## 1. Introduction

Ever since the initial discovery of Archaea by Woese and Fox [1], the number of recognized lineages comprising the Archaea domain has rapidly increased. Currently, five superphyla (and 35-38 phyla) of Archaea are described in the NCBI database [2-4]. Of these, the recently discovered Asgard include several phyla and are the most diverse [5, 6]. To emphasize this, they have been referred to as a “superphylum” [7].

Amongst the three domains, there are approximately 102 recognized ribosomal protein (r-protein) families [8-11]. Essentially universal amongst these are seventeen large subunit and nineteen small subunit r-proteins. Many of these 36 r-proteins likely had an ancient origin possibly before the last universal common ancestor (LUCA) [10-16]. However, this conclusion is dependent on how the gene content of LUCA is identified. At least eight of these genes are found in multiple versions of LUCA (Rivas personal communication). The Archaea and the Eukarya

also share eleven large subunit (LSU) r-proteins and twenty small subunit (SSU) r-proteins. However, neither group shares additional r-proteins with the Bacteria. For this reason, it has been hypothesized [17, 18] that the eukaryotic translation system originated from an earlier archaeal version and was subsequently expanded.

In Bacteria, approximately 32 r-proteins from both subunits and two translation-related proteins are grouped into seven well-studied clusters. They are, viz., *alpha*, *beta/L10*, *L11*, *S10*, *S20*, *str*, and the *spc* clusters [19, 20]. In *E. coli* and other well-studied organisms, these clusters have been experimentally shown to be operons [21]. In general, we will refer to them as clusters with the implication that many of them could likely be operons. The *S10* and the *spc* operons (clusters) are the largest, each typically encoding ten-twelve r-proteins [21, 22]. Gene expression when known is frequently regulated by one of the r-protein components within the cluster. Besides these clusters, there are several smaller clusters as well. The r-protein clusters are widely conserved amongst bacterial species and remnants, and Bacteria-like clusters are also found in some chloroplasts and mitochondria [23]. However, it is not clear if the clusters seen in mitochondria are functional. If they are not, then their presence or absence is likely a mere reflection of their nonessentiality.

The universality of r-protein gene clusters is often linked with a long history of their association. They also represent ancient/regulatory relationships, with implications for early gene clusters in mini chromosomes. Eukaryotic transcription usually involves a single transcript. Thus, eukaryotic operons/clusters of genes are restricted to few organisms [24, 25]. While eukaryotes do have meaningful gene clusters, these clusters do not include r-protein genes [24, 26–29].

Earlier studies of non-Asgard Archaea showed that the universal r-protein genes are in clusters similar or identical to those found in Bacteria [20]. While three archaeal r-protein genes *S4e*, *L32e*, and *L19e* were found associated with the archaeal version of the *spc* operon, the nonuniversal archaeal r-protein gene *L18e* was part of the conserved *L13* cluster. 17 r-protein genes were part of one of the ten previously unrecognized gene clusters. These clusters were found to be associated with genes involved in the initiation of protein synthesis, transcription, or other cellular processes. Since such associations in the universal clusters could not be found, it was posited that the ribosome had its own independent line of evolution [20].

It has been clear since their initial discovery that archaeal ribosomes more closely resemble those of Eukarya than Bacteria as has been documented in detail [30–32]. However, the non-Asgard Archaea have not been shown to have any strong specific relationship to Eukarya, and the matter was not initially aggressively pursued. Efforts to understand the relationship between the domains of life were rekindled when it was discovered that the then newly discovered Asgard Archaea have many nonribosomal features that are shared with Eukarya [7, 31, 32]. This has fostered the hypothesis that Eukarya may have descended from the Archaea [33]. Asgard Archaea were proposed to represent a historical link or bridge between Archaea and Eukarya [34–41]. A recent report has suggested that the genomic

material is condensed and spatially distinct from the riboplasm within certain Loki- and Heimdallarchaeota cells, as further proof of the role of Asgard Archaea in eukaryogenesis [42]. Amongst the Asgard group, the Heimdallarchaeota have been proposed to be the closest to the Eukarya [43, 44]. Other analyses based on multiple gene sequences placed the Asgard between the TACK group of Archaea and Eukarya [7, 36, 40]. In contrast, one branch of the Asgard, namely, the Lokiarchaeota, and their close relatives were proposed to be closer to the Euryarchaeota, than to Eukarya [45].

Genes that are translated together frequently share similar origins, physical interactions, functions, or regulatory mechanism(s) [46]. The arrangement of genes in genomes is thus a window to understand how organisms are related. Assuming there was a universal ancestor of the three domains, it likely had a gene order/arrangement that possibly was a precursor of operons. Its evolutionary origins may have dated back to four billion years ago [47–51]. While some portions of this gene order underwent shuffling over evolutionary timescales [52, 53], the arrangement of some gene clusters continues to be conserved [54]. In particular, the translation machinery and the genes that encode the same are highly conserved. Their association frequently has had a long evolutionary history [55–58]. The origin of the ribosome has been posited to be strongly coupled with the early history and even origins of life [56, 59–63], which is further supported by recent papers by Bose et al. [64] and Bose et al. [65].

In order to gain further insight into the r-protein gene clusters in Asgard Archaea, the genome order of r-protein coding genes was analyzed and compared with non-Asgard archaeal and bacterial genomes.

## 2. Materials and Methods

**2.1. Retrieval of Sequences of Genomes and Protein Coding Sequences.** The data used herein are based on the availability of sequences as of July 2023. The feature table and GenBank/RefSeq sequences of the genomes of Asgard Archaea were obtained from the public databases of the National Center for Biotechnology Information (NCBI) [66–68] and the Integrated Microbial Genomes and Microbiomes (IMG/M) system of the DOE's Joint Genome Institute (JGI) [69, 70]. All the raw genomes and the annotated gene and protein sequences of the same were saved as such and used. When required, r-protein sequences were obtained from the ProteoVision server [71]. Genomes that were not annotated and deposited as raw sequences were excluded.

**2.2. Mapping of the r-protein Gene Cluster(s).** The feature table of each genome was first checked for the presence/absence of genes in any given cluster. The closest available homologs of those genes that appeared to be missing in a given Asgard genome were used to ascertain the presence/absence of the same using stand-alone blast [72]. Missing genes, which were found to be misannotated as “hypothetical proteins,” were identified using blastX search and/or gene sequence alignment and included in the cluster map(s) (Figures 1–5).

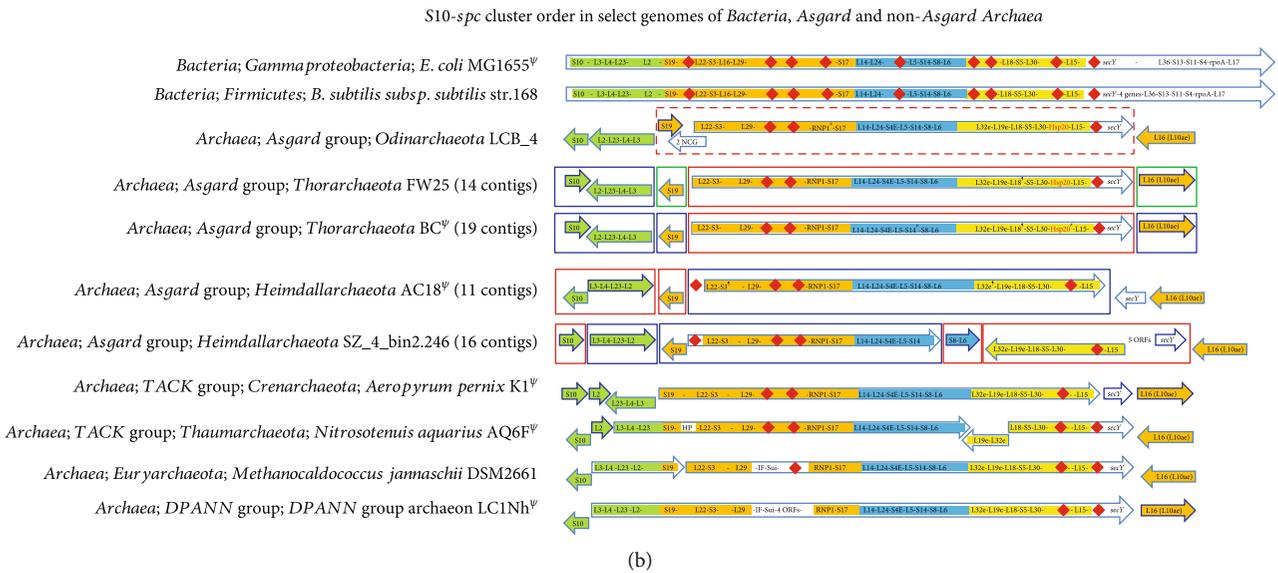
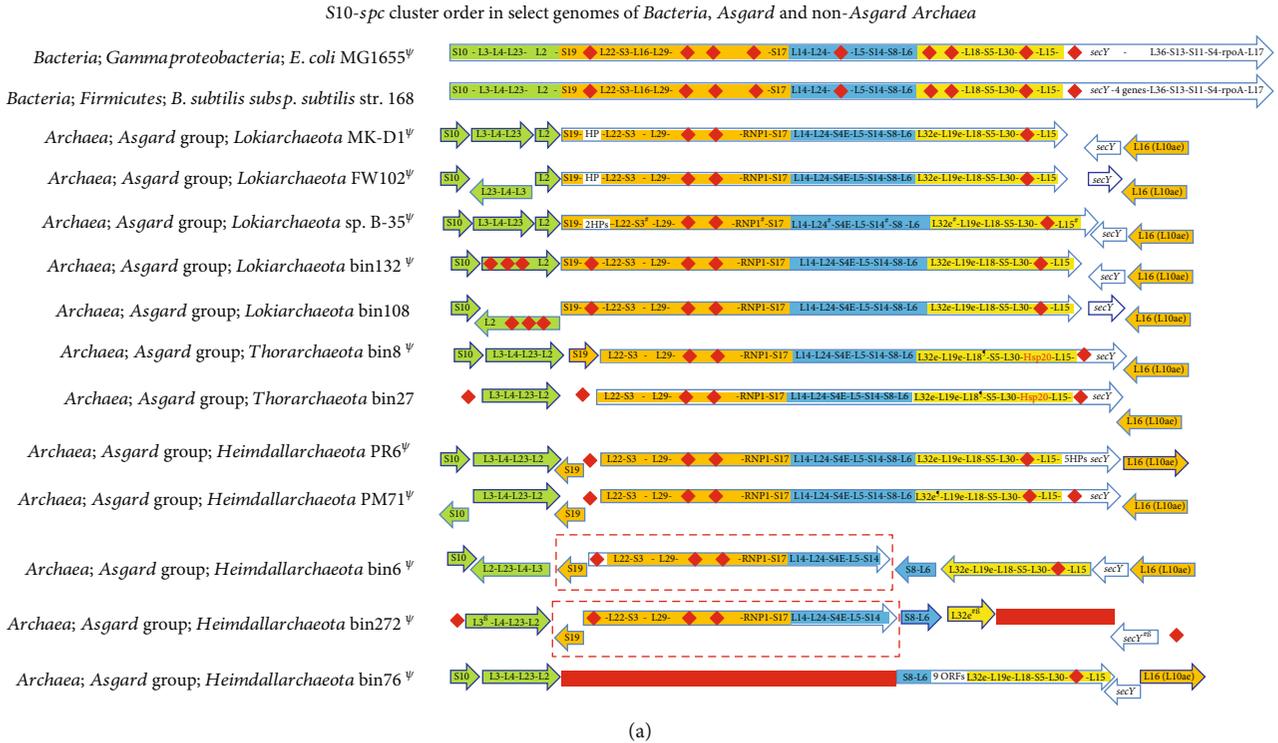


FIGURE 1: (a) <sup>ψ</sup>The orientation of all the genes in the entire cluster flipped for aligning with the other genomes. Each arrow represents a different location on a single complete scaffold; adjacent arrows do not indicate their order on the genome; genes within an arrow are contiguous. <sup>¶</sup>Pseudogene; red diamond and red rectangle represent gene(s) absent in the corresponding genome/location; arrows within dashed red boxes are contiguous; HP: hypothetical protein; RNP1: ribonuclease P protein component 1; ORF: open reading frame/gene; <sup>#</sup>ORF(s) annotated as HP(s); Hsp20: heat shock protein 20; <sup>β</sup>partial gene. (b) <sup>ψ</sup>The orientation of all the genes in the entire cluster flipped for aligning with the other genomes. In *Thorarchaeota* and *Heimdallarchaeota*, each colored (blue, green, or red) box represents a different contig; adjacent arrows do not indicate their order on the genome unless they are within a dashed red box; genes within an arrow are contiguous. <sup>¶</sup>Pseudogene; red diamond represents gene(s) absent in the corresponding genome/location; arrows within dashed red boxes are contiguous; HP: hypothetical protein; RNP1: ribonuclease P protein component 1; ORF: open reading frame/gene; <sup>#</sup>ORF(s) annotated as HP(s); Hsp20: heat shock protein 20; <sup>β</sup>partial gene; IF Sui: protein translation factor SUI1 homolog; NCG: noncluster genes.

### 3. Results

A total of ~552 *Asgard* genomes grouped under the various phyla that were available on the NCBI server ([https://www.](https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=1935183)

[ncbi.nlm.nih.gov/datasets/genome/?taxon=1935183](https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=1935183)) as of July 2023 were examined [36, 73]. Of all the *Asgard* genomes, only thirteen occur as a single complete (scaffold) genome(s) (Table 1). They will henceforth be referred to as

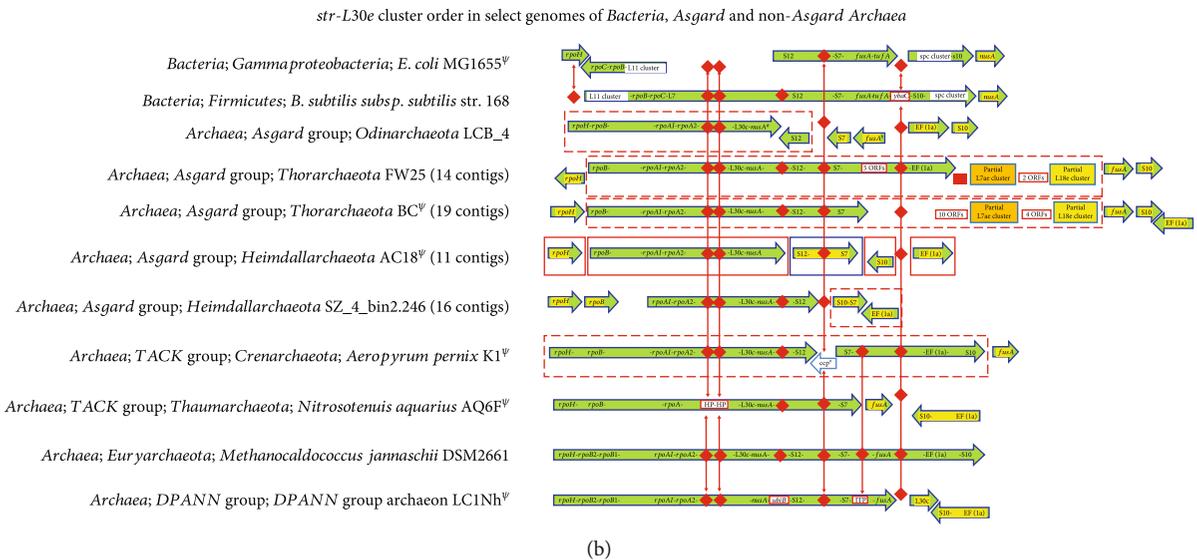
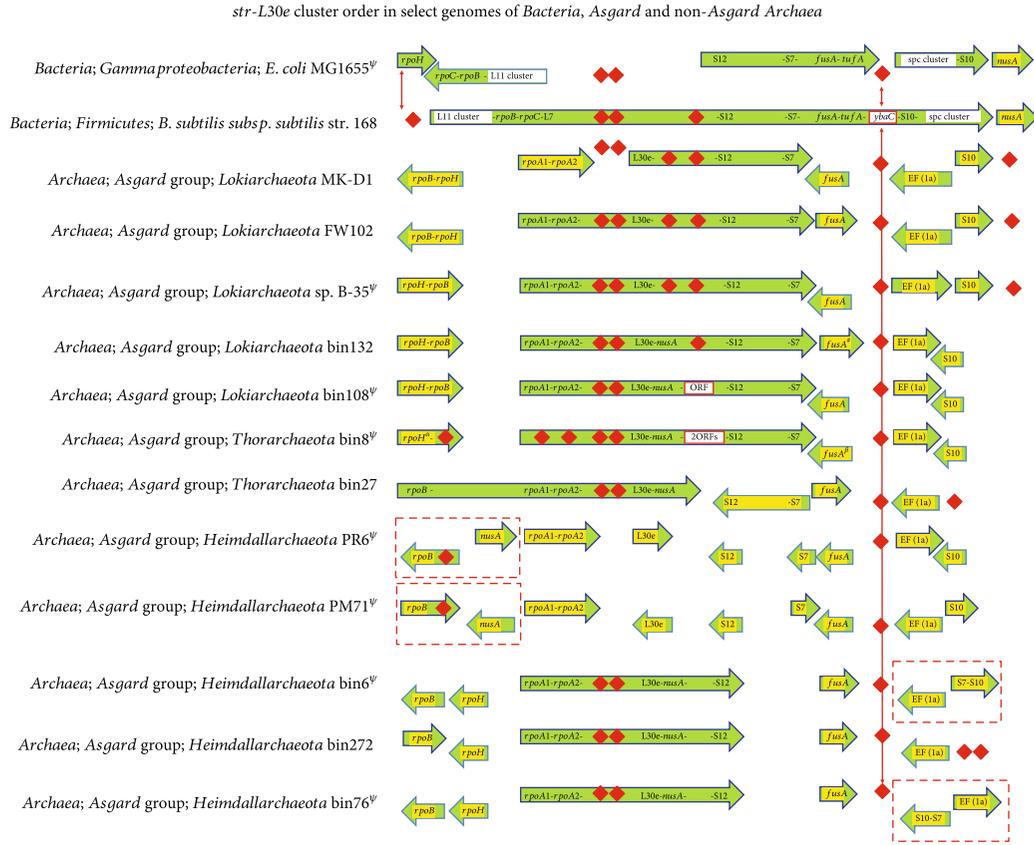
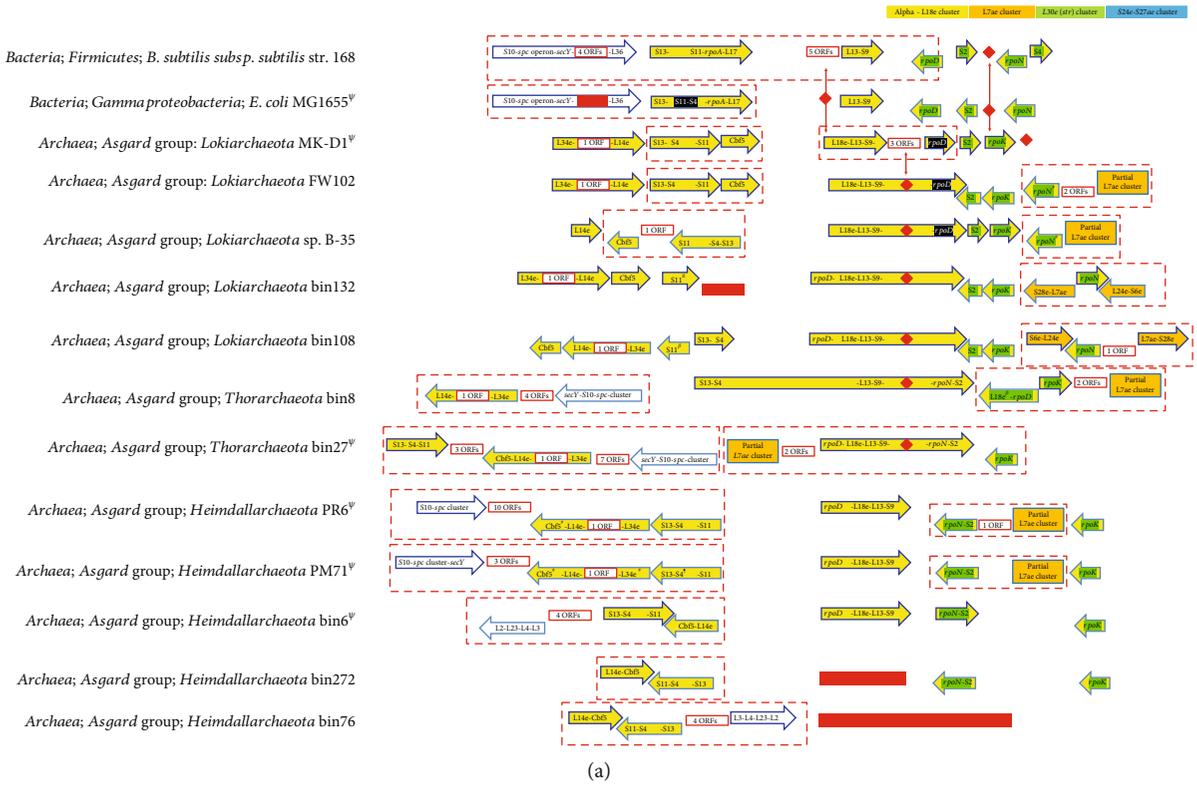


FIGURE 2: (a) <sup>ψ</sup>The orientation of all the genes in the entire cluster flipped for aligning with the other genomes. Each arrow represents a different location on the single complete scaffold/genome; adjacent arrows do not indicate their order on the genome; arrows within dashed red boxes are contiguous; genes within an arrow are contiguous; red diamond represents gene(s) absent in the corresponding genome/location. *ybaC*: proline iminopeptidase; HP: hypothetical protein; <sup>#</sup>ORF(s) annotated as HP(s); gene in yellow color occurs outside the main cluster in that genome; <sup>β</sup>partial gene; <sup>α</sup>duplicate copies of the gene adjacent to each other. (b) <sup>ψ</sup>The orientation of all the genes in the entire cluster flipped for aligning with the other genomes. Adjacent arrows do not indicate their order on the genome; arrows within dashed red boxes are contiguous; genes within an arrow are contiguous; red diamond and red rectangle represent gene(s) absent in the corresponding genome/location; in *Heimdallarchaeota* AC18, each colored (blue, green, or red) box represents a different contig; *ybaC*: proline iminopeptidase; HP: hypothetical protein; <sup>#</sup>ORF(s) annotated as HP(s); genes in yellow color occur outside the main cluster in that genome; <sup>β</sup>partial gene; <sup>α</sup>duplicate copies of the gene adjacent to each other; ccp: crenarchaeal conserved protein; *ubiB*: ubiquinone biosynthesis protein coding gene.

(Alpha operon) L18e cluster order in select genomes of *Asgard* and non-*Asgard* Archaea



(Alpha operon) L18e cluster order in select genomes of *Asgard* and non-*Asgard* Archaea

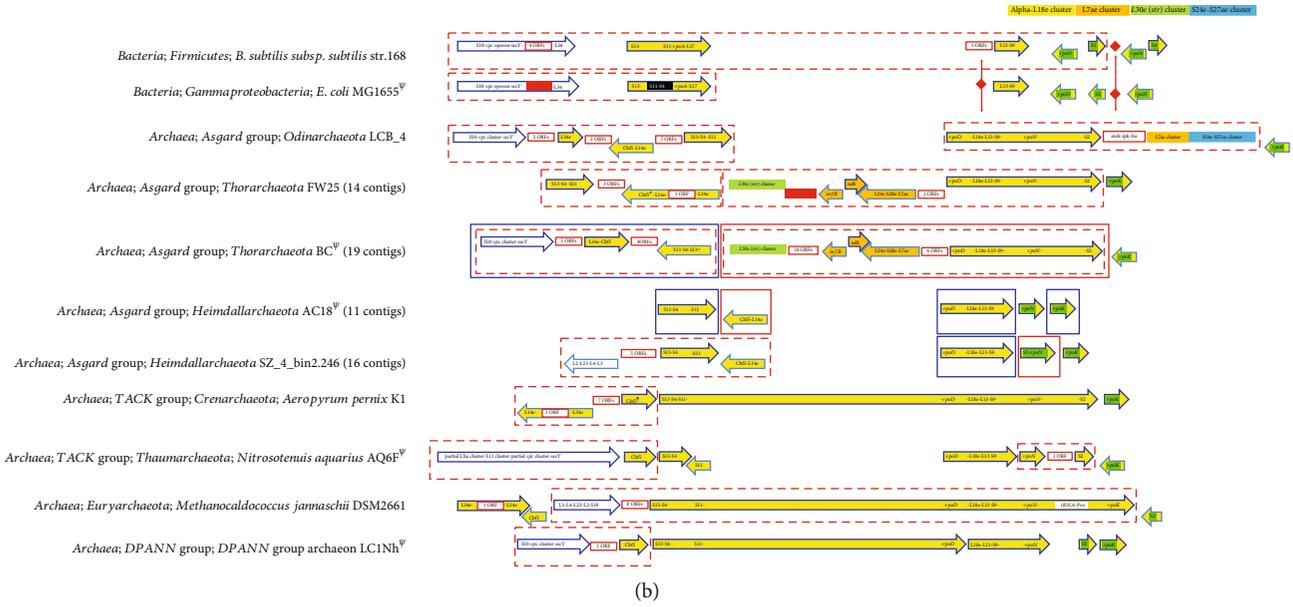
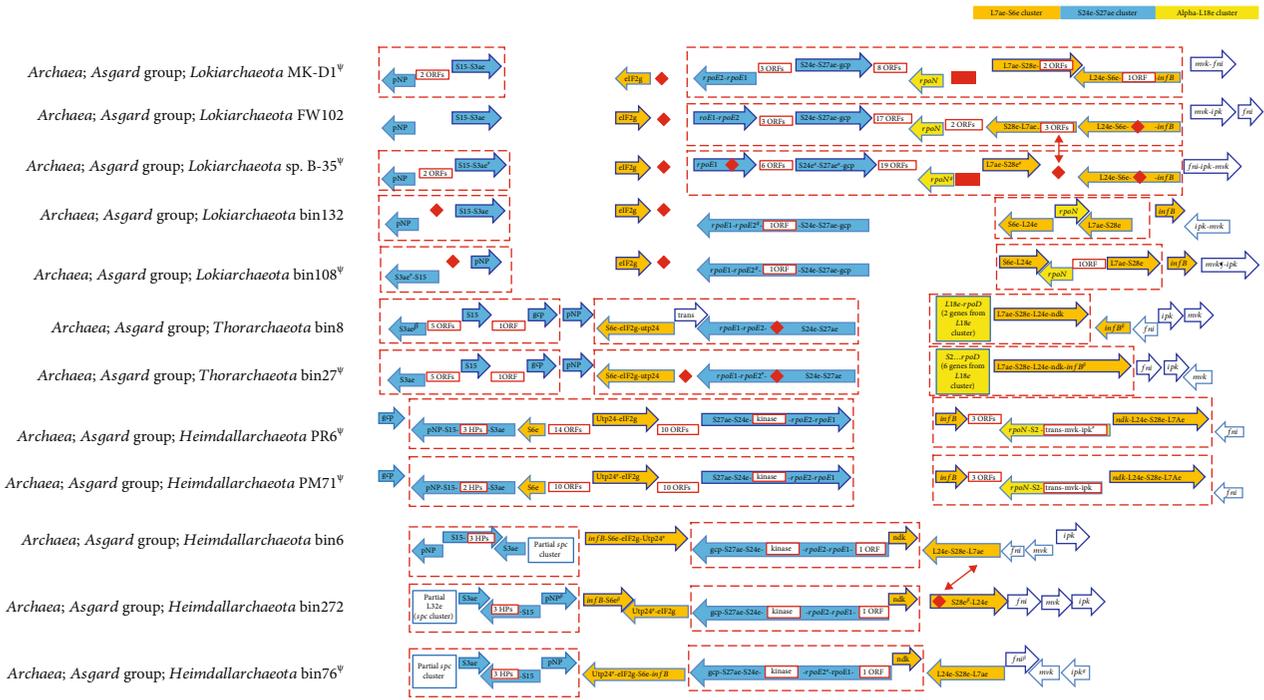


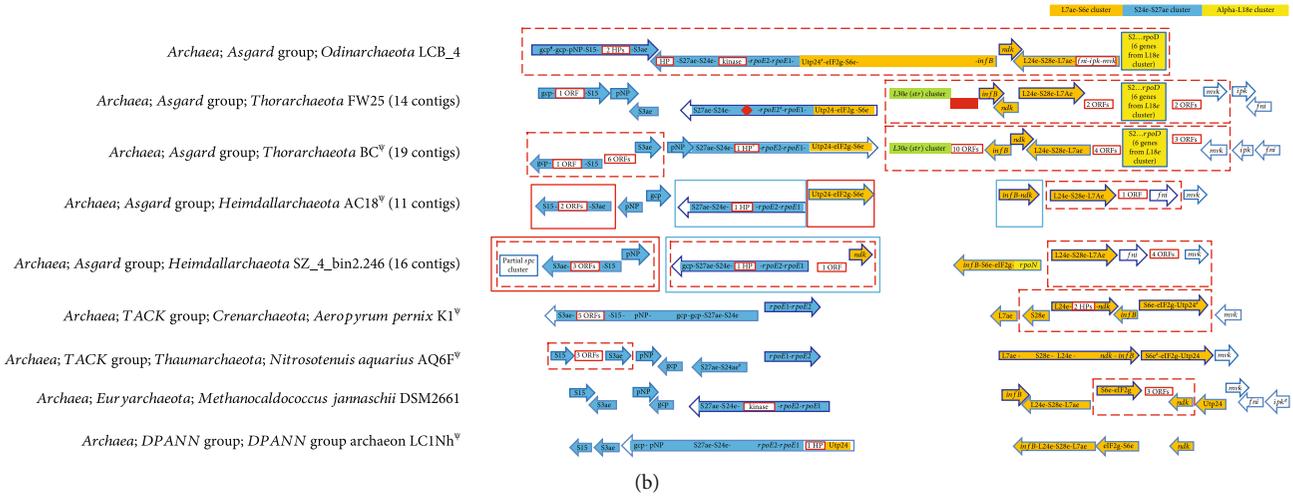
FIGURE 3: (a) <sup>ψ</sup>The orientation of all the genes in the entire cluster flipped for aligning with the other genomes. Adjacent arrows do not indicate their order in the genome; genes within an arrow are contiguous; red diamond and red rectangle represent gene(s) absent in the corresponding genome/location; arrows within dashed red boxes are contiguous; genes in white color (black background) are not in the order as found in the other genomes; genes in green color occur outside the main cluster in that genome; <sup>#</sup>ORF(s) annotated as HP(s); <sup>†</sup>pseudogene. (b) <sup>ψ</sup>The orientation of all the genes in the entire cluster flipped for aligning with the other genomes. Adjacent arrows do not indicate their order on the genome; <sup>^</sup>begins a contig; genes within an arrow are contiguous; red diamond and red rectangle represent gene(s) absent in the corresponding genome/location; arrows within dashed red boxes are contiguous; in Thorarchaeote BC and Heimdallarchaeotes AC18 and SZ\_4\_bin2.246, each colored (blue, green, or red) box represents a different contig; genes in white color (black background) are not in the order as found in the other genomes; <sup>†</sup>pseudogene; genes in green color occur outside the main cluster in that genome; <sup>#</sup>ORF(s) annotated as HP(s).

S24e-S27ae-rpoE1 cluster order in select genomes of *Asgard* and non-*Asgard* Archaea



(a)

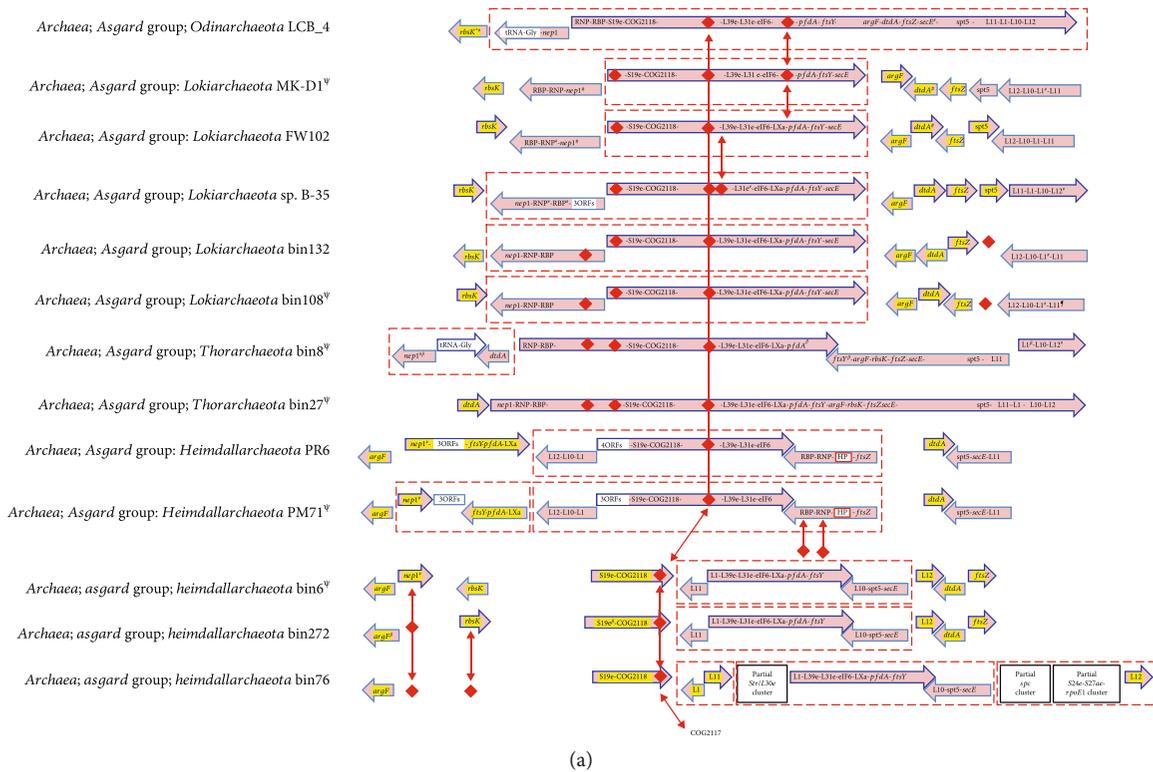
S24e-S27ae-rpoE1 cluster order in select genomes of *Asgard* and non-*Asgard* Archaea



(b)

FIGURE 4: (a) *S24e-S27ae-rpoE1* cluster. <sup>ψ</sup>The orientation of all the genes in the entire cluster flipped for aligning with the other genomes. Adjacent arrows do not indicate their order on the genome except when they are boxed inside dashed red boxes; genes within an arrow are contiguous; red diamond and red rectangle represent gene(s) absent in corresponding genome/location; arrows within dashed red boxes are contiguous. HP: hypothetical protein; pNP: noncanonical purine NTP pyrophosphatase; InfB: translation initiation factor IF-2; ndk: nucleoside diphosphate kinase; Utp24: 30S proteasome protein; <sup>#</sup>ORF(s) annotated as HP(s); gcp: bifunctional tRNA threonylcarbamoyladenosine biosynthesis protein; *fni-iph-mvk*: archaeal lipid pathway genes; <sup>β</sup>partial gene. (b) *S24e-S27ae-rpoE1* cluster. <sup>ψ</sup>The orientation of all the genes in the entire cluster flipped for aligning with the other genomes. Adjacent arrows do not indicate their order on the genome except when they are boxed inside dashed red boxes; genes within an arrow are contiguous; in Heimdallarchaeotes AC18 and SZ\_4\_bin2.246, each colored (blue, green, or red) box represents a different contig. HP: hypothetical protein; pNP: noncanonical purine NTP pyrophosphatase; InfB: translation initiation factor IF-2; ndk: nucleoside diphosphate kinase; Utp24: 30S proteasome protein; <sup>#</sup>ORF(s) annotated as HP(s); gcp: bifunctional tRNA threonylcarbamoyladenosine biosynthesis protein; *fni-iph-mvk*: archaeal lipid pathway genes.

L31e and L12-L10-L11-L1 cluster order in select genomes of Asgard and non-Asgard Archaea



L31e and L12-L10-L11-L1 cluster order in select genomes of Asgard and non-Asgard Archaea

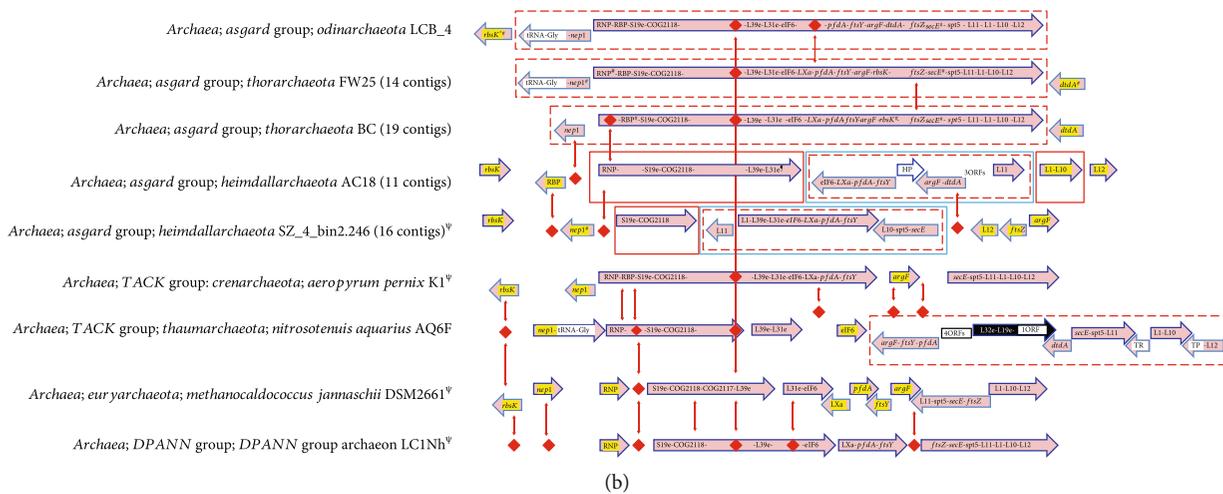


FIGURE 5: (a) *L31e* and *L12-L10-L11-L1* cluster. <sup>ψ</sup>The orientation of all the genes in the entire cluster flipped for aligning with the other genomes. Adjacent arrows do not indicate their order on the genome except when they are boxed inside dashed red boxes; genes within an arrow are contiguous; red diamond represents gene(s) absent in the corresponding genome/location; arrows within dashed red boxes are contiguous; genes/ORFs not colored are not the usual components of this cluster. TP: trimeric intracellular cation-selective channel proteinellular cation-selective channel protein; HP: hypothetical protein; \*gene begins or terminates a contig; RNP: ribonuclease P protein component 4; RBP: RNA binding protein; *nep1*: ribosomal RNA small subunit methyltransferase gene; <sup>#</sup>ORF(s) annotated as HP(s); TR: transcriptional regulator; genes in yellow color occur outside the core/cluster. (b) *L31e* and *L12-L10-L11-L1* cluster. <sup>ψ</sup>The orientation of all the genes in the entire cluster flipped for aligning with the other genomes. Adjacent arrows do not indicate their order on the genome except when they are boxed inside dashed red boxes; genes within an arrow are contiguous; red diamond represents gene(s) absent in the corresponding genome/location; arrows within dashed red boxes are contiguous; genes/ORFs not colored are not the usual components of this cluster; in the two Heimdallarchaeotes, each colored (blue, green, or red) box represents a different contig. TP: trimeric intracellular cation-selective channel protein; HP: hypothetical protein; \*genes begins or terminates a contig; RNP: ribonuclease P protein component 4; RBP: RNA binding protein; *nep1*: ribosomal RNA small subunit methyltransferase gene; <sup>#</sup>ORF(s) annotated as HP(s); genes in white (black background) are part of the big *S10-spc* cluster, but occurring with the *secE-sps5-L11* segment in the Thaumarchaeote *N. aquarius* AQ6f; TR: transcriptional regulator; genes in yellow color occur outside the core/cluster.

TABLE 1: List of Asgard genomes used in this study.

Archaea (Asgard group)	Accession number	No. of contigs	Assembly	GenBank sequence/INSDC
Lokiarchaeota				
MK-D1	GCA_008000775.1	*	ASM800077v1	CP042905.1
<i>H. repetitus</i> FW102	GCA_021498095.1	*	ASM2149809v1	JAIZWK010000001.1
B-35	GCA_025839675.1	*	ASM2583967v1	CP104013.1
bin132	GCA_020343655.1	*	ASM2034365v1	CP070805.1
bin108	GCA_020344955.1	*	ASM2034495v1	CP070831.1
Odinarchaeota LCB_4	GCA_001940665.1	*	ASM194066v1	MDVT00000000.1
Thorarchaeota				
bin27	GCA_020348985.1	*	ASM2034898v1	CP070761.1
bin8	GCA_020355105.1	*	ASM2035510v1	CP070658.1
FW25	GCA_021498125.1	12	ASM2149812v1	JAIZWL01
BC	GCA_008080745.1	19	ASM808074v1	SHMX01
Heimdallarchaeota				
PM71	GCA_021513695.1	*	ASM2151369v1	CP084166.1
PR6	GCA_021513715.1	*	ASM2151371v1	CP084167.1
bin6	GCA_020353515.1	*	ASM2035351v1	CP070695.1
bin76	GCA_020351745.1	*	ASM2035174v1	CP070665.1
bin272	GCA_020348965.1	*	ASM2034896v1	CP070760.1
AC18	GCA_021498085.1	11	ASM2149808v1	JAIZWM01
SZ_4_bin2.246	GCA_011364965.1	16	ASM1136496v1	RDOB01

\*Complete genome.

MKD1, FW102, B-35, bin132, bin108 (Lokiarchaeota), LCB4 (Odinarchaeota), bin27 and bin8 (Thorarchaeota), PM71, PR6, bin6, bin76, and bin272 (Heimdallarchaeota). The next genomes that are of reasonably good quality (assembled into less than 20 contigs) with annotated genes/proteins are those of the Thorarchaeote strains FW25 (twelve contigs) and BC (nineteen contigs) and the Heimdallarchaeote strains AC18 (eleven contigs) and SZ\_4 bin2.246 (sixteen contigs). They will henceforth be referred to as FW25, BC, AC18, and bin2.246, respectively. These genomes provide useful information for some of the clusters of interest but typically not all.

The remaining Asgard genomes each consist of at least twenty contigs and frequently many more. Given the incompleteness of these genomes, the r-protein coding genes in a given cluster were often found in multiple contigs. Some of them, as illustrated in Supplementary Figures 1a-b, either end or begin a contig. The genomes (contigs) of many other genomes have not been annotated into ORFs. Hence, their gene order is not available and therefore was not considered when mapping the clusters. In summary, only the complete genomes and those available with less than twenty contigs, as defined above, were used for the comparative analysis of r-protein gene arrangement. Thus, the choice of non-Asgard archaeal genomes for representation in the figures for comparison is based on the four major groups classified under Archaea [74], viz., the Asgard, the DPANN, the TACK, and the Euryarchaeota [2, 7, 34, 36, 40, 41, 75–78]. Furthermore, the phylogenetic relatedness as described in the literature

was also taken into account [7, 34, 37, 39, 41, 79]. *E. coli* and *B. subtilis* were used as representatives of Bacteria.

**3.1. Big Cluster Comprising Genes from S10 and spc.** A core of 14 cooccurring genes (*uL22-uS3-uL29-uS17-uL14-uL24-uL5-uS14-uS8-uL6-uL18-uS5-uL30-uL15*) belonging to the segment of the *S10-spc* cluster is characterized and established as operons in *E. coli* and other Bacteria. In the archaeal (both Asgard and non-Asgard) genomes, this cluster core has 4 additional Archaea-Eukarya-specific genes (*RNP1* (ribonuclease protein 1), *S4e*, *L32e*, and *L19e*) in Asgard and non-Asgard genomes (Figures 1(a) and 1(b)). The main core (of 14 genes) was identified as conserved in gene order, arrangement, and genome location in Asgard Archaea, non-Asgard Archaea, and Bacteria (Figure 1). The one exception is in the Thaumarchaeote *N. aquarius* Aq6f (Figure 1(b)). In this genome, a part of this cluster, viz., the segments *L19e-L32e* and *uL18-uS5-uL15*, occurs in the immediate neighborhood of the *L11-L1-L10-L12* cluster.

In Lokiarchaeote sp. B-35, six genes of this cluster (*uS3*, *RNP1*, *uL24*, *uS14*, *L32e*, and *uL15*) are incorrectly annotated as hypothetical proteins. Nevertheless, this core occurs as a complete set in the complete genomes of Asgard Archaea (listed in Table 1) except four Heimdallarchaeote genomes (bin6, bin272, bin76, and bin2.246) (Figures 1(a) and 1(b)).

In the Thorarchaeotes bin8, bin27, FW25, and BC and the Odinarchaeote LCB, a gene for a heat shock protein occurs as an addition to this cluster (Figures 1(a) and 1(b)). A domain analysis using the Conserved Domain Database (CDD) [80]

showed that this protein belongs to the family of HSP20 (small heat shock protein IbpA) family (COG0071) involved in post-translational modification, protein turnover, and chaperones. Acquisition of this gene in these Asgard Archaea could be an unusual instance of horizontal gene transfer occurring within a highly conserved cluster. Additionally, in Thorarchaeotes bin8 and bin27, *uL18* is annotated as a pseudogene.

In the Heimdallarchaeote bin6, the cluster is split into three segments found on different parts of the complete genome. Ten of these genes (*uL22..uS14*) (segment 1) are in one contiguous section, while *uS8-uL6* (segment 2) occurs separately in a second segment. Six genes (*L32e-L19e-uL18-uS5-uL30-uL15*) (segment 3) are in another contiguous section. In bin272, segments 1 and 2 are similarly found separately on the genome. However, only a partial version of the gene *L32e* from segment 3 is present, while the rest of segment 3 is missing. In bin76, segment 1 is missing in entirety, while segments 2 and 3 are present. Additionally, in Lokiarchaeote bin108 and Heimdallarchaeotes bin76 and bin272, 10%, 15%, and 21% of the genes are partial. In the case of the Asgard genomes, which are only represented as contigs, the core occurs as a contiguous set on one contig in FW25, BC, and AC18. Only in the case of Heimdallarchaeote SZ\_4\_bin2.246 (sixteen contigs), segment 1 is on one contig, while segments 2 and 3 are both at different locations on another contig. This pattern is observed even in the Asgard genomes that consist of more than twenty contigs. In these genomes, the *S10-spc* r-protein cluster organization, despite being found in multiple contigs, is considered indicative of cooccurrence, if the genomes were complete (Supplementary figure 1). In all the non-Asgard Archaea examined, the entire cluster occurs as one contiguous set of genes (with the exception of the Thaumarchaeote *N. aquarius* AQ6f).

**3.2. Small Cluster of Genes Normally Found in the *S10* Operon.** A second smaller cluster comprising the homologs of the most conserved genes of the bacterial *S10* operon/cluster, namely, *uS10-uL3-uL4-uL23-uL2*, occurs independently on the Asgard genomes, separate from the rest. This feature is shared by some non-Asgard Archaea as well (Figures 1(a) and 1(b)). Part of the small cluster, *uL3-uL4-uL23*, always cooccurs. In contrast, *uS10* is always found to occur independently from this cluster in both the Asgard and non-Asgard Archaea. In bin8, bin27, FW25, and BC (Thorarchaeotes) and PR6, PM71, bin6, bin272, AC18, and SZ\_4\_bin2.246 (Heimdallarchaeotes), *uL3-uL4-uL23-uL2* occurs as one whole segment. In MKD1, FW102, and B-35 (Lokiarchaeotes), *uL2* occurs independently, separate from the remainder of the small cluster. However, *uL3-uL4-uL23* is entirely missing in bin132 and bin108 (Lokiarchaeotes).

In the bacterial genomes, *E. coli* and *B. subtilis*, the *S10* and *spc* operons (clusters) cooccur as one contiguous cluster of twenty-one genes, including this smaller cluster of five genes. This is then followed by the smaller cluster of *bL36-uS13-uS11-uS4-rpoA-bL17*, which is part of another cluster, namely, the Alpha operon/*L18e* cluster, described below. Thus, in the bacterial genomes, the *S10-spc* operon/cluster is contiguous

with a part of the *S13...S9* cluster. *L18e* being Archaea-Eukarya-specific is not found in this segment.

**3.3. *Str/L30e* Cluster.** The core of this cluster is comprised of nine contiguous genes, which is observed to cooccur with the *L11* cluster in *B. subtilis* (Bacteria). In the *E. coli* genome, the cluster is split into two sections. One section comprising *rpoB-rpoC* cooccurs with the *L11* cluster, while the other section is *S12..tufA*. Each section is found in different parts of the genome (Figures 2(a) and 2(b)). The equivalent of the *str/L30e* cluster in non-Asgard Archaea (*A. pernix* K1 (TACK group; Crenarchaeota), *M. jannaschii* DSM2661 (Euryarchaeota)) and the DPANN group Archaea (Figure 2) has 9-10 genes, viz., *rpoH-rpoB-rpoA-L30e-nusA-S12-S7-(fusA)-EF(1a)-S10*. In the Thaumarchaeote *N. aquarius* AQ6f and DPANN group archaeon LC1Nh, *EF(1a)-S10fusA* occurs independent of this section. In the genomes of the Lokiarchaeotes FW102, B-35, bin132, bin108, and bin8; the Thorarchaeotes FW25 and BC; the Odinararchaeote LCB; and the Heimdallarchaeotes (bin6, bin272, bin76, and bin2.246), a part of this cluster, viz., *rpoA1-rpoA2-L30e-nusA-S12*, occurs as a conserved core. In the other Asgard genomes, this core is split into two parts, comprised of *rpoA1-rpoA2-L30e-nusA* (five genes) and *S12-S7* (two genes), respectively, which are found on different regions of the genome(s). Notable are the complete genomes of the Heimdallarchaeotes PR6 and PM71, in which the entire cluster is dispersed with no conservation of the gene order (Figure 2(a)). Additionally, in the Thorarchaeote genomes FW 25 and BC, the *str/L30e* cluster cooccurs with a portion each of the *L7ae* (a new cluster described below) and the *L18e* clusters. Despite these two genomes occurring in fourteen and nineteen contigs, respectively, the cluster is fairly contiguous in one contig. The arrangement of the *str/L30e* cluster in the Asgard Thorarchaeota most closely resembles the arrangement found in the non-Asgard Archaea.

**3.4. Alpha Operon/*L18e* Cluster.** This cluster is an operon (*S13-S11-S4-rpoA-L17*) in bacterial genomes and is found immediately downstream of the *S10-spc* operon (Figure 3). In the non-Asgard Archaea, *A. pernix* K1 (TACK group) and *M. jannaschii* DSM 2661 (Euryarchaeota), this cluster is comprised of a contiguous core of eight genes, viz., *S13-S4-S11-rpoD-L18e-L13-S9-rpoN* (Figure 3(b)). Only in *M. jannaschii* DSM 2661, this cluster is in the genomic neighborhood of the *uL3-uL4-uL23-uL2* section of the *S10* cluster.

In the Asgard genomes, the *L18e* cluster is split into two main sections. The first is comprised of *S13-S4-S11*, which, in five Asgard genomes, is found cooccurring with *L34e-L14e-Cbf5* in the genomic neighborhood of the big *S10-spc* cluster (Figures 3(a) and 3(b)). *Cbf5* encodes an RNA-guided pseudouridylation complex pseudouridine synthase subunit. In three Asgard genomes, the *S13-S4-S11* group is in the immediate genomic neighborhood of *uL3-uL4-uL23-uL2*.

The second section is comprised of *rpoD-L18e-L13-S9* (Figures 3(a) and 3(b)) and is found independent of *S13-S4-S11* in the genomes of Lokiarchaeota, Heimdallarchaeota, and three Thorarchaeotes. The gene for the DNA-directed RNA polymerase subunit N (*rpoN*), when not found in the

main cluster, is often associated with the *L7ae* cluster. In two Asgard Thorarchaeote genomes, part of the *L18e* cluster cooccurs with the *str/L30e* cluster. In all the Asgard Thorarchaeote genomes examined, a partial portion of the *L7ae* cluster cooccurs with either a part of or the whole of the *rpoD-L18e-L13-S9* segment.

**3.5. *S24e-S27ae* Cluster.** This cluster has a core of eight genes (*gcp-pNP-S15-S3ae-S27ae-S24e-rpoE2-rpoE1*). It cooccurs and is contiguous with the *L7ae* and the *Alpha-L18e* clusters in the Asgard Odinararchaeota (Figure 4(b)). *gcp* and *pNP* encode a bifunctional tRNA threonylcarbamoyladenine biosynthesis protein and a noncanonical purine NTP pyrophosphatase, respectively.

Based on how it is found in the rest of the archaeal genomes, the *S24e* cluster can be split into two sections, one containing the three genes, *pNP-S15-S3ae*, and the other with five genes, *rpoE1-rpoE2-S24e-S27ae-gcp*. In all five complete Lokiarchaeote genomes, the *pNP-S15-S3ae* section occurs independently. In three of the Lokiarchaeote genomes, the second segment cooccurs with a part of the *L7ae* cluster. A section comprising *S27ae-S24e-rpoE2-rpoE1* cluster cooccurs with the *S6e-eIF2g-utp24* section of *L7ae* cluster in all the Asgard Thorarchaeota examined. In the two complete Asgard Heimdallarchaeota (PR6 and PM71), the two sections of the *S24e* cluster “sandwich” a small segment of the *L7ae* cluster, viz., *S6e* and *Utp24-eIF2g*. In four Heimdallarchaeotes, the *pNP-S15-S3ae* section is found in the immediate neighborhood of one section of the *spc* cluster. Amongst the non-Asgard Archaea, the *S24e-S27ae-rpoE1* cluster is contiguous with the *L7ae* cluster, the *Alpha-L18e*, and the *str/L30e* clusters in the Desulfurococcales (Crenarchaeota) Archaea, namely, *D. amylolyticus* 1221n and *S. hellenicus* DSM 12710 (Supplementary Figure 2).

**3.6. Identification of a New Cluster of r-protein Coding Genes.** A new cluster *L7ae* comprised of eight genes (*L7ae-S28e-L24e-ndk-infB-S6e-eIF2g-Utp24*) was identified in the Asgard Odinararchaeota (Figure 4). The genes *ndk*, *infB*, *eIF2g*, and *Utp24* encode a nucleoside diphosphate kinase, a translation initiation factor IF-2, the gamma subunit of translation initiation factor 2, and a 30S proteasome protein, respectively. In the Odinararchaeote genome, this cluster cooccurs with the entire *S24e-S27ae* cluster, in the following order: *S24e-S27ae* cluster, *L7ae* cluster, archaeal mevalonate pathway genes [81] and a partial *L18e* cluster. Such a contiguous arrangement of multiple clusters was also found in the genomes of *D. amylolyticus* 1221n and *S. hellenicus* DSM 12710 (Desulfurococcales (Crenarchaeota)) (Supplementary Figure 2). In three Lokiarchaeote genomes, *L7ae.infB* (5 genes, since *ndk* could not be found in these genomes using a BLAST search) cooccur with a portion of the *S24e-S27ae* cluster. In the two complete genomes of Heimdallarchaeotes PR6 and PM71, 3 genes, *S6e-eIF2g-utp24*, are sandwiched in between two parts of the *S24e-S27ae* cluster (Figure 4). In the two complete Asgard Thorarchaeote genomes, this 3-gene section *S6e-eIF2g-utp24* cooccurs with a five-gene segment of the *S24e-S27ae* cluster. In the other two Thorarchaeote genomes, a partial

portion (5 genes) of this new cluster is found sandwiched between the *str/L30e* and a partial section (6 genes) of the *L18e* clusters.

**3.7. *L31e* and *L11-L1-L10-L12* Clusters.** The genes for the Archaea unique r-proteins, namely, L39e, L31e, S19e, and LXa, have been described to be part of a cluster that includes the conserved segment of 4 genes, viz., *S19e-COG2118-COG2117-L39e* [20]. This 4-gene conserved segment occurs as such only in the non-Asgard Archaea *M. jannaschii*. COG2117 is missing in all the other non-Asgard and Asgard Archaea examined (Figure 5). In the Asgard genomes of Odinararchaeota LCB4 and the three Thorarchaeote strains, this segment is part of a larger contiguous segment that includes the four gene *L11-L1-L10-L12* cluster, which is described as an operon [82, 83]. The genes in this cluster include those encoding RNP (ribonuclease P protein component 4), RBP (an RNA binding protein), Nep1 (ribosomal RNA small subunit methyltransferase), Spt5 (transcription elongation factor), SecE (gamma subunit of the protein translocase SEC61 complex), RbsK (a carbohydrate kinase), FtsZ (a cell division protein), FtsY (a signal recognition particle-docking protein), and DtdA (D-tyrosyl-tRNA(Tyr) deacylase).

In the Asgard genomes examined in this study, except four Asgard genomes, all the other genomes show a conserved 5-gene core cluster, *S19e-COG2118-L39e-L31e* (Figures 5(a) and 5(b)). The remaining genes of the *L31e* and *L11-L1-L10-L12* clusters show significant rearrangements in the Asgard Archaea. Overall, it can be divided into four tentative segments, viz., *nep1-RNP-RBP-S19e-COG2118-L39e-L31e-eIF6*, *LXa-pfdA-ftsY*, *ftsZ-secE-spt5*, and *L11-L1-L10-L12*.

In the Asgard Lokiarchaeotes MKD1 and FW102, part of the large segment, *S19e.eIF6*, is contiguous with *LXa...secE*, and in B-35, bin132, and bin108, *nep1...secE* occurs as a whole unit. The *L11-L1-L10-L12* segment is on a different location in these genomes. In all the Asgard Heimdallarchaeotes examined, *L11* is split from its operon/cluster (*L11-L1-L10-L12*) and found in other parts of the cluster. In the two complete Heimdallarchaeotes PR6 and PM71, the remainder of the operon *L12-L10-L1* while on the negative strand is contiguous with the section containing *S19e.eIF6* on the positive strand, followed by *RBP.ftsZ* on the negative strand. The three-gene segment *ftsY-pfdA-LXa* occurs independently. This pattern of rearrangement is identical in these two Asgard Heimdallarchaeotes, with the only difference being the strand orientation of all these segments is the opposite relative to each other. In the incomplete Heimdallarchaeote bin2.246 genome (16 contigs) and three Heimdallarchaeote genomes (bin6, bin272, and bin76), parts of these segments occur as *L11*, contiguous with *L39e-L31e-eIF6-LXa-ftsY*, followed by *L10-spt5-secE*. These three genomes also show truncated versions of the *S10-spc* cluster (Figure 1).

Additionally, an annotation error was observed in the Asgard Thorarchaeote BC genome. A hypothetical protein coding gene (GenBank accession number: TXT56565.1) is downstream of *rbsK* (GenBank accession number: TXT56564.1, which is also incorrectly annotated as a hypothetical protein). This gene is likely an artifact of erroneous annotation, because the coordinates of the same (gene for TXT56565.1: 125555-125674) overlap the coordinates of the gene *ftsZ* immediately

downstream by 28 bases (gene for TXT56566.1: 125646-126779). Overall, the *S19e-COG2118-L39e* order is conserved in all the genomes with the exception of the non-Asgard Archaea Thaumarchaeota (Figure 5).

#### 4. Discussion

Classical operons are found to occur in both Bacteria and Archaea [84, 85]. The conservation of the arrangement of the r-protein gene cluster is in congruence with how they are conserved in sequence [21, 22, 55]. A large chunk of 14 genes from the most highly conserved *S10-spc* cluster of r-protein coding genes was thus found to be similarly conserved in arrangement, whether it was Bacteria, Asgard, or non-Asgard Archaea. The arrangement of the four r-protein coding gene clusters as observed in the 17 Asgard archaeal genomes (Lokiarchaeota, Odinararchaeota, Heimdallarchaeota, and Thorarchaeota) suggests that apart from the big cluster (*S10-spc*), the arrangement/order of the rest of the r-protein clusters is not conserved and is dispersed. In the context of one section of the *Alpha* operon found in the vicinity of the *S10-spc* cluster, the conservation of the location resembles what is observed in bacterial genomes. Assuming these Asgard genomes have been correctly assembled, a portion of the *Alpha* operon (*L18e*) cluster associated with the highly conserved and universal *S10-spc* cluster is likely a reflection of their association despite the phylogenetic distance between Bacteria and Asgard Archaea.

It has been suggested that the Heimdallarchaeota group is the most likely sister group of Eukarya [7, 36, 40, 43, 44]. The dispersive nature of the *str/L30e* cluster order and the unique arrangement of the *S24e-S27ae*, the *L7ae*, the *L31e*, and the *L12-L10-L11-L1* clusters are particularly evident in the case of the Heimdallarchaeote genomes. The unique arrangement of the r-protein clusters, as observed in the case of the Heimdallarchaeota members of the Asgard in this study, may corroborate this. This could possibly be a marker for evolutionary divergence within the Asgard genomes. Additional complete Asgard genomes, when available, may help understand such divergences better. Examination of the plastid genome of *A. thaliana* and the gene order of (*S10-spc* cluster)-*L36-S11* shows similarities with the order in Bacteria (*E. coli*, *B. subtilis*), though a section of this cluster is missing from the same (data not shown).

The dispersion of gene clusters in eukaryotes has been suggested to be directly correlated with higher rates of chromosomal/genome evolution [26, 86]. It remains to be seen if Asgard Archaea show a higher rate of genome evolution as compared to the non-Asgard Archaea.

Genome organization seldom remains in stasis with the phenomenon of lateral gene transfer constantly changing the dynamics of the same. It has been hypothesized that the Asgard proteome, particularly the eukaryotic signature protein coding genes, evolved through extensive horizontal gene transfer [36]. It is not clear from the few available complete Asgard genomes if the arrangement of the r-protein clusters is a result of this phenomenon. Nevertheless, the roles of gene

order or arrangement in fashioning cellular physiology and genome evolution remain to be explored further [87].

It should be noted that the arrangement of the r-protein coding gene clusters in the Asgard could be in some cases an artifact of poor genome completion. The occurrence of several partial genes, a universal gene annotated as a pseudogene, and several universal gene segments missing in the Asgard genomes in this study (Figures 1–5) suggests that the genomes in which they are found may have been incorrectly assembled. The noncontiguous/disjointed occurrence of the otherwise conserved segment of 14 cooccurring genes of the *S10-spc* cluster in the three Heimdallarchaeotes (bin6, bin272, and bin76) (Figure 1(a)) and the absence of three of the most universal and highly conserved genes *uL3-uL4-uL23* from the genomes of Lokiarchaeotes bin132 and bin108 (Figure 1) suggest that the assembly of these genomes is likely not accurate. The incompleteness or misassembly of genomic contigs has been observed previously [88–92]. In fact, binning errors were implicated in the absence of r-protein genes from metagenome-assembled genomes (MAGs), and hence, analysis of genomics and phylogeny of uncultivated microbes, using MAGs, may be fraught with erroneous interpretations [93, 94].

However, we did observe one interesting feature in a non-Asgard archaeal MAG. Of the three non-Asgard archaeal genomes assembled from metagenomes that are available in the NCBI database [88], only the genome of DPANN archaeon GW2011\_AR10 (GenBank accession: CP010424) is complete as a single scaffold/chromosome. Examining the order/arrangement of the r-protein gene clusters in this genome showed that four of the five clusters reported in this paper cooccur contiguously. The arrangement of these clusters in this genome is as follows: *str-L30e* cluster (*rpoH-rpoB-HP-rpoA1-rpoA2-L30e-nusA-S12-S7-fusA-EF(1a)-S10*)-*5HPs-S10-spc* cluster (*L3-L4-L23-L2-S19-L22-S3-L29-IF(Sui)-RNP1-S17-L14-L24-S4e-L5-S14-S8-L6-L32e-L19e-L18-S5-L30-L15-secY*)-*HP-L34e-L14e-L18e (Alpha operon)* cluster (*S13-S4-S11-rpoD-tRNA(leu)-L18e-L13-S9-rpoN-S2*)-*HP-gltX-HP-L7ae* cluster (*L7ae-S28e-L24e-ndk-dUTP diphosphatase-infB-S6e-eIF2g-Utp24-rpoE1-rpoE2-S24e-S27ae-10 ORFs-gcp-S15-HP-S3ae*). The *L31e* and *L11-L10-L12* clusters are not part of this section and occur separately as two distinct sections elsewhere in the genome. Only in *B. subtilis* (Bacteria), a partially similar feature is observed with the *L11* cluster found immediately upstream of the *S10-spc* cluster.

The second non-Asgard MAG, that of DPANN *Candidatus Micrarchaeum acidiphilum* ARMAN-2 (GenBank assembly: GCA\_009387755.1) is in 8 contigs, with the r-protein gene clusters split and not conserved as reported for some of the Asgard genomes in the paper. The third non-Asgard MAG, that of DPANN *Candidatus Forterreia multitransposorum* archaeon (GenBank accession no.: CP045477), has not been annotated into ORFs (open reading frames)/genes etc. Thus, these sequences derived from metagenome-assembled genomes may not be sound.

It is not clear if the cooccurrence of four r-protein clusters in the MAG of DPANN archaeon GW2011\_AR10 is a result of the accuracy of the assembly and annotation/curation of the genome. If it is indeed an accurate assemble, then

the assembly of other genomes in the database may need to be revisited for improving better quality which may help in an improved understanding of arrangement of these clusters. Thus, in the absence of good quality genomes, ascertaining the order of highly conserved genes of r-proteins can be challenging, and thus, well-curated genomes are vital towards understanding genome organization. The importance of a more robust curation of MAGS as well publicly available Asgard genomes [6, 95] cannot be emphasized enough to get a better picture of the genome organization of the same.

## 5. Conclusions

The discovery of the Asgard Archaea has redefined our understanding of the phylogenetic positions of the various groups that comprise the Archaea. The genomic arrangement of the r-protein coding genes in the Asgard superphylum within the domain Archaea suggests that the order may follow the phylogeny. The closer an Asgard is to eukaryotes, the more dispersed is the arrangement of the r-protein gene clusters. This feature suggests a possible ancient gain or loss of the operon strategy of gene regulation in the early history of life. Given that very few Asgard genomes are currently complete, new data and better means of genome assembly will facilitate a better understanding of genome order in the Asgard Archaea in the future.

## Abbreviations

MKD1:	Lokiarchaeote strain <i>Candidatus Prometheoarchaeum syntrophicum</i> MK-D1
FW102:	Lokiarchaeote strain <i>Harpocratesius repetitus</i> FW102
B-35:	Lokiarchaeote strain B-35
bin132:	Lokiarchaeote strain bin132
bin108:	Lokiarchaeote strain bin108
LCB4:	Odinarchaeote strain LCB_4
bin27:	Thorarchaeote bin27
bin8:	Thorarchaeote bin8
FW25:	Thorarchaeote strain FW25
BC:	Thorarchaeote strain BC
PM71:	Heimdallarchaeote strain PM71
PR6:	Heimdallarchaeote strain PR6
bin6:	Heimdallarchaeote bin6
bin76:	Heimdallarchaeote bin76
bin272:	Heimdallarchaeote bin272
AC18:	Heimdallarchaeote AC18
bin2.246:	Heimdallarchaeote SZ_4_bin2.246
LUCA:	Last universal common ancestor
LSU:	Large subunit
SSU:	Small subunit
r-protein:	Ribosomal protein
mtDNA:	Mitochondrial DNA
NCBI:	National Center for Biotechnology Information
rRNA:	Ribosomal RNA
MAGs:	Metagenome-assembled genomes
IMG/M:	Integrated Microbial Genomes and Microbiomes
DOE:	Department of Energy
JGI:	Joint Genome Institute.

## Data Availability

The datasets used and analysed within the current study are available from the NCBI website as referenced in the paper.

## Disclosure

A preprint has previously been published [96].

## Conflicts of Interest

All authors declare they have no competing interests. LAK and ELS were student volunteers on this work. LAK was a fresh graduate from Clements High School, and ELS had just completed his 10th grade at Obra D. Tompkins High School, who volunteered with the group of Dr. George E. Fox at the University of Houston.

## Authors' Contributions

MRT and GEF conceived and designed the study. MRT obtained the sequences. MRT, RVS, LAK, and ELS compared the organization of the ribosomal protein coding genes in the genomes. RVS, LAK, and ELS prepared the comparative chart for the gene clusters. MRT prepared the figures. MRT and GEF prepared the manuscript paper that was finalized with the help from all the authors. All authors read and approved the final manuscript.

## Acknowledgments

This work was supported in part by NASA (Contract 80NSSC18K1139) under the Center for Origin of Life, Georgia Institute of Technology, to George E. Fox and partly by the NSF (NSF-MCB-EAGER 2227347) to Madhan R. Tirumalai.

## Supplementary Materials

Supplementary Figure 1: (a) the *S10-spc* cluster distributed in multiple contigs in Asgard Archaea. In *Odinarchaeota* LCB\_4, arrows represent different locations on a single contig; adjacent arrows do not indicate their order on the genome; genes within an arrow are contiguous; the direction of the arrow denotes the strand orientation. \*Begins or ends a contig. Hsp20: heat shock protein 20 (a, partial gene; d, pseudogene). Red diamond represents gene(s) absent in the corresponding genome/location; arrows within dashed red boxes are contiguous. HP: hypothetical protein; RNP1: ribonuclease P protein component 1; NCG: noncluster genes. #ORF(s) annotated as HP(s). Hsp20: heat shock protein 20. Blank box represents the entire gene set missing in corresponding genome location. (b) The *S10-spc* cluster distributed in multiple contigs in Asgard Archaea. Adjacent arrows do not indicate their order on the genome; genes within an arrow are contiguous; the direction of the arrow denotes the strand orientation. \*Begins or ends a contig. Hsp20: heat shock protein 20 (a, partial gene; d, pseudogene). Red diamond represents gene(s) absent in the corresponding genome/location; arrows within dashed red boxes

are contiguous. HP: hypothetical protein; RNPI: ribonuclease P protein component 1; NCG: noncluster gene(s). #ORF(s) annotated as HP(s). Hsp20: heat shock protein 20. Blank box represents the entire gene set missing in corresponding genome location. Supplementary Figure 2: the *S24e-S27ae-rpoE1* cluster is contiguous with *L7ae...Utp24*, the *Alpha-L18e*, and the *str/L30e* clusters in the Desulfurococcales (Crenarchaeota) Archaea, namely, *D. amylolyticus* 1221n and *S. hellenicus* DSM 12710. Genes within an arrow are contiguous; red diamond represents gene(s) absent in the corresponding genome/location. ORFs: open reading frames. Arrows within dashed red boxes are contiguous. HP: hypothetical protein. (*Supplementary Materials*)

## References

- [1] C. R. Woese and G. E. Fox, "Phylogenetic structure of the prokaryotic domain: the primary kingdoms," *Proceedings of the National Academy of Sciences*, vol. 74, no. 11, pp. 5088–5090, 1977.
- [2] B. J. Baker, V. De Anda, K. W. Seitz, N. Dombrowski, A. E. Santoro, and K. G. Lloyd, "Diversity, ecology and evolution of Archaea," *Nature Microbiology*, vol. 5, no. 7, pp. 887–900, 2020.
- [3] P. Forterre, "Archaea: a goldmine for molecular biologists and evolutionists," in *Archaea. Methods in Molecular Biology*, vol. 2522, pp. 1–21, Humana, New York, NY, 2022.
- [4] N. O. Medina-Chávez and M. Trivisano, "Archaeal communities: the microbial phylogenomic frontier," *Frontiers in Genetics*, vol. 12, 2022.
- [5] C. J. Castelle and J. F. Banfield, "Major new microbial groups expand diversity and alter our understanding of the tree of life," *Cell*, vol. 172, no. 6, pp. 1181–1197, 2018.
- [6] F. MacLeod, G. S. Kindler, H. L. Wong, R. Chen, and B. P. Burns, "Asgard Archaea: diversity, function, and evolutionary implications in a range of microbiomes," *AIMS Microbiology*, vol. 5, no. 1, pp. 48–61, 2019.
- [7] K. Zaremba-Niedzwiedzka, E. F. Caceres, J. H. Saw et al., "Asgard archaea illuminate the origin of eukaryotic cellular complexity," *Nature*, vol. 541, no. 7637, pp. 353–358, 2017.
- [8] N. Ban, R. Beckmann, J. H. Cate et al., "A new system for naming ribosomal proteins," *Current Opinion in Structural Biology*, vol. 24, pp. 165–169, 2014.
- [9] D. J. Klein, P. B. Moore, and T. A. Steitz, "The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit," *Journal of Molecular Biology*, vol. 340, no. 1, pp. 141–177, 2004.
- [10] N. A. Kovacs, A. S. Petrov, K. A. Lanier, and L. D. Williams, "Frozen in time: the history of proteins," *Molecular Biology and Evolution*, vol. 34, no. 5, pp. 1252–1260, 2017.
- [11] O. Lecompte, R. Ripp, J. C. Thierry, D. Moras, and O. Poch, "Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale," *Nucleic Acids Research*, vol. 30, no. 24, pp. 5382–5390, 2002.
- [12] G. E. Fox, "The evolutionary history of the ribosome," in *The Genetic Code and the Origin of Life*, R. d. P. L., Ed., pp. 92–105, Landes Bioscience, 2004.
- [13] C. Hsiao, T. K. Lenz, J. K. Peters et al., "Molecular paleontology: a biochemical model of the ancestral ribosome," *Nucleic Acids Research*, vol. 41, no. 5, pp. 3373–3385, 2013.
- [14] M. Root-Bernstein and R. Root-Bernstein, "The ribosome as a missing link in the evolution of life," *Journal of Theoretical Biology*, vol. 367, pp. 130–158, 2015.
- [15] R. Root-Bernstein and M. Root-Bernstein, "The ribosome as a missing link in prebiotic evolution III: over-representation of tRNA- and rRNA-like sequences and plieofunctionality of ribosome-related molecules argues for the evolution of primitive genomes from ribosomal RNA modules," *International Journal of Molecular Sciences*, vol. 20, no. 1, p. 140, 2019.
- [16] P. Vishwanath, P. Favaretto, H. Hartman, S. C. Mohr, and T. F. Smith, "Ribosomal protein-sequence block structure suggests complex prokaryotic evolution with implications for the origin of eukaryotes," *Molecular Phylogenetics and Evolution*, vol. 33, no. 3, pp. 615–625, 2004.
- [17] H. Hartman, P. Favaretto, and T. F. Smith, "The archaeal origins of the eukaryotic translational system," *Archaea*, vol. 2, Article ID 431618, 9 pages, 2006.
- [18] E. Schmitt, P. D. Coureux, R. Kazan, G. Bourgeois, C. Lazenec-Schurdevin, and Y. Mechulam, "Recent advances in archaeal translation initiation," *Frontiers in Microbiology*, vol. 11, article 584152, 2020.
- [19] T. Coenye and P. Vandamme, "Organisation of the S10, spc and alpha ribosomal protein gene clusters in prokaryotic genomes," *FEMS Microbiology Letters*, vol. 242, no. 1, pp. 117–126, 2005.
- [20] J. Wang, I. Dasgupta, and G. E. Fox, "Many nonuniversal archaeal ribosomal proteins are found in conserved gene clusters," *Archaea*, vol. 2, pp. 241–251, 2009.
- [21] J. L. Siefert, K. A. Martin, F. Abdi, W. R. Widger, and G. E. Fox, "Conserved gene clusters in bacterial genomes provide further support for the primacy of RNA," *Journal of Molecular Evolution*, vol. 45, no. 5, pp. 467–472, 1997.
- [22] M. R. Tirumalai, D. Anane-Bediakoh, S. Rajesh, and G. E. Fox, "Net charges of the ribosomal proteins of the S10 and spc clusters of halophiles are inversely related to the degree of halotolerance," *Microbiology Spectrum*, vol. 9, no. 3, article e0178221, 2021.
- [23] A. M. Hauth, U. G. Maier, B. F. Lang, and G. Burger, "The Rhodomonas salina mitochondrial genome: bacteria-like operons, compact gene arrangement and complex repeat region," *Nucleic Acids Research*, vol. 33, no. 14, pp. 4433–4442, 2005.
- [24] Y. Ben-Shahar, K. Nannapaneni, T. L. Casavant, T. E. Scheetz, and M. J. Welsh, "Eukaryotic operon-like transcription of functionally related genes in *Drosophila*," *Proceedings of the National Academy of Sciences*, vol. 104, no. 1, pp. 222–227, 2007.
- [25] P. Ganot, T. Kallesøe, R. Reinhardt, D. Chourrout, and E. M. Thompson, "Spliced-leader RNA trans splicing in a chordate, *Oikopleura dioica*, with a compact genome," *Molecular and Cellular Biology*, vol. 24, pp. 7795–7805, 2004.
- [26] J. M. Lee and E. L. Sonnhammer, "Genomic gene clustering analysis of pathways in eukaryotes," *Genome Research*, vol. 13, no. 5, pp. 875–882, 2003.
- [27] A. E. Osbourn and B. Field, "Operons," *Cellular and Molecular Life Sciences*, vol. 66, no. 23, pp. 3755–3775, 2009.
- [28] C. Petibon, M. Malik Ghulam, M. Catala, and S. Abou Elela, "Regulation of ribosomal protein genes: an ordered anarchy," *WIREs RNA*, vol. 12, no. 3, article e1632, 2021.

- [29] J. Spieth, G. Brooke, S. Kuersten, K. Lea, and T. Blumenthal, "Operons in *C. elegans*: Polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions," *Cell*, vol. 73, no. 3, pp. 521–532, 1993.
- [30] M. Jüttner and S. Ferreira-Cerca, "Looking through the lens of the ribosome biogenesis evolutionary history: possible implications for archaeal phylogeny and eukaryogenesis," *Molecular Biology and Evolution*, vol. 39, no. 4, 2022.
- [31] I. Kisly and T. Tamm, "Archaea/eukaryote-specific ribosomal proteins - guardians of a complex structure," *Computational and Structural Biotechnology Journal*, vol. 21, pp. 1249–1261, 2023.
- [32] T. F. Smith, J. C. Lee, R. R. Gutell, and H. Hartman, "The origin and evolution of the ribosome," *Biology Direct*, vol. 3, no. 1, p. 16, 2008.
- [33] B. P. Frohn, T. Härtel, J. Cox, and P. Schwill, "Tracing back variations in archaeal ESCRT-based cell division to protein domain architectures," *PLoS One*, vol. 17, no. 3, article e0266395, 2022.
- [34] S. Albers, J. Ashmore, T. Pollard, A. Spang, and J. Zhou, "Origin of eukaryotes: what can be learned from the first successfully isolated Asgard archaeon," *Faculty Reviews*, vol. 11, p. 3, 2022.
- [35] A. Du Toit, "Profilin(g) Asgard archaea," *Nature Reviews Microbiology*, vol. 16, p. 717, 2018.
- [36] Y. Liu, K. S. Makarova, W. C. Huang et al., "Expanded diversity of Asgard Archaea and their relationships with eukaryotes," *Nature*, vol. 593, no. 7860, pp. 553–557, 2021.
- [37] P. López-García and D. Moreira, "Cultured Asgard Archaea shed light on eukaryogenesis," *Cell*, vol. 181, no. 2, pp. 232–235, 2020.
- [38] L. Manoharan, J. A. Kozłowski, R. W. Murdoch et al., "Metagenomes from coastal marine sediments give insights into the ecological role and cellular features of *Loki* and *Thorarchaeota*," *mBio*, vol. 10, article e02039, 2019.
- [39] P. I. Penev, S. Fakhretaha-Aval, V. J. Patel et al., "Supersized ribosomal RNA expansion segments in Asgard Archaea," *Genome Biology and Evolution*, vol. 12, no. 10, pp. 1694–1710, 2020.
- [40] A. Spang, J. H. Saw, S. L. Jorgensen et al., "Complex archaea that bridge the gap between prokaryotes and eukaryotes," *Nature*, vol. 521, pp. 173–179, 2015.
- [41] R. Xie, Y. Wang, D. Huang et al., "Expanding Asgard members in the domain of Archaea sheds new light on the origin of eukaryotes," *Science China Life Sciences*, vol. 65, no. 4, pp. 818–829, 2022.
- [42] B. Avci, J. Brandt, D. Nachmias et al., "Spatial separation of ribosomes and DNA in Asgard archaeal cells," *The ISME Journal*, vol. 16, no. 2, pp. 606–610, 2022.
- [43] L. Eme, D. Tamarit, E. F. Caceres et al., "Inference and reconstruction of the heimdallarchaeal ancestry of eukaryotes," *Nature*, vol. 618, no. 7967, pp. 992–999, 2023.
- [44] T. A. Williams, C. J. Cox, P. G. Foster, G. J. Szöllösi, and T. M. Embley, "Phylogenomics provides robust support for a two-domains tree of life," *Nature Ecology & Evolution*, vol. 4, no. 1, pp. 138–147, 2020.
- [45] V. Da Cunha, M. Gaia, D. Gadelle, A. Nasir, and P. Forterre, "Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes," *PLOS Genetics*, vol. 13, no. 6, article e1006810, 2017.
- [46] T. Dandekar, B. Snel, M. Huynen, and P. Bork, "Conservation of gene order: a fingerprint of proteins that physically interact," *Trends in Biochemical Sciences*, vol. 23, no. 9, pp. 324–328, 1998.
- [47] H. C. Betts, M. N. Puttick, J. W. Clark, T. A. Williams, P. C. J. Donoghue, and D. Pisani, "Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin," *Nature Ecology & Evolution*, vol. 2, no. 10, pp. 1556–1562, 2018.
- [48] A. Booth, C. Mariscal, and W. F. Doolittle, "The modern synthesis in the light of microbial genomics," *Annual Review of Microbiology*, vol. 70, no. 1, pp. 279–297, 2016.
- [49] A. J. Crapitto, A. Campbell, A. Harris, and A. D. Goldman, "A consensus view of the proteome of the last universal common ancestor," *Ecology and Evolution*, vol. 12, no. 6, article e8930, 2022.
- [50] H. N. Nguyen, A. Jain, O. Eulenstein, and I. Friedberg, "Tracing the ancestry of operons in bacteria," *Bioinformatics*, vol. 35, no. 17, pp. 2998–3004, 2019.
- [51] C. R. Woese, "Interpreting the universal phylogenetic tree," *Proceedings of the National Academy of Sciences*, vol. 97, no. 15, pp. 8392–8396, 2000.
- [52] E. V. Koonin, "Horizontal gene transfer: essentiality and evolvability in prokaryotes, and roles in evolutionary transitions," *F1000Research*, vol. 5, p. 1805, 2016.
- [53] R. L. Tatusov, A. R. Mushegian, P. Bork et al., "Metabolism and evolution of *Haemophilus influenzae* deduced from a whole-genome comparison with *Escherichia coli*," *Current Biology*, vol. 6, no. 3, pp. 279–291, 1996.
- [54] G. Brandis, S. Cao, and D. Hughes, "Operon concatenation is an ancient feature that restricts the potential to rearrange bacterial chromosomes," *Molecular Biology and Evolution*, vol. 36, no. 9, pp. 1990–2000, 2019.
- [55] J. C. Bowman, A. S. Petrov, M. Frenkel-Pinter, P. I. Penev, and L. D. Williams, "Root of the tree: the significance, evolution, and origins of the ribosome," *Chemical Reviews*, vol. 120, no. 11, pp. 4848–4878, 2020.
- [56] G. E. Fox, "Origins and early evolution of the ribosome," in *Evolution of the Protein Synthesis Machinery and Its Regulation*, G. H. R. Jagus, Ed., pp. 31–60, Springer International Publishing AG, 2016.
- [57] A. S. Petrov, B. Gulen, A. M. Norris et al., "History of the ribosome and the origin of translation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, pp. 15396–15401, 2015.
- [58] M. R. Tirumalai, M. Rivas, Q. Tran, and G. E. Fox, "The peptidyl transferase center: a window to the Past," *Microbiology and Molecular Biology Reviews*, vol. 85, no. 4, article e0010421, 2021.
- [59] E. Camprubí, J. W. de Leeuw, C. H. House et al., "The emergence of life," *Space Science Reviews*, vol. 215, p. 56, 2019.
- [60] G. E. Fox, "Origin and evolution of the ribosome," *Cold Spring Harbor Perspectives in Biology*, vol. 2, no. 9, article a003483, 2010.
- [61] M. Rivas and G. E. Fox, "Further characterization of the pseudo-symmetrical ribosomal region," *Life*, vol. 10, no. 9, p. 201, 2020.
- [62] M. Rivas and G. E. Fox, "How to build a protoribosome: structural insights from the first protoribosome constructs that have proven to be catalytically active," *RNA*, vol. 29, no. 3, pp. 263–272, 2023.

- [63] E. Fer, B. Cuevas-Zuñivira, A. D. Goldman, Z. R. Adam, and B. Kacar, "Origin and Evolution of Translation: A Unifying Perspective Across Time," *EcoEvoRxiv*, 2023.
- [64] T. Bose, G. Fridkin, A. Bashan, and A. Yonath, "Origin of Life: Chiral Short RNA Chains Capable of Non-Enzymatic Peptide Bond Formation," *Israel Journal of Chemistry*, vol. 61, pp. 863–872, 2021.
- [65] T. Bose, G. Fridkin, C. Davidovich et al., "Origin of life: protoribosome forms peptide bonds and links RNA and protein dominated worlds," *Nucleic Acids Research*, vol. 50, no. 4, pp. 1815–1828, 2022.
- [66] N. R. Coordinators, "Database resources of the National Center for Biotechnology Information," *Nucleic Acids Research*, vol. 44, no. D1, pp. D7–D19, 2016.
- [67] N. A. O'Leary, M. W. Wright, J. R. Brister et al., "Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation," *Nucleic Acids Research*, vol. 44, no. D1, pp. D733–D745, 2016.
- [68] K. D. Pruitt, T. Tatusova, and D. R. Maglott, "NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins," *Nucleic Acids Research*, vol. 35, pp. D61–D65, 2007.
- [69] L.-X. Chen, K. Anantharaman, A. Shaiber, A. M. Eren, and J. F. Banfield, "Accurate and complete genomes from metagenomes," *Genome Research*, vol. 30, no. 3, pp. 315–333, 2020.
- [70] S. Mukherjee, D. Stamatis, J. Bertsch et al., "Genomes OnLine Database (GOLD) v.8: overview and updates," *Nucleic Acids Research*, vol. 49, no. D1, pp. D723–D733, 2021.
- [71] P. I. Penev, H. M. McCann, C. D. Meade et al., "ProteoVision: web server for advanced visualization of ribosomal proteins," *Nucleic Acids Research*, vol. 49, no. W1, pp. W578–W588, 2021.
- [72] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403–410, 1990.
- [73] J. Sun, P. N. Evans, E. J. Gagen et al., "Recoding of stop codons expands the metabolic potential of two novel Asgard archaeota lineages," *ISME Communications*, vol. 1, no. 1, p. 30, 2021.
- [74] R. Méheust, C. J. Castelle, A. L. Jaffe, and J. F. Banfield, "Conserved and lineage-specific hypothetical proteins may have played a central role in the rise and diversification of major archaeal groups," *BMC Biology*, vol. 20, no. 1, p. 154, 2022.
- [75] P. S. Adam, G. Borrel, C. Brochier-Armanet, and S. Gribaldo, "The growing tree of Archaea: new perspectives on their diversity, evolution and ecology," *The ISME Journal*, vol. 11, no. 11, pp. 2407–2425, 2017.
- [76] N. Dombrowski, J. H. Lee, T. A. Williams, P. Offre, and A. Spang, "Genomic diversity, lifestyles and evolutionary origins of DPANN archaea," *FEMS Microbiology Letters*, vol. 366, no. 2, 2019.
- [77] C. Petitjean, P. Deschamps, P. López-García, and D. Moreira, "Rooting the domain Archaea by phylogenomic analysis supports the foundation of the new kingdom Proteoarchaeota," *Genome Biology and Evolution*, vol. 7, no. 1, pp. 191–204, 2014.
- [78] C. Rinke, P. Schwientek, A. Sczyrba et al., "Insights into the phylogeny and coding potential of microbial dark matter," *Nature*, vol. 499, no. 7459, pp. 431–437, 2013.
- [79] T. Rodrigues-Oliveira, F. Wollweber, R. I. Ponce-Toledo et al., "Actin cytoskeleton and complex cell architecture in an Asgard archaeon," *Nature*, vol. 613, no. 7943, pp. 332–339, 2023.
- [80] S. Lu, J. Wang, F. Chitsaz et al., "CDD/SPARCLE: the conserved domain database in 2020," *Nucleic Acids Research*, vol. 48, pp. D265–D268, 2020.
- [81] R. Yoshida, T. Yoshimura, and H. Hemmi, "Reconstruction of the "Archaeal" mevalonate pathway from the methanogenic archaeon *Methanosarcina mazei* in *Escherichia coli* cells," *Applied and Environmental Microbiology*, vol. 86, no. 6, article e02889, 2020.
- [82] M. Nomura, R. Gourse, and G. Baughman, "Regulation of the synthesis of ribosomes and ribosomal components," *Annual Review of Biochemistry*, vol. 53, no. 1, pp. 75–117, 1984.
- [83] C. Ramírez, L. C. Shimmin, P. Leggatt, and A. T. Matheson, "Structure and Transcription of the L11-L1-L10-L12 Ribosomal Protein Gene Operon from the Extreme Thermophilic Archaeon *Sulfolobus acidocaldarius*," *Journal of Molecular Biology*, vol. 244, no. 2, pp. 242–249, 1994.
- [84] M. N. Price, A. P. Arkin, and E. J. Alm, "The life-cycle of operons," *PLoS Genetics*, vol. 2, no. 6, article e96, 2006.
- [85] Y. I. Wolf, I. B. Rogozin, A. S. Kondrashov, and E. V. Koonin, "Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context," *Genome Research*, vol. 11, no. 3, pp. 356–372, 2001.
- [86] J. M. Ranz, F. Casals, and A. Ruiz, "How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus *Drosophila*," *Genome Research*, vol. 11, no. 2, pp. 230–239, 2001.
- [87] L. Larotonda, D. Mornico, V. Khanna et al., "Chromosomal position of ribosomal protein genes affects long-term evolution of *Vibrio cholerae*," *mBio*, vol. 14, no. 2, article e0343222, 2023.
- [88] I.-M. A. Chen, K. Chu, K. Palaniappan et al., "The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities," *Nucleic Acids Research*, vol. 49, no. D1, pp. D751–D763, 2021.
- [89] N. Nagarajan and M. Pop, "Sequencing and genome assembly using next-generation technologies," in *Computational Biology. Methods in Molecular Biology*, vol. 673, no. 673, pp. 1–17, Humana Press, Totowa, NJ, 2010.
- [90] V. G. Stepanov, M. R. Tirumalai, S. Montazari, A. Checinska, K. Venkateswaran, and G. E. Fox, "Bacillus pumilus SAFR-032 genome revisited: sequence update and re-annotation," *PLoS One*, vol. 11, no. 6, article e0157331, 2016.
- [91] M. R. Tirumalai, V. G. Stepanov, A. Wunsche et al., "Bacillus safensis FO-36b and Bacillus pumilus SAFR-032: a whole genome comparison of two spacecraft assembly facility isolates," *BMC Microbiology*, vol. 18, no. 1, p. 57, 2018.
- [92] M. R. Tirumalai and G. E. Fox, "An ICEBs1-like element may be associated with the extreme radiation and desiccation resistance of *Bacillus pumilus* SAFR-032 spores," *Extremophiles*, vol. 17, no. 5, pp. 767–774, 2013.
- [93] S. G. Garg, N. Kapust, W. Lin et al., "Anomalous phylogenetic behavior of ribosomal proteins in metagenome-assembled Asgard Archaea," *Genome Biology and Evolution*, vol. 13, no. 1, 2021.
- [94] K. Mise and W. Iwasaki, "Unexpected absence of ribosomal protein genes from metagenome-assembled genomes," *ISME Communications*, vol. 2, no. 1, p. 118, 2022.

- [95] H. L. Wong, R. A. White 3rd, P. T. Visscher, J. C. Charlesworth, X. Vázquez-Campos, and B. P. Burns, “Disentangling the drivers of functional complexity at the metagenomic level in Shark Bay microbial mat microbiomes,” *The ISME Journal*, vol. 12, no. 11, pp. 2619–2639, 2018.
- [96] M. R. Tirumalai, G. E. Fox, R. V. S. Jr, L. A. Kutty, and E. L. Song, *The ribosomal protein cluster organization in Asgard Archaea—an analysis*, Authorea, 2022.