Eukaryotic-like ribosomal RNA region in Lokiarchaeota

Authors: Petar I. Penev[1,2], Sara Fakhretaha-Aval[1,3], Vaishnavi J. Patel[4], Jamie J. Cannone[4], Robin R. Gutell[4], Anton S. Petrov[1,3*], Loren Dean Williams[1,2,3*], Jennifer B. Glass[1,2,5*]

Affiliations:

[1]NASA Center for the Origin of Life, Georgia Institute of Technology, Atlanta, GA 30332-0400, USA

[2]School of Biological Sciences, Georgia Institute of Technology, North Avenue, Atlanta, GA 30332, USA

[3]School of Chemistry and Biochemistry, Georgia Institute of Technology, 901 Atlantic Dr, Atlanta, GA 30332, USA

[4]Department of Integrative Biology, The University of Texas at Austin, 2415 Speedway #C0930, Austin, TX 78712, USA

[5]School of Earth and Atmospheric Sciences, Georgia Institute of Technology, 311 Ferst Dr, Atlanta, GA 30332, USA

*Correspondence to: anton.petrov@biology.gatech.edu; loren.williams@chemistry.gatech.edu; jennifer.glass@eas.gatech.edu

Abstract

The ribosome's common core connects all life back to a common ancestor and serves as a window to relationships among organisms. In eukaryotes, the common core contains expansion segments (ES's) that vastly increase ribosomal RNA size. Supersized ES's have not been observed previously in Bacteria or Archaea, and the origin of eukaryotic ES's remains enigmatic. We discovered that the large subunit rRNA of *Lokiarchaeota*, the closest modern cell lineage to the last common ancestor of Archaea and Eukarya, bridges the gap in size between prokaryotic and eukaryotic rRNA. The long large subunit rRNA in *Lokiarchaeota* is largely due to the presence of two eukaryotic-like, supersized ES's, ES9 and ES39, which are transcribed *in situ*. We applied computational models, covariation analysis, and chemical footprinting experiments to study the structure and evolution of *Lokiarchaeota* ES9 and ES39. We also defined the eukaryotic ES39 fold for comparison. We found that *Lokiarchaeota* and eukaryotic ES's are structurally distinct: *Lokiarchaeota* ES39 has more and longer helices than the eukaryotic ES39 fold. Despite their structural differences, we found that *Lokiarchaeota* and eukaryotic ES's originated from a common ancestor that was "primed" for evolution of larger and more complex rRNAs than those found in Bacteria and other archaea.

Introduction

The ribosome connects all life on Earth back to the Last Universal Common Ancestor (LUCA) (Woese and Fox 1977). The small ribosomal subunit (SSU) decodes mRNA and the large ribosomal subunit (LSU) links amino acids together to produce coded protein. Both subunits are made of ribosomal RNA (rRNA) and ribosomal protein (rProtein). All cytoplasmic ribosomes contain a structurally conserved universal common core, comprised of 2800 nucleotides and 28 rProteins, and including the peptidyl transferase center (PTC) in the LSU and the decoding center (DCC) in the SSU (Melnikov, et al. 2012; Bernier, et al. 2018). The rRNA of the common core is a reasonable approximation of the rRNA in LUCA and is most similar to rRNA of extant bacteria (Melnikov, et al. 2012; Petrov, et al. 2014b; Bernier, et al. 2018).

In Eukarya, the rRNA of the common core is elaborated by expansion segments (ES's, **Fig. 1**) (Veldman, et al. 1981; Clark, et al. 1984; Hassouna, et al. 1984; Gonzalez, et al. 1985; Michot and Bachellerie 1987; Bachellerie and Michot 1989; Gutell 1992; Lapeyre, et al. 1993; Gerbi 1996; Schnare, et al. 1996). ES's emerge from a small number of conserved sites on the common core and are excluded from regions of essential ribosomal function such as the DCC, the PTC and the subunit interface (Ben-Shem, et al. 2010; Anger, et al. 2013). Expansion segments are larger and more numerous on the LSU than on the SSU; across phylogeny, size variation of the SSU rRNA is around 10% that of LSU rRNA (Gutell 1992; Gerbi 1996; Bernier, et al. 2018). Metazoan rRNAs contain supersized ES's of hundreds of nucleotides (nts).

The recent discovery and characterization of the Asgard archaeal superphylum suggests that the last archaeal and eukaryotic common ancestor (LAECA) contained key components of eukaryotic cellular systems (Spang, et al. 2015; Klinger, et al. 2016; Eme, et al. 2017; Zaremba-Niedzwiedzka, et al. 2017; Narrowe, et al. 2018; Spang, et al. 2019). Eukaryotic signature proteins (ESPs) found in Asgard archaea are involved in cytoskeleton, trafficking, ubiquitination, and translation. Asgard archaea also contain several homologs of eukaryotic ribosomal proteins (Hartman and Fedorov 2002; Spang, et al. 2015; Zaremba-Niedzwiedzka, et al. 2017). Before our work here, it was not known if Asgard rRNAs could contain eukaryotic-like features such as supersized ES's. Eukaryotic-like, supersized ES's have not been observed previously in Bacteria or Archaea and were considered unique to eukaryotes (Ware, et al. 1983; Clark, et al. 1984; Hassouna, et al. 1984; Gerbi 1996; Melnikov, et al. 2012).

Here, we apply computation and experiment to study the structure and evolution of Asgard rRNA. We find that LSU rRNA of the Asgard phylum, *Lokiarchaeota*, contains an archaeal common core and supersized eukaryotic-like ES's. No es's were found in SSU rRNA of *Lokiarchaeota*.In size and complexity, *Lokiarchaeota* LSU ES's exceed those of protists rRNAs and rival those of metazoan rRNAs. Our data suggests that the large ES's of *Lokiarchaeota* and Eukarya can be traced back to a common ancestor.

Results

Comparative analysis reveals broad patterns of LSU rRNA size relationships. Previously, we developed the SEREB MSA (Sparse and Efficient Representation of Extant Biology, Multiple Sequence Alignment) as a tool for comparative analysis of rRNA and rProtein sequences (Bernier, et al. 2018). The SEREB MSA is a structure-informed alignment of a sparse and unbiased group of sequences including all major phyla. The MSA was manually curated and extensively cross-validated. The SEREB MSA is useful as a seed to study a variety of evolutionary phenomena. Previously, we augmented the SEREB MSA to include additional metazoan sequences, allowing us to characterize ES's and their evolution in metazoans (Mestre-Fos, et al. 2019a; Mestre-Fos, et al. 2019b). Here, we augmented the SEREB MSA to include 21 sequences from the Asgard superphylum (**supplementary datasets S1,2**).

The SEREB MSA indicates that size relationships of LSU rRNAs follow the general pattern: Bacteria (2725-2960 nts, n=61 [n is number of species]) < Archaea (2886 to 3094 nts, n=48, excludes *Lokiarchaeota*) < Eukarya (3300-5200 nts, n=30; **Fig. 2**). Archaeal rRNAs frequently contain micro-expansion segments (μ-ES's; stem loops of less than 20 nts) at positions of eukaryotic ES's. Archaeal LSU rRNAs commonly contain μ-ES's at the sites of attachment of ES9 and ES39 in eukaryotes. For example, in the archaeon *P. furiosus*, μ-ES9 is 33 nts and μ-ES39 is 45 nts (**Fig. 1C, supplementary figure S2)**. The archaeon *Haloarcula marismortui* contains μ-ES9 but lacks μ-ES39 (not shown).

***Lokiarchaeota* bridge Eukarya and Archaea in LSU rRNA size.** The Asgard augmentation of the SEREB MSA reveals unexpectedly large *Lokiarchaeota* LSU rRNAs. *Lokiarchaeota* LSU rRNAs range from 3100 to 3300 nts (n=7). *Lokiarchaeota* rRNAs are close to or within the observed size range of eukaryotic LSU rRNAs (**Fig. 2**). The Asgard-augmented

SEREB MSA reveals supersized ES's in *Lokiarchaeota* spp. These supersized ES's attach to the universal common core rRNA at the sites of attachment of eukaryotic ES9 and ES39 and archaeal μ-ES9 and μ-ES39 (**Fig. 1**). Here we explored the Asgard augmentation of the SEREB MSA to investigate the structure, distribution, and evolution of rRNA expansions of Asgard archaea.**ES9 and ES39 in some *Lokiarchaeota* are larger than μ-ES's of other archaea and ES's of protists.** The MSA shows that ES39 in *Lokiarchaeota* ranges in size from 95 to 200 nts, compared to 138 nts in *Saccharomyces cerevisiae*, 178 nts in *Drosophila melanogaster*, and 231 nts in *Homo sapiens* (**Fig. 3)**. For *Candidatus* Lokiarchaeota archaeon 1244-F3-H4-B5 (*Lokiarchaeota* F3H4_B5), the primary focus of our work here, ES39 is 191 nts **(Figs. 3, 4**). ES9 in *Lokiarchaeota* ranges from 29 to 103 nts, and in some species is larger than any known eukaryotic ES9 except the ES9 *Guillardia theta* (29 nts in *S. cerevisiae*, 44 nts in *D. melanogaster*, 44 nts in *H. sapiens*, and 111 nts in *G. theta*; **supplementary figure S2**). ES9 is 86 nts in *Lokiarchaeota* F3H4_B5 **(supplementary figure S2**). ES9 and ES39 contribute to the large size of *Lokiarchaeota* LSU rRNAs compared to the LSU rRNAs of other archaea. Outside of *Lokiarchaeota*, archaea lack supersized ES's. Some *Lokiarchaeota* also lack supersized ES's.

***Lokiarchaeota* LSU rRNA contain the common core.** We have determined the extent of structural similarity of *Lokiarchaeota* LSU rRNAs with those of various eukaryotes. We combined computational and experimental methods to characterize the LSU rRNA secondary structure of *Lokiarchaeota* F3H4_B5 (**Fig. 1E; supplementary figure S1**). Secondary and three-dimensional structures are known for ribosomes of several eukaryotes and archaea, providing a basis for modeling by homology. Like all other LSU rRNAs, *Lokiarchaeota* LSU rRNA contains the rRNA common core, which is trivial to model because the backbone atoms of the common core are highly conserved in all cytosolic ribosomes.

ES39 has a well-defined fold in eukaryotes. To determine similarities and differences between ES39's in various eukaryotes and archaea, we investigated the extent of conservation of eukaryotic ES39 over phylogeny. We compared experimental three-dimensional structures of rRNAs of species ranging from protists to primates (Ben-Shem, et al. 2010; Klinge, et al. 2011; Khatter, et al. 2015; Li, et al. 2017). The ES39 fold consists of H98 (20-30 nts), helix *b* (40-50 nts), and the linkage of H98 and helix *b* by three unpaired segments of rRNA, which are each around 15 nts in length (**Fig. 6; supplementary figure S3)**. The ES39 fold is conserved in structure but not in sequence.

The ES39 fold has complex evolutionary history. In general, ES's have increased in size over evolutionary history via accretion. Growth by addition of one rRNA helix to another is commonly marked an insertion fingerprint (Petrov, et al. 2014b; Petrov, et al. 2015). Eukaryotic ES39 lacks insertion fingerprints. Instead, the ES39 fold has three long non-duplex rRNA strands at its base that are tightly associated with the ribosomal surface. These unpaired regions are unique to eukaryotic ES39 and do not appear in archaeal ES39. Two of the unpaired regions interact with eukaryotic-specific helical extensions on rProteins uL13 and eL14 **(supplementary figure S4)** and the third one interacts with ES7 and rProtein aL33 **(supplementary figures S3, S5, S6)** (Khatter, et al. 2015). These eukaryotic-specific unpaired regions indicate a complex evolutionary history in which changes in ES39 structure were coupled with changes in other ribosomal components.

The ES39 fold is decorated by a variable length helix. Many, but not all, eukaryotes possess a third helix (helix *a*) that extends from the ES39 fold **(Fig. 6A)**. Helix *a* expands in size from simple unicellular eukaryotes, such as *Tetrahymena thermophila* (no helix) and *Toxoplasma gondii* (10 nts), to more complex eukaryotes such the fungus *S. cerevisiae* (18 nts) and the insect *Drosophila melanogaster* (20 nts). Helix *a* is largest in the eukaryotic phylum *Chordata* (106 nts for *H. sapiens*; **supplementary figure S3**).

Initial *Lokiarchaeota* ES39 secondary models were predicted by two methods. One preliminary secondary structural model of ES39 of *Lokiarchaeota* F3H4_B5 was generated using mfold (Zuker 2003) (**Fig. 4**). Mfold predicts a minimum free energy secondary structures using experimental nearest-neighbor parameters. We selected the mfold model with lowest free energy for further studies. A second model forced *Lokiarchaeota* ES39 to conform to the *H. sapiens* secondary structure. The mfold model was confirmed to be correct by covariation analysis and SHAPE reactivity data, below.

Covariation supports the mfold model for secondary structure of ES39 of *Lokiarchaeota* F3H4_B5. Covariation, or cooperative changes of base-paired nucleotides across phylogeny, can help reveal RNA secondary structure (Levitt 1969; Ninio, et al. 1969; Woese, et al. 1980; Noller, et al. 1981; Gutell, et al. 1993; Gutell, et al. 1994). Base-pairs can be detected through covariation analysis. We calculated base-pairing conservation predicted by helical regions of both the mfold model and the *H. sapiens* homology model using available sequence data. Helical regions of the *Lokiarchaeota* ES39 secondary structural model predicted by mfold

show covarying nucleotides **(Fig. 5D)**, with conservation of base-pairing (**Fig. 5C**). The observation of these covarying nucleotides supports the model determined by mfold.

Chemical footprinting confirms the mfold model for secondary structure of ES39 of *Lokiarchaeota* F3H4_B5. We further tested the secondary structural model of *Lokiarchaeota* F3H4_B5 ES39 using Selective 2' Hydroxyl Acylation analyzed by Primer Extension (SHAPE). This experimental method provides data on RNA flexibility at the nucleotide level (Merino, et al. 2005; Wilkinson, et al. 2006). SHAPE reactivity is generally high in unpaired RNA, which is more flexible, and low in paired RNA, which is less flexible. SHAPE has been widely used to probe rRNA (Leshin, et al. 2011; Lavender, et al. 2015; Gomez Ramos, et al. 2017; Lenz, et al. 2017) and other RNAs (Wilkinson, et al. 2005; Gilbert, et al. 2008; Stoddard, et al. 2008; Watts, et al. 2009; Novikova, et al. 2012; Spitale, et al. 2013; Huang, et al. 2014). The SHAPE results from *Lokiarchaeota* F3H4_B5 ES39 rRNA (**supplementary figure S7**) are in agreement with the secondary structure based on co-variation and mfold. Reactivity is low for paired nucleotides and is high in loops and bulges (**Fig. 5B**). The resolution and accuracy of the SHAPE data are supported by observation of relatively high reactivity at the vast majority of unpaired nucleotides and low reactivity for most paired nucleotides of the *Lokiarchaeota* ES39 secondary structure. The *Lokiarchaeota* SHAPE data are not consistent with models that force *Lokiarchaeota* ES39 to conform to the *H. sapiens* secondary structure.

***Lokiarchaeota* and Asgard ES39 deviate from the eukaryotic ES39 fold.** The eukaryotic ES39 junction of helices H98, *a*, and *b* contains significant extent of unpaired nucleotides; it consists of three 15-nt unpaired regions. By contrast, ES39 in *Lokiarchaeota* F3H4_B5 contains more paired nucleotides than in eukaryotes and lacks unpaired regions greater than 8 nts **(Figs. 1E, 4**). *Lokiarchaeota* F3H4_B5 ES39 is composed of four short helical regions (H98, *a1*, *a2*, *b*; each up to 38 nts) and one long helical region (helix *a*: 72 nts). H98 and helix *b* connect in a three-way junction with helix *a* at the base of ES39. Helices *a1* and *a2* split helix *a* at the top of ES39 in a three-way junction.

We modeled and visualized (Cannone, et al. 2002) secondary structures of ES39 sequences from additional Asgard species (**supplementary figure S8**). None of these modeled structures exhibited long unpaired regions. ES39 of all modelled Asgard archaea contain a three-way junction that connects helices H98, *a* and *b*. This three-way junction is similar to the one seen in *Lokiarchaeota* F3H4_B5 (**Fig. 4**). Additionally, some members of Asgard archaea

revealed an additional branching of helix *a* into *a1* and *a2* mirroring the morphology of ES39 in *Lokiarchaeota* F3H4_B5. Despite the common branching morphology, the length of the individual helices substantially varies between different species (**Fig. 5D**).

Supersized ES's of Lokiarchaeota are transcribed *in situ*. To assess whether *Lokiarchaeota* ES's are transcribed, we assembled metatranscriptomic reads from sediment from the Gulf of Mexico (Yergeau, et al. 2015; Cai, et al. 2018). We found multiple transcripts from *Lokiarchaeota*-like LSU ribosomes that contain ES9 and ES39 sequences, confirming that *Lokiarchaeota* ES's are indeed transcribed *in situ* **(Fig. 5D**; **supplementary dataset S2)**.

Discussion

The recent discovery of the archaeal *Lokiarchaeota* phylum, which contain multitudes of ESPs, has redefined our understanding of eukaryotic evolutionary history (Spang, et al. 2015; Zaremba-Niedzwiedzka, et al. 2017). The incorporation of *Lokiarchaeota* sequences into phylogenies has brought Archaea and Eukarya close together in the tree of life (Hug, et al. 2016; Fournier and Poole 2018). Here, we extend the molecular comparison by identifying commonalities of rRNA of Eukarya and *Lokiarchaeota*.

***Lokiarchaeota* rRNA has unique eukaryotic-like features.** The ribosome has been extensively studied as both an evolving system (Agmon, et al. 2005; Smith, et al. 2008; Bokov and Steinberg 2009; Fox 2010; Petrov, et al. 2015; Melnikov, et al. 2018) and as a window to relationships among organisms (Woese and Fox 1977; Hillis and Dixon 1991; Olsen and Woese 1993; Fournier and Gogarten 2010; Hug, et al. 2016). Previous work revealed robust patterns that govern ribosomal variation over phylogeny (Hassouna, et al. 1984; Gerbi 1996; Melnikov, et al. 2012) and suggest mechanisms of ribosomal change over evolution (Petrov, et al. 2014b; Petrov, et al. 2015; Kovacs, et al. 2017; Melnikov, et al. 2018). Here, we extend structure-based methods of comparative analysis to *Lokiarchaeota* rRNA and demonstrate its distinctive eukaryotic-like features. We provide mechanistic models for the evolution of common rRNA features of eukaryotes and *Lokiarchaeota*.

We assessed the extent to which *Lokiarchaeota* ribosomes follow or deviate from previously established patterns of ribosomal structure. We found that *Lokiarchaeota* ribosomes follow several established patterns.

Lokiarchaeota ribosomes:

- contain the universal common core of rRNA and rProteins (this work) (Spang, et al. 2015),
- confine rRNA diversity of structure and size to ES's/μ-ES's (this work),
- restrict ES's to universally conserved sites on the common core (Ware, et al. 1983; Clark, et al. 1984; Hassouna, et al. 1984; Michot and Bachellerie 1987; Bachellerie and Michot 1989; Lapeyre, et al. 1993; Gerbi 1996),
- avoid ES attachment from the ribosomal interior or near functional centers (Ben-Shem, et al. 2010; Anger, et al. 2013), and
- concentrate diversity in structure and size on LSU rRNA, not SSU rRNA (Gerbi 1996; Bernier, et al. 2018).

Lokiarchaeota ribosomes deviate from several previous patterns of variation of ribosome structure over phylogeny. *Lokiarchaeota* LSU rRNAs are larger than their place on the archaeal domain of the tree of life would predict. Excluding *Lokiarchaeota*, rRNA length increases in the order: Bacteria < Archaea ≪ protists ≪ metazoans (Melnikov, et al. 2012; Petrov, et al. 2014b; Bernier, et al. 2018). *Lokiarchaeota* rRNA is eukaryotic-like in length, eclipsing the rRNA of many protists. *Lokiarchaeota* ES39 is larger than ES39 in protists and some metazoans. ES9 of *Lokiarchaeota* is larger than ES9 of any system except *Guillardia theta*. Both ES9 and ES39 of *Lokiarchaeota* are larger than ES9 and ES39 of any other archaeal phylum known to date.

***Lokiarchaeota* ES39 is located within an archaeal structural environment in the ribosome.** ES39 in Eukarya protrudes from helices 94 and 99 of the ribosomal common core **(Fig. 6B)**. In three dimensions, ES39 is close to ES7 and rProteins uL13, eL14, and aL33 (**supplementary figures S3, S4, S5**). These elements in *Lokiarchaeota* are more similar to Archaea than to Eukarya. Additionally, *Lokiarchaeota*, like all Archaea, contain helix 1 (H1), which is in direct contact with H98 at the base of ES39, whereas eukaryotes lack H1 (**Fig 4B**). Combined with the eukaryotic-like size of *Lokiarchaeota* ES39, these characteristics predict that *Lokiarchaeota* ribosomes have a unique structure in this region.

The pathway of ES39 evolution is unique. The ribosome has grown in size by a process of accretion (Petrov, et al. 2014b). Basal structure is preserved when new rRNA is acquired. For instance, ES7 shows continuous growth over phylogeny, expanding from LUCA to Archaea to protists to metazoans to mammals (Petrov, et al. 2014b; Bernier, et al. 2018). The accretion model predicts that H98, at the base of ES39, would superimpose in Bacteria, Archaea, and

Eukarya, but in fact H98 does not overlap in superimposed 3D structures (**Fig. 6B**). The archaeon *P. furiosus* has a slightly extended and bent H98 compared to the bacterium *E. coli* (**Fig. 6B**). This spatial divergence is likely due to the difference in how *E. coli* H98 and *P. furiosus* H98 interact with H1 of the LSU. *E. coli* H98 interacts within the H1 minor groove through an A-minor interaction, while in *P. furiosus* H98 is positioned on top of H1 (**Fig. 6B**). H1 is absent in eukaryotes (**Fig. 1**), allowing H98 to occupy the position of H1 (**Fig. 6B**).

LAECA likely had a large ES39. The observation of supersized ES39's in species with and without H1 suggests that ES39 growth is independent of the presence or absence of H1. Whether LAECA had a large or small ES39 is difficult to ascertain because the Asgard superphylum shows wide size variability in ES39 (**Fig. 3**). However, the accretion model suggests that LAECA contained a large ES39, which fulfilled a patching role upon the loss of H1. In this model, ES39 was remodeled upon loss of H1. ES39 underwent strand dissociation in Eukarya to fill the space left by the deletion of H1; this unpaired ES39 structure was further stabilized by eukaryotic extensions of rProteins uL13 and eL14 (**supplementary figure S4**). This pattern of structural patching has been observed in mitoribosomes (Petrov, et al. 2019). If ES39 grew to its eukaryotic size after the loss of H1, one would not expect remodeling to form the unpaired structure; ES39 would have gradually accreted helices like other parts of the ribosome.

***Lokiarchaeota* ES39 may extend in a different direction than eukaryotic ES39.** *Lokiarchaeota* spp. have larger ES39 than other archaea (**Fig. 3**) and possess H1, unlike Eukarya (**Fig 4B**). We predict that *Lokiarchaeota* ES39 has an archaeal-like interaction with H1 through H98 and helix *b* (**Fig. 6B)**. *Lokiarchaeota* helix *a* likely grows out from the three-way junction between H98 and helix *b*, perpendicular to the eukaryotic helix *a* (**Fig. 6B)**. While helix *a* of eukaryotic ES39 is pointed in the direction of the sarcin-ricin loop, helix *a* of *Lokiarchaeota* is likely pointed in the direction of the central protuberance or the exit tunnel.

Growth of *Lokiarchaeota* ES9 is unrestricted due to absence of ES15. To date, ES9 in *Lokiarchaeota* is longer than ES9 in any organism except the alga *Guillardia theta* (**supplementary figure S2**). In *H. sapiens*, ES9 forms a kissing-loop structure with ES15 (**supplementary figure S6**). ES15 is not present in *Lokiarchaeota*, which may permit unrestricted growth of ES9.

***Lokiarchaeota* have higher complexity of rRNA than other archaea, consistent with higher *Lokiarchaeota* molecular complexity.** Previously, we demonstrated that biological complexity is correlated with LSU rRNA sizes (Petrov, et al. 2014b). Here, we found that *Lokiarchaeota* spp. have ES39's with lengths that eclipse many eukaryotes and are close to those of Chordates, resulting in larger LSU rRNAs in *Lokiarchaeota* than other archaea. This size pattern is consistent with the idea that LSU rRNA size tracks organismal complexity because *Lokiarchaeota* have more complex cellular infrastructure than other archaea (Akıl and Robinson 2018; Imachi, et al. 2019).

The specific roles of μ-ES's and ES's over phylogeny are unknown but are likely complex, polymorphic, and pleotropic. The observation of μ-ES's in Archaea, ES's in Eukarya, and supersized ES's in *Lokiarchaeota* suggest that (i) the roots of ribosomal complexity are inextricably intertwined with the roots of Eukarya, and (ii) the roots of Eukarya penetrate more deeply into the archaeal domain than previously recognized, conclusions that are consistent with recent phylogenetic results that root Eukarya within the archaeal branch (Williams, et al. 2020).

Conclusions

Lokiarchaeota ribosomes contain supersized ES39's with structures that are distinct from eukaryotic ES39's. *Lokiarchaeota* ES9's are larger than eukaryotic ES9's. To date, *Lokiarchaeota* is the only prokaryotic phylum with supersized ES's, bringing the size range of archaeal LSU close to those of eukarya. *Lokiarchaeota* ES39 likely grows outward from the ribosomal surface in a different direction than eukaryotic ES39's. Our findings raise the possibility that eukaryotic-sized ES's existed on the ribosomal surface before LECA, suggesting that ribosomal complexity is more deeply rooted than previously known.

Materials and Methods

Genome sequencing, assembly, and binning

Sample collection. Sediments were cored from deep seafloor sediment at ODP site 1244 (44°35.1784´N; 125°7.1902´W; 895 m water depth) on the eastern flank of Hydrate Ridge ~3 km northeast of the southern summit on ODP Leg 204 in 2002 (Tréhu, et al. 2003) and stored at -80°C at the IODP Gulf Coast Repository.

DNA extraction. DNA was extracted from sediment from F3-H4 (18.10 meters below the seafloor) using a MO-BIO PowerSoil total RNA Isolation Kit with the DNA Elution Accessory Kit, following the manufacturer protocol without beads. Approximately 2 grams of sediments were used for the extraction from six extractions (12 g total) and DNA pellets from the two replicates from each depth were pooled together. DNA concentrations were measured using a Qubit 2.0 fluorometer with dsDNA High Sensitivity reagents (Invitrogen, Grand Island, NY, USA). DNA yield was 7.5 ng per gram of sediment.

Multiple displacement amplification, library preparation, and sequencing. Genomic DNA was amplified using a REPLI-g Single Cell Kit (Qiagen, Germantown, MD, USA) using UV-treated sterile plasticware and reverse transcription-PCR grade water (Ambion, Grand Island, NY, USA). Quantitative PCR showed that the negative control began amplifying after 5 hr of incubation at 30°C, and therefore, the 30°C incubation step was shortened to 5 hr using a Bio-Rad C1000 Touch thermal cycler (Bio-Rad, Hercules, CA, USA). DNA concentrations were measured by Qubit. Two micrograms of MDA-amplified DNA were used to generate genome libraries using a TruSeq DNA PCR-Free Kit following the manufacturer's protocol (Illumina, San Diego, CA, USA). The resulting libraries were sequenced using a Rapid-Run on an Illumina HiSeq 2500 to obtain 100 bp paired-end reads. Metagenomic sequences were deposited into NCBI as accession numbers SAMN07256342-07256348 (BioProject PRJNA390944).

Metagenome assembly, binning, and annotation. Demultiplexed Illumina reads were mapped to known adapters using Bowtie2 in local mode to remove any reads with adapter contamination. Demultiplexed Illumina read pairs were quality trimmed with Trim Galore (Martin 2011) using a base Phred33 score threshold of Q25 and a minimum length cutoff of 80 bp. Paired-end reads were then assembled into contigs using SPAdes assembler (Bankevich, et al. 2012) with --meta option for assembling metagenomes, iterating over a range of k-mer values

(21,27,33,37,43,47,51,55,61,65,71,75,81,85,91,95). Assemblies were assessed with reports generated with QUAST (Gurevich, et al. 2013). Features on contigs were predicted through the Prokka pipeline with RNAmmer for rRNA, Aragorn for tRNA, Infernal and Rfam for other non-coding RNA and Prodigal for protein coding genes. Each genomic bin was searched manually for 23S rRNA sequences. The *Lokiarchaeota* F3-H4-B5 bin (estimated 2.8% completeness and 0% contamination) was found to contain a 3300 nt 23S rRNA sequence. The *Lokiarchaeota* F3-H4-B5 bin was deposited into NCBI as BioSample SAMN13223206 and GenBank genome accession number WNEK00000000.

Environmental 23S rRNA transcript assembly and analysis

Assembly. Publicly available environmental meta-transcriptomic reads were downloaded from NCBI BioProject PRJNA288120 (Yergeau, et al. 2015). Quality evaluation of the reads was performed with FastQC (Andrews 2012) and trimming was done with TrimGalore (Martin 2011). Assembly of SRR5992925 was done using the SPADES (Bankevich, et al. 2012) assembler with --meta and --rna options, to evaluate which performs better. Basic statistic measures such as Nx, contig/transcript coverage and length were compared (**Supplementary Datasets S3, S4**) yielding better results for the rnaspades assembler. All subsequent meta-transcriptomic datasets were assembled with rnaspades.

Identifying ribosomal RNA sequences. BLAST databases were constructed (Altschul, et al. 1990) from the resulting contig files and they were queried for ribosomal regions characteristic of the Asgardian clade (ES39/ES9 sequences from GC14_75). Additionally, the program quast (Gurevich, et al. 2013) with --rna-finding option was used.

SEREB Multiple Sequence Alignment (MSA) augmentation. High scoring transcripts, as well as genomic sequences with Asgard origin, were included in the SEREB MSA (Bernier, et al. 2018) using the program mafft (Katoh and Standley 2013) with the --add option. Known intronic regions (Cannone, et al. 2002) were removed from new sequences. The highly variable region of ES39 was manually aligned using structural predictions from mfold (Zuker 2003).

LSU size comparison. The LSU size comparison was based on the transcribed gene for the LSU, which is comprised of a single uninterrupted rRNA sequence for bacteria and archaea (**Fig. 1A,C,E**), and is comprised of multiple concatenated rRNA sequences for the fragmented eukaryotic rRNA gene (**Fig. 1B,D,F**). The 5S rRNA, which is essentially constant, is excluded

from the size calculation. The comparison excludes rRNAs of endosymbionts and pathogens, which tend to contain reduced genomes, metabolisms, and translation systems (Peyretaillade, et al. 1998; Moran 2002; McCutcheon and Moran 2011).

Secondary structure models. To model the secondary structure of *Candidatus* Lokiarchaeota archaeon 1244-F3-H4-B5 LSU rRNA, we used the secondary structure of *P. furiosus* (Petrov, et al. 2014a) and a multiple sequence alignment (MSA) of archaeal LSU rRNAs broadly sampled over the phylogeny (**supplementary dataset S1**). Locations of expansion segments were unambiguously identified from the MSA. Due to the low percent identity (<50%) (Bernhart and Hofacker 2009) we applied *ab initio* modelling for ES regions. The secondary structures of the ES's were predicted by mfold (Zuker 2003).

Covariation. To verify the secondary structures of the highly variable ES regions base-pairing conservation was calculated with the program Jalview (Waterhouse, et al. 2009). Gaps from the MSA were ignored in the calculation to produce comparable results about available regions. The base-pairing model of secondary structures of ES9 (**supplementary figure S9**) and ES39 (**Fig. 4C,D**) was generated in the Jalview annotation format and used for the base-pairing conservation calculation.

Defining the eukaryotic ES39 fold. To identify the structurally invariant part of ES39 in Eukaryotes, we used superimposition based on the common core within domain VI of the ribosomal structures from 4 eukaryotes (*Tetrahymena thermophila*, *Toxoplasma gondii*, *Saccharomyces cerevisiae*, *Homo sapiens*; **supplementary figure S3**). Initially the *Drosophila melanogaster* ribosomal structure (PDB ID: 4V6W) was used in identifying the core. However, as it has additional loops elongating the unpaired regions, we excluded it from our analysis. *Drosophila melanogaster* is known to have AU-enriched ES's; therefore, it is not surprising that it has perturbations in its ES39.

ES39 rRNA SHAPE analysis

Synthesis of *Lokiarchaeota* ES39 rRNA. pUC57 constructs containing T7 promoter and the gene encoding *Lokiarchaeota* ES39 rRNA was linearized using HindIII restriction enzyme. *Lokiarchaeota ES39* rRNA was synthesized by *in vitro* transcription using HiScribe™ T7 High Yield RNA Synthesis Kit; New England Biolabs. RNA was then precipitated in ethanol/ammonium acetate and purified by G25 size exclusion chromatography

(illustraTMNAPTM-10, GE Healthcare). RNA purity was assayed by denaturing gel electrophoresis.

SHAPE reaction. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE; Wilkinson, et al. 2006) was performed to chemically probe local nucleotide flexibility in ES39 rRNA. *In vitro*-transcribed ES39 rRNA was added to folding buffer (180mM NaOAc, 50mM Na-HEPES (pH 8.0) and 1 mM 1,2- diaminocyclohexanetetraacetic acid (DCTA))to obtain 400nM RNA in total volume of 80 µL. RNA was annealed by cooling from 75 °C to 25 °C at 1 °C/min. RNA modification reaction was performed with final concentration of 100mM benzoyl cyanide (Sigma) prepared in dimethyl sulfoxide (DMSO). Non-modified RNA samples were incubated with DMSO. Reactions were carried out for 2 min at room temperature. Modified RNAs and control sample were purified by precipitation in ethanol and ammonium acetate at 20 °C for 2 hr. RNA samples were centrifuged at 4°C for 10 min. The RNA pellets were washed out with 100 µL of 80% ethanol for two times and dried out using Speedvac. 22 µL of TE buffer [1mM EDTA and 10mM Tris-Cl (pH 8.0)] were added into each samples and pellet were resuspended.

Reverse transcription. Reverse transcription was conducted on 20 µL of modified RNAs and unmodified RNA sample as a control, in presence of 8 pmol 5'[6-FAM] labeled primer (5'-GAACCGGACCGAAGCCCG-3'), 2 mM DTT, 625 µM of each deoxynucleotidetriphosphate (dNTP), and 5 µL of reverse transcription (RT) 5X first-strand buffer [250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM $MgCl_2$]. To anneal the primer, samples were heated at 95°C for 30 secs, held at 65°C for 3 min, and then 4°C for 10 min. RT mixtures were incubated at 52°C for 2 min before addition of 1 µL(200 U) of Superscript III Reverse transcriptase (Invitrogen) and reactions were incubated at 55°C for 2 hr. later, RT terminated by heating smaples at 70°C for 15 min. Chain termination sequencing reaction was performed on 10 pmol unmodified RNA prepared in TE buffer, 8 pmol 5'[6-FAM] labeled primer, with a ratio of 1:1 dideoxynucleotide (ddNTP) to dNTP. A sequencing reaction was performed with the same condition without ddNTPs.

Capillary electrophoresis of RT reaction products and data analysis. Capillary electrophoresis of RT reactions was performed as described previously (Hsiao, et al. 2013). For each reaction 0.6 µl DNA size standard (Geneflo™ 625), 17.4 µl Hi-Di Formamide (Applied Biosystems), and 2 µl of RT reaction mixture were loaded in a 96-well plate. Samples were

heated at 95°C for 5 min before electrophoresis and the RT products were resolved using applied biosystems. SHAPE data were processed using a Matlab scripts as described previously (Athavale, et al. 2012). SHAPE profile was mapped onto ES39 rRNA secondary structure with the RiboVision program (Bernier, et al. 2014).

Acknowledgments: We thank Cecilia Kretz and Piyush Ranjan for sample preparation and genome binning. We thank Brett Baker for providing an unpublished *Heimdallarchaeota* sequence (Fig. S8B). We thank Jessica Bowman, Santi Mestre-Fos, Aaron Engelhart, Anthony Poole, and Betül Kaçar for helpful conversations.

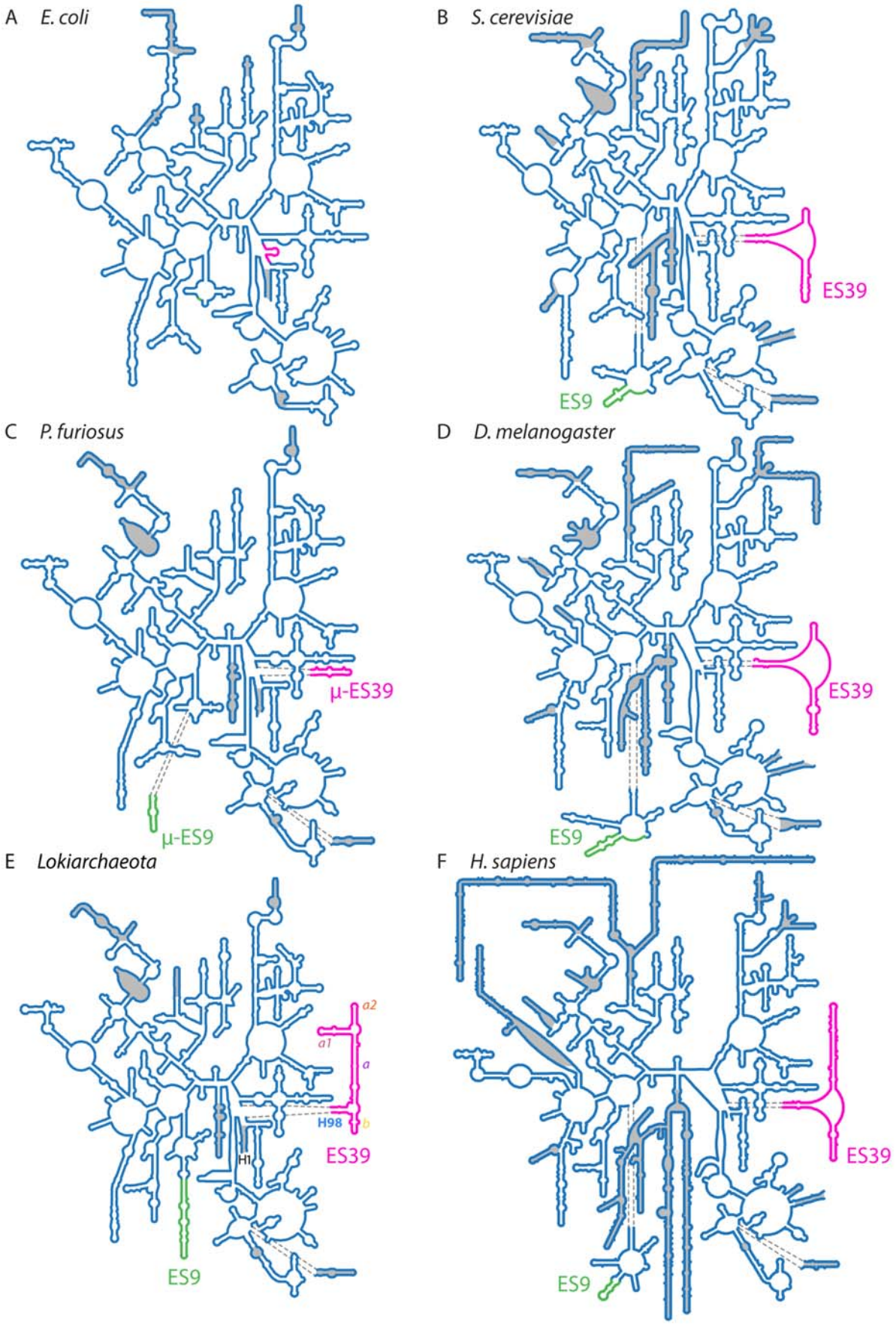
Funding: This research was supported by National Aeronautics and Space Administration grants NNX16AJ29G, the Center for the Origin of Life grant 80NSSC18K1139, and a Center for Dark Energy Biosphere Investigations (C-DEBI) Small Research Grant. This is C-DEBI contribution [will be provided upon paper acceptance]. This research was supported in part through research cyberinfrastructure resources and services provided by the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology, Atlanta, Georgia, USA.

Author contributions: ASP, LDW, and JBG conceptualized the research. PIP curated the data. ASP, LDW, and JBG acquired the funding and administered the project. SF-A performed the SHAPE experiments. JJC and VJP performed analyses. ASP, LDW, RRG, and JBG supervised the research. RRG validated the research. PIP and VJP prepared the figures. PIP, JBG, and LDW wrote the manuscript with input from all authors.

Competing interests: Authors declare no competing interests.

Data and materials availability: Metagenomic sequences were deposited into NCBI as accession numbers SAMN07256342-07256348 (BioProject PRJNA390944). The *Lokiarchaeota* F3-H4-B5 bin was deposited into NCBI as BioSample SAMN13223206 and GenBank genome accession number WNEK00000000. The *Lokiarchaeota* F3H4_B5 23S rRNA gene is in the reverse complement of contig WNEK01000002.1, nucleotide positions 251-3579.

Figures



446 **Figure 1: Secondary structures of the LSU rRNA from Bacteria, Archaea, and Eukarya**. (A)
447 *Escherichia coli* (Bacteria), (B) *Saccharomyces cerevisiae* (fungus, Eukarya); (C) *Pyrococcus furiosus*
448 (Archaea); (D) *Drosophila melanogaster* (insect, Eukarya); (E) *Lokiarchaeota* F3H4_B5 (Archaea); (F)
449 *Homo sapiens* (primate, Eukarya). Secondary structures in panels A, B, C, D, and F are taken from
450 *Petrov, et al. (2014a)*. Secondary structure in panel E is from this study. Universal rRNA common core is
451 shown in blue lines (not shaded). ES9 is shown with a green line. ES39 is shown with a magenta line. H1
452 and ES39 helices are labeled on *Lokiarchaeota* (panel E). ES's and helices not present in the common
453 core are shaded in gray. Sizes of secondary structures are to scale. The numbering scheme of *Noller, et al.*
454 *(1981)* and *Leffers, et al. (1987)* were used to label the helices and ES's.
455

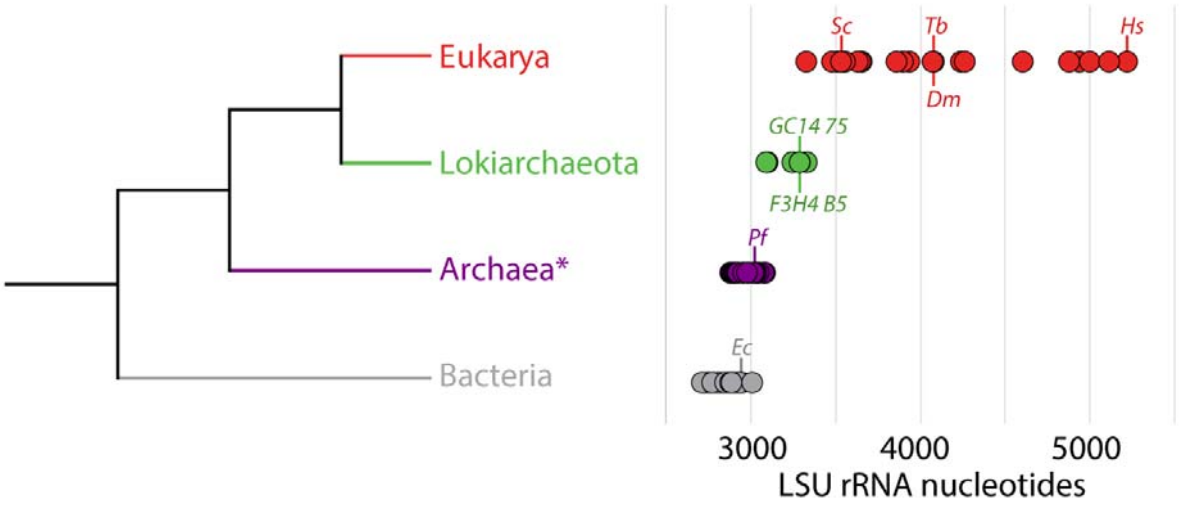


Figure 2. Length of LSU rRNA increases from Bacteria, to Archaea (excluding Asgard archaea), to Lokiarchaeota, to Eukarya. LSU rRNA lengths were obtained from the updated SEREB database. Abbreviations: *Ec, Escherichia coli; Pf, Pyrococcus furiosus; F3H4 B5, Lokiarchaeota F3H4_B5; GC14 75, Lokiarchaeota GC14_75; Tb, Trypanosoma brucei; Sc, Saccharomyces cerevisiae; Dm, Drosophila melanogaster; Hs, Homo sapiens.*

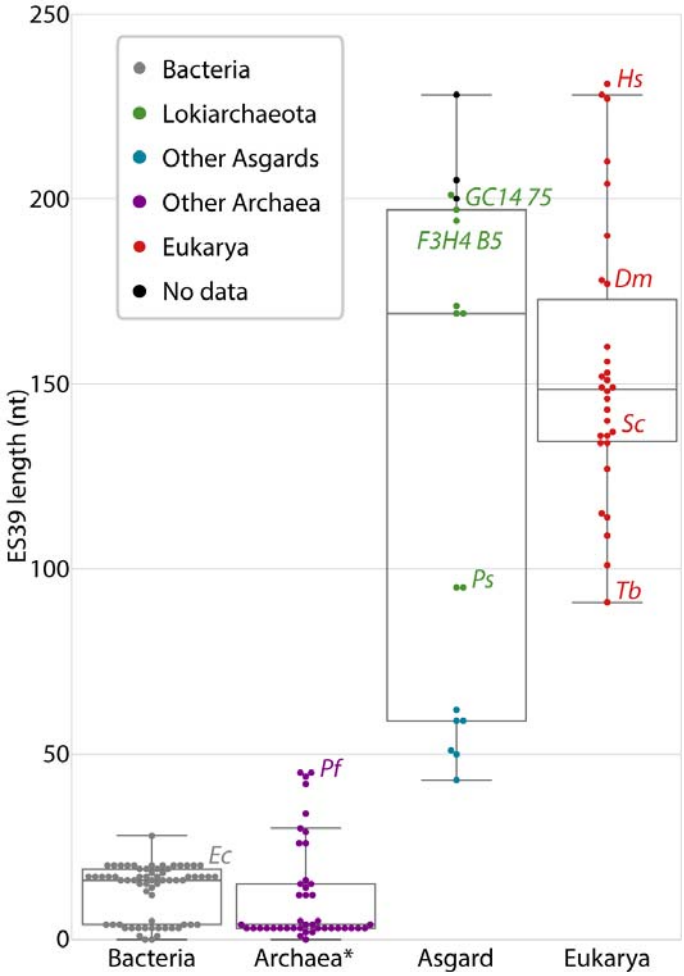
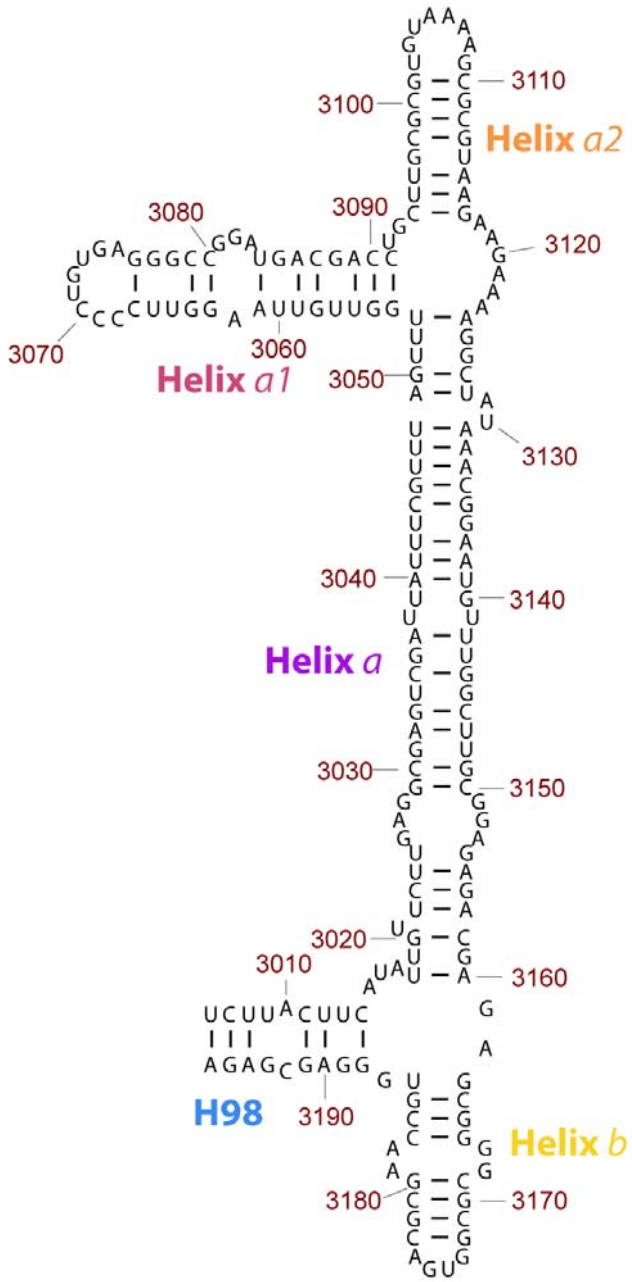


Figure 3: Distribution of ES39 lengths from the three domains of life, including Asgard archaea. The number of nucleotides calculated between alignment positions 4891 and 5123 (*H. sapiens* numbering) of the LSU alignment for each species (**supplementary dataset S1**). The box shows the quartiles of the dataset. Whiskers extend to show the rest of the distribution, except for points that are determined to be outliers using a function of the inter-quartile range. Bacteria sequences are gray, *Lokiarchaeota* sequences are green, other Asgard sequences are blue, other archaeal sequences are purple, eukaryotic sequences are red, sequences from metatranscriptomic contigs (**supplementary dataset S2**) for which there is no species determination are black. Abbreviations: *Ps: Prometheoarchaeum syntrophicum;* the rest are described in **Figure 2**.



475

476 **Figure 4**: **Secondary structure model of *Lokiarchaeota* F3H4_B5 ES39.** The secondary structure of
477 ES39 spans nucleotide positions 3006-3196 of the F3H4_B5 LSU rRNA sequence. Canonical base-
478 pairing positions are indicated with black lines. Helices are annotated with colored labels: blue – H98,
479 purple – Helix *a*, pink – Helix *a1*, orange – Helix *a2*, gold – Helix *b*. Figure was generated with
480 RiboVision (Bernier, et al. 2014).

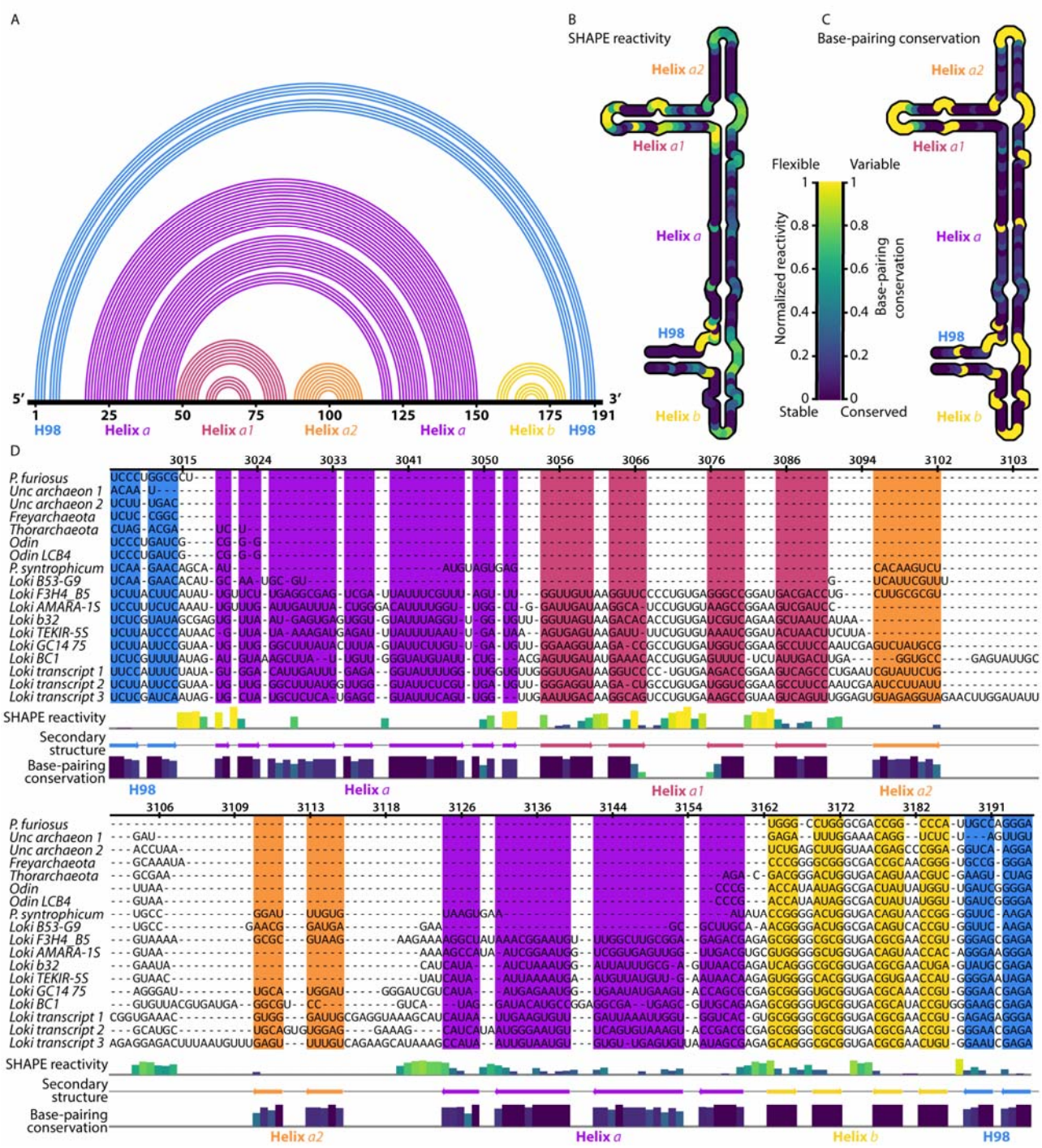
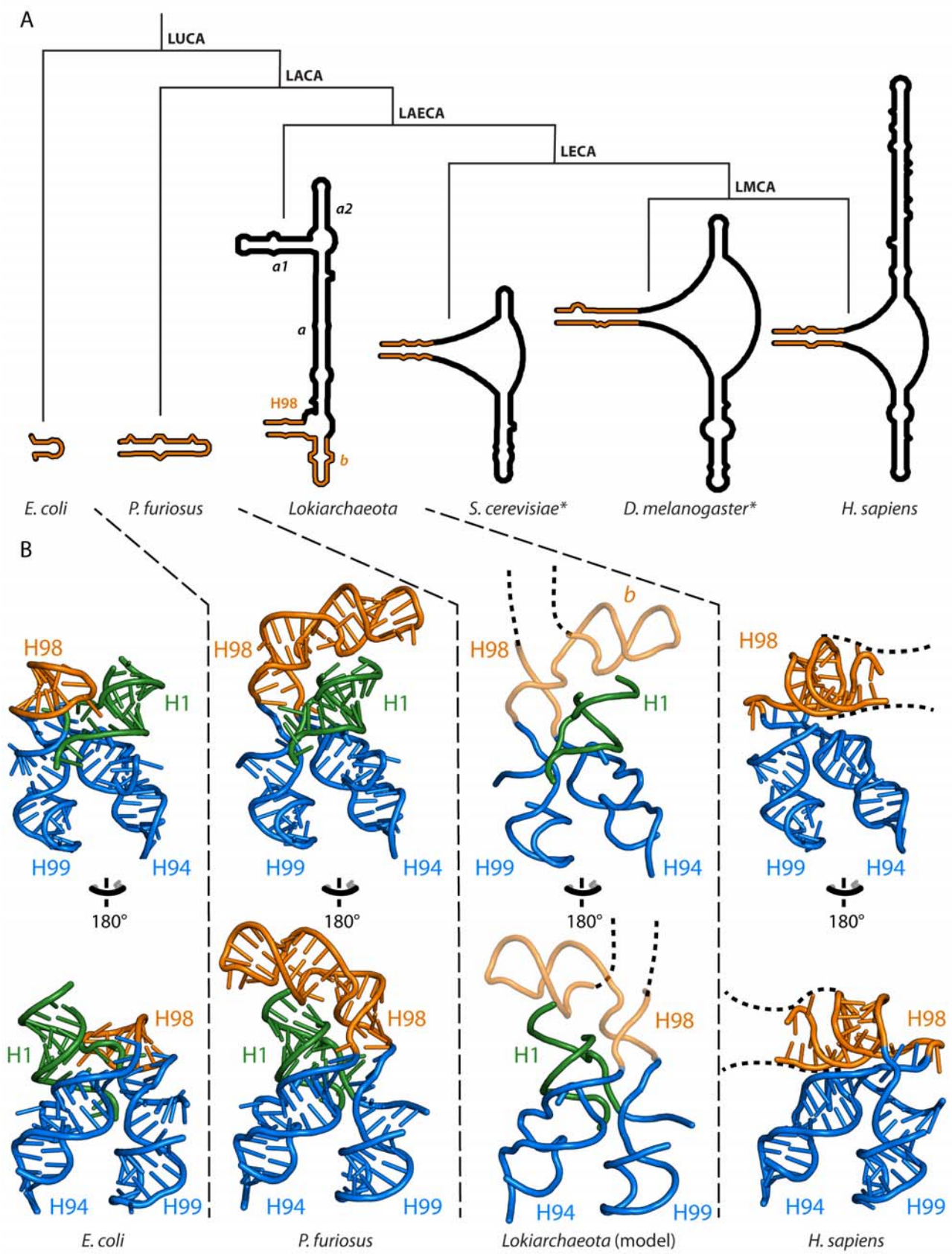


Figure 5: **Secondary structure of *Lokiarchaeota* F3H4_B5 ES39 from experiment and computation.** (A) 1-D topology map of base pairs. The primary sequence of ES39 is on the horizontal line, Arcs indicate base pairs. Each helix is a distinct color. (B) SHAPE reactivity for ES39 mapped onto the secondary structure. Darker color indicates less flexible (paired) rRNA. (C) Base pairing covariation within the Asgard superphylum mapped on the secondary structure. Darker color indicates covarying

(paired) rRNA across Asgard. Unpaired rRNA, for which no covariation data can be calculated, is gold. (D) SHAPE reactivity and base-pairing conservation mapped onto the ES39 MSA of Asgard sequences. *Lokiarchaeota* F3H4_B5 numbering was used. The secondary structure is indicated with colored arrows bellow the alignment and as colored background. SHAPE reactivity is indicated with a bar graph above the secondary structure annotation, colors of the bars are consistent with panel B. Base-pairing conservation is indicated with a bar graph bellow the secondary structure annotation; colors of the bars are consistent with panel C. Panel D was generated with Jalview (Waterhouse, et al. 2009). Helices are labeled with colored text in each panel; blue, H98; violet, helix *a*; pink, helix *a1*; orange, helix *a2*; yellow, helix *b*. Full sequence names and sequencing project identifiers are available in **supplementary dataset S2**. Both SHAPE reactivity and covariation are normalized.



498

Figure 6: **Secondary and tertiary structures of ES39 mapped on the tree of life** (A) secondary structures of *Escherichia coli* (Bacteria), *Pyrococcus furiosus* (Archaea), *Lokiarchaeota* F3H4_B5 (Archaea), *Saccharomyces cerevisiae* (Eukarya), *Drosophila melanogaster* (Eukarya), and *Homo sapiens* (Eukarya). Ancestral clades on the phylogenetic tree are labeled as LUCA: last universal common ancestor; LACA: Last Archaeal Common Ancestor; LAECA: Last Archaeal and Eukaryotic Common Ancestor; LECA: Last Eukaryotic Common Ancestor; LMCA: Last Metazoan Common Ancestor. H98 is highlighted in orange. (B) Three dimensional structures of ES39 and its neighborhood. H98 is orange, H1 is green, H94 and H99 are blue. The *P. furiosus* structure is used as model for the *Lokiarchaeota* structure. Likely position and direction of the *Lokiarchaeota* ES39 continuation is indicated with a black dashed line. Direction of eukaryotic ES39 continuation is indicated with a black dashed line. *3D structures for *S. cerevisiae* and *D. melanogaster* are identical to *H. sapiens* in this region and are not shown.

References:

Agmon I, Bashan A, Zarivach R, Yonath A. 2005. Symmetry at the active site of the ribosome: structural and functional implications. In. Biological Chemistry. p. 833.

Akıl C, Robinson RC. 2018. Genomes of Asgard archaea encode profilins that regulate actin. Nature 562:439-443.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403-410.

Andrews S. 2012. FastQC: a quality control tool for high throughput sequence data. Available at https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Anger AM, Armache JP, Berninghausen O, Habeck M, Subklewe M, Wilson DN, Beckmann R. 2013. Structures of the human and Drosophila 80S ribosome. Nature 497:80-85.

Athavale SS, Gossett JJ, Hsiao C, Bowman JC, O'Neill E, Hershkovitz E, Preeprem T, Hud NV, Wartell RM, Harvey SC, et al. 2012. Domain III of the T. thermophilus 23S rRNA folds independently to a near-native state. RNA 18:752-758.

Bachellerie JP, Michot B. 1989. Evolution of large subunit rRNA structure. The 3' terminal domain contains elements of secondary structure specific to major phylogenetic groups. Biochimie 71:701-709.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology : a journal of computational molecular cell biology 19:455-477.

Ben-Shem A, Jenner L, Yusupova G, Yusupov M. 2010. Crystal structure of the eukaryotic ribosome. Science 330:1203-1209.

Bernhart SH, Hofacker IL. 2009. From consensus structure prediction to RNA gene finding. Briefings in Functional Genomics 8:461-471.

Bernier CR, Petrov AS, Kovacs NA, Penev PI, Williams LD. 2018. Translation: The Universal Structural Core of Life. Mol Biol Evol. 35:2065-2076.

Bernier CR, Petrov AS, Waterbury CC, Jett J, Li F, Freil LE, Xiong X, Wang L, Migliozzi BL, Hershkovits E, et al. 2014. RiboVision suite for visualization and analysis of ribosomes. Faraday Discuss 169:195-207.

Bokov K, Steinberg SV. 2009. A hierarchical model for evolution of 23S ribosomal RNA. Nature 457:977-980.

Cai M, Liu Y, Zhou Z, Yang Y, Pan J, Gu J-D, Li M. 2018. Asgard archaea are diverse, ubiquitous, and transcriptionally active microbes. bioRxiv:374165.

Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Muller KM, et al. 2002. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics 3:2.

Clark CG, Tague BW, Ware VC, Gerbi SA. 1984. Xenopus laevis 28S ribosomal RNA: a secondary structure model and its evolutionary and functional implications. Nucleic Acids Res. 12:6197-6220.

Eme L, Spang A, Lombard J, Stairs CW, Ettema TJG. 2017. Archaea and the origin of eukaryotes. Nature Rev Microbiol. 15:711-723.

Fournier GP, Gogarten JP. 2010. Rooting the Ribosomal Tree of Life. Mol Biol Evol. 27:1792-1801.

Fournier GP, Poole AM. 2018. A Briefly Argued Case That Asgard Archaea Are Part of the Eukaryote Tree. Frontiers in Microbiology 9.

Fox GE. 2010. Origin and evolution of the ribosome. Cold Spring Harb Perspect Biol. 2:a003483.

Gerbi S. 1996. Expansion segments: regions of variable size that interrupt the universal core secondary structure of ribosomal RNA. Ribosomal RNA structure, evolution, processing, and function in protein biosynthesis 71:87.

Gilbert SD, Rambo RP, Van Tyne D, Batey RT. 2008. Structure of the SAM-II riboswitch bound to S-adenosylmethionine. Nat Struct Mol Biol. 15:177-182.

Gomez Ramos LM, Degtyareva NN, Kovacs NA, Holguin SY, Jiang L, Petrov AS, Biesiada M, Hu MY, Purzycka KJ, Arya DP, et al. 2017. Eukaryotic Ribosomal Expansion Segments as Antimicrobial Targets. Biochemistry 56:5288-5299.

Gonzalez IL, Gorski JL, Campen TJ, Dorney DJ, Erickson JM, Sylvester JE, Schmickel RD. 1985. Variation among human 28S ribosomal RNA genes. Proc Natl Acad Sci USA. 82:7666-7670.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072-1075.

Gutell RR. 1992. Evolutionary Characteristics of 16S and 23S rRNA Structure. In: H. Hartman KM, editor. The Origin and Evolution of the Cell. p. 243-309.

Gutell RR, Gray MW, Schnare MN. 1993. A compilation of large subunit (23S and 23S-like) ribosomal RNA structures: 1993. Nucleic Acids Res. 21:3055-3074.

Gutell RR, Larsen N, Woese CR. 1994. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. Microbiological reviews 58:10-26.

578 Hartman H, Fedorov A. 2002. The origin of the eukaryotic cell: a genomic investigation. Proc
579 Natl Acad Sci USA. 99:1420-1425.

580 Hassouna N, Mithot B, Bachellerie J-P. 1984. The complete nucleotide sequence of mouse 28S
581 rRNA gene. Implications for the process of size increase of the large subunit rRNA In higher
582 eukaryotes. Nucleic Acids Res. 12:3563-3583.

583 Hillis DM, Dixon MT. 1991. Ribosomal DNA: Molecular Evolution and Phylogenetic Inference.
584 The Quarterly Review of Biology 66:411-453.

585 Hsiao C, Lenz TK, Peters JK, Fang P-Y, Schneider DM, Anderson EJ, Preeprem T, Bowman JC,
586 O'Neill EB, Lie L, et al. 2013. Molecular paleontology: a biochemical model of the ancestral
587 ribosome. Nucleic Acids Res. 41:3373-3385.

588 Huang H, Suslov NB, Li NS, Shelke SA, Evans ME, Koldobskaya Y, Rice PA, Piccirilli JA.
589 2014. A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. Nat
590 Chem Biol. 10:686-691.

591 Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN,
592 Hernsdorf AW, Amano Y, Ise K, et al. 2016. A new view of the tree of life. Nature Microbiology
593 1:16048.

594 Imachi H, Nobu MK, Nakahara N, Morono Y, Ogawara M, Takaki Y, Takano Y, Uematsu K,
595 Ikuta T, Ito M, et al. 2019. Isolation of an archaeon at the prokaryote-eukaryote interface.
596 bioRxiv:726976.

597 Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7:
598 Improvements in Performance and Usability. Mol Biol Evol. 30:772-780.

599 Khatter H, Myasnikov AG, Natchiar SK, Klaholz BP. 2015. Structure of the human 80S
600 ribosome. Nature 520:640.

601 Klinge S, Voigts-Hoffmann F, Leibundgut M, Arpagaus S, Ban N. 2011. Crystal Structure of the
602 Eukaryotic 60S Ribosomal Subunit in Complex with Initiation Factor 6. Science 334:941-948.

603 Klinger CM, Spang A, Dacks JB, Ettema TJG. 2016. Tracing the Archaeal Origins of Eukaryotic
604 Membrane-Trafficking System Building Blocks. Mol Biol Evol. 33:1528-1541.

605 Kovacs NA, Petrov AS, Lanier KA, Williams LD. 2017. Frozen in Time: The History of
606 Proteins. Mol Biol Evol. 34:1252-1260.

607 Lapeyre B, Michot B, Feliu J, Bachellerie J-P. 1993. Nucleotide sequence of the
608 Schizosaccharomyces pombe 25S ribosomal RNA and its phylogenetic implications. Nucleic
609 Acids Res. 21:3322-3322.

610 Lavender CA, Lorenz R, Zhang G, Tamayo R, Hofacker IL, Weeks KM. 2015. Model-Free RNA
611 Sequence and Structure Alignment Informed by SHAPE Probing Reveals a Conserved Alternate
612 Secondary Structure for 16S rRNA. PLoS Comp Biol. 11:e1004126-e1004126.

Leffers H, Kjems J, Østergaard L, Larsen N, Garrett RA. 1987. Evolutionary relationships amongst archaebacteria: A comparative study of 23 S ribosomal RNAs of a sulphur-dependent extreme thermophile, an extreme halophile and a thermophilic methanogen. J Mol Biol. 195:43-61.

Lenz TK, Norris AM, Hud NV, Williams LD. 2017. Protein-free ribosomal RNA folds to a near-native state in the presence of Mg2+. RSC Advances 7:54674-54681.

Leshin JA, Heselpoth R, Belew AT, Dinman J. 2011. High throughput structural analysis of yeast ribosomes using hSHAPE. RNA biology 8:478-487.

Levitt M. 1969. Detailed molecular model for transfer ribonucleic acid. Nature 224:759-763.

Li Z, Guo Q, Zheng L, Ji Y, Xie Y-T, Lai D-H, Lun Z-R, Suo X, Gao N. 2017. Cryo-EM structures of the 80S ribosomes from human parasites Trichomonas vaginalis and Toxoplasma gondii. Cell Res. 27:1275-1288.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011 17:3.

McCutcheon JP, Moran NA. 2011. Extreme genome reduction in symbiotic bacteria. Nature Rev Microbiol. 10:13-26.

Melnikov S, Ben-Shem A, Garreau de Loubresse N, Jenner L, Yusupova G, Yusupov M. 2012. One core, two shells: bacterial and eukaryotic ribosomes. Nat Struct Mol Biol. 19:560.

Melnikov S, Manakongtreecheep K, Söll D. 2018. Revising the Structural Diversity of Ribosomal Proteins Across the Three Domains of Life. Mol Biol Evol. 35:1588-1598.

Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. 2005. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). J Am Chem Soc. 127:4223-4231.

Mestre-Fos S, Penev PI, Richards JC, Dean WL, Gray RD, Chaires JB, Williams LD. 2019a. Profusion of G-quadruplexes on both subunits of metazoan ribosomes. PLoS One 14:e0226177.

Mestre-Fos S, Penev PI, Suttapitugsakul S, Hu M, Ito C, Petrov AS, Wartell RM, Wu R, Williams LD. 2019b. G-Quadruplexes in Human Ribosomal RNA. J Mol Biol. 431:1940-1955.

Michot B, Bachellerie JP. 1987. Comparisons of large subunit rRNAs reveal some eukaryote-specific elements of secondary structure. Biochimie 69:11-23.

Moran NA. 2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell 108:583-586.

Narrowe AB, Spang A, Stairs CW, Caceres EF, Baker BJ, Miller CS, Ettema TJG. 2018. Complex Evolutionary History of Translation Elongation Factor 2 and Diphthamide Biosynthesis in Archaea and Parabasalids. Genome biology and evolution 10:2380-2393.

Ninio J, Favre A, Yaniv M. 1969. Molecular model for transfer RNA. Nature 223:1333-1335.

Noller HF, Kop J, Wheaton V, Brosius J, Gutell RR, Kopylov AM, Dohme F, Herr W, Stahl DA, Gupta R, et al. 1981. Secondary structure model for 23S ribosomal RNA. Nucleic Acids Res. 9:6167-6189.

Novikova IV, Hennelly SP, Sanbonmatsu KY. 2012. Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. Nucleic Acids Res. 40:5034-5051.

Olsen GJ, Woese CR. 1993. Ribosomal RNA: a key to phylogeny. The FASEB Journal 7:113-123.

Petrov AS, Bernier CR, Gulen B, Waterbury CC, Hershkovits E, Hsiao C, Harvey SC, Hud NV, Fox GE, Wartell RM, et al. 2014a. Secondary Structures of rRNAs from All Three Domains of Life. PLoS One 9:e88222.

Petrov AS, Bernier CR, Hsiao C, Norris AM, Kovacs NA, Waterbury CC, Stepanov VG, Harvey SC, Fox GE, Wartell RM, et al. 2014b. Evolution of the ribosome at atomic resolution. Proc Natl Acad Sci USA. 111:10251-10256.

Petrov AS, Gulen B, Norris AM, Kovacs NA, Bernier CR, Lanier KA, Fox GE, Harvey SC, Wartell RM, Hud NV, et al. 2015. History of the ribosome and the origin of translation. Proc Natl Acad Sci USA. 112:15396-15401.

Petrov AS, Wood EC, Bernier CR, Norris AM, Brown A, Amunts A. 2019. Structural Patching Fosters Divergence of Mitochondrial Ribosomes. Mol Biol Evol. 36:207-219.

Peyretaillade E, Biderre C, Peyret P, Duffieux F, Méténier G, Gouy M, Michot B, Vivarès CP. 1998. Microsporidian Encephalitozoon cuniculi , a unicellular eukaryote with an unusual chromosomal dispersion of ribosomal genes and a LSU rRNA reduced to the universal core. Nucleic Acids Res. 26:3513-3520.

Schnare MN, Damberger SH, Gray MW, Gutell RR. 1996. Comprehensive Comparison of Structural Characteristics in Eukaryotic Cytoplasmic Large Subunit (23 S-like) Ribosomal RNA. J Mol Biol. 256:701-719.

Smith TF, Lee JC, Gutell RR, Hartman H. 2008. The origin and evolution of the ribosome. Biology Direct 3:16.

Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJ. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521:173-179.

Spang A, Stairs CW, Dombrowski N, Eme L, Lombard J, Caceres EF, Greening C, Baker BJ, Ettema TJG. 2019. Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. Nature Microbiology 4:1138-1148.

Spitale RC, Crisalli P, Flynn RA, Torre EA, Kool ET, Chang HY. 2013. RNA SHAPE analysis in living cells. Nat Chem Biol. 9:18-20.

Stoddard CD, Gilbert SD, Batey RT. 2008. Ligand-dependent folding of the three-way junction in the purine riboswitch. RNA 14:675-684.

Tréhu AM, Bohrmann G, Rack FR, Torres ME, Richter C, Bangs NL, Barr SR, Borowski WS, Claypool GE, Collett TS, et al. 2003. ODP Leg 204; gas hydrate distribution and dynamics beneath southern Hydrate Ridge. JOIDES Journal 29:5.

Veldman GM, Klootwijk J, de Regt VC, Planta RJ, Branlant C, Krol A, Ebel JP. 1981. The primary and secondary structure of yeast 26S rRNA. Nucleic Acids Res. 9:6935-6952.

Ware VC, Tague BW, Graham Clark C, Gourse RL, Brand RC, Gerbi SA. 1983. Sequence analysis of 28S ribosomal DNA from the amphibian Xenopus laevis. Nucleic Acids Res. 11:7795-7817.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189-1191.

Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Jr., Swanstrom R, Burch CL, Weeks KM. 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature 460:711-716.

Wilkinson KA, Merino EJ, Weeks KM. 2005. RNA SHAPE Chemistry Reveals Nonhierarchical Interactions Dominate Equilibrium Structural Transitions in tRNAAsp Transcripts. J Am Chem Soc. 127:4659-4667.

Wilkinson KA, Merino EJ, Weeks KM. 2006. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat Protoc. 1:1610-1616.

Williams TA, Cox CJ, Foster PG, Szöllősi GJ, Embley TM. 2020. Phylogenomics provides robust support for a two-domains tree of life. Nat Ecol Evol. 4:138-147.

Woese CR, Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci USA. 74:5088-5090.

Woese CR, Magrum LJ, Gupta R, Siegel RB, Stahl DA, Kop J, Crawford N, Brosius J, Gutell R, Hogan JJ, et al. 1980. Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. Nucleic Acids Res. 8:2275-2293.

Yergeau E, Maynard C, Sanschagrin S, Champagne J, Juck D, Lee K, Greer CW. 2015. Microbial community composition, functions and activities in the Gulf of Mexico, one year after the Deepwater Horizon accident. Appl Environ Microbiol. 81:5855-5866.

714 Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, Vancaester E, Seitz
715 KW, Anantharaman K, Starnawski P, Kjeldsen KU, et al. 2017. Asgard archaea illuminate the
716 origin of eukaryotic cellular complexity. Nature 541:353.

717 Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic
718 Acids Res. 31:3406-3415.

719