ARJ

Information Processing Differences Between Archaea and Eukaraya—Implications for Homologs and the Myth of Eukaryogenesis

Change Tan, Division of Biological Sciences, University of Missouri, Columbia, Missouri.
Jeffrey Tomkins, Institute for Creation Research, 1806 Royal Lane, Dallas, Texas.

Abstract

In the grand schema of evolution, a mythical prokaryote to eukaryote cellular transition allegedly gave rise to the diversity of eukaryotic life (eukaryogenesis). One of the key problems with this idea is the fact that the prokaryotic world itself is divided into two apparent domains (bacteria and archaea) and eukarya share similarities to both domains of prokaryotes while also exhibiting many major innovative features found in neither. In this article, we briefly review the current landscape of the controversy and show how the key molecular features surrounding DNA replication, transcription, and translation are fundamentally distinct in eukarya despite superficial similarities to prokaryotes, particularly archaea. These selected discontinuous molecular chasms highlight the impossibility for eukarya having evolved from archaea. In a separate paper, we will address alleged similarities between eukarya and bacteria.

Keywords: First eukaryotic common ancestor, FECA, Last eukaryotic common ancestor, LECA, Eukaryogenesis, bacteria, archaea, eukaryote, DNA replication, transcription, translation, molecular evolution

Disclaimer: Regarding author C. Tan, the opinions expressed in this article are the author's own and not necessarily those of the University of Missouri.

Introduction

Eukarya are organisms with cells much larger than prokaryotes that possess nuclei and other membrane enclosed intracellular organelles. Thus, many of their processes are highly compartmentalized and more complicated than those of the most complex prokaryotes. In fact, a typical eukaryotic cell is about a thousand times larger in volume than a typical bacterial or archaeal cell and a fundamental eukaryote-prokaryote dichotomy clearly exists in regards to intracellular organization, complexity, and innovation. Besides, eukarya themselves are highly diverse and comprise three kingdoms of multicellular life that include plants, animals, and fungi. They also comprise a diverse array of unicellular organisms with extremely complex genomic features called protists.

In the grand evolutionary paradigm, the origin of the eukaryotic cell represents one of the great mysteries and key hypothetical transitions of life that is alleged to have occurred over one billion years ago—termed eukaryogenesis. Fossils offer little support to the eukaryogenesis model as one-celled eukarya from alleged strata of this age are already incredibly diversified—exhibiting complicated cellular innovations typical of extant species (Knoll et al. 2006). Of course, such fossils offer little information as to the specific nature of their cellular machinery

and genetic systems. Thus, the purely hypothetical field of molecular paleontology has arisen that seeks to infer evolutionary events from the study of extant genomes and the genes they encode.

One of the key problems with the whole idea of eukaryogenesis is the fact that the prokaryotic world itself is divided into two apparent domains (bacteria and archaea) and extant one-celled eukarya share molecular similarities to both domains of prokaryotes while also exhibiting major innovative features found in neither (Lake et al. 2009; O'Malley and Koonin 2011; Zimmer 2009). Archaea and bacteria share extensive similarities for many metabolic genes, but differ significantly in the types of genes that encode the cellular machinery for information transfer processes such as those associated with DNA replication, transcription, and translation. The paradox is that eukarya share considerably more similarity in their information processing proteins to archaea while more similarity in their metabolic proteins to bacteria.

Molecular Discontinuity in the Three Domains of Life

Not surprisingly, genome-wide sequence comparisons have found that the majority of eukaryotic genes are unique to the eukaryotic domain itself, without identifiable homologs in the other two domains of life—bacteria and archaea,

each of which includes many diverse single-celled microorganisms (Dagan and Martin 2006; Esser et al. 2004). Interestingly, of the eukaryotic genes that have prokaryotic homologs, the vast majority either only have bacterial homologs, or are more similar in sequence to bacterial genes than to archaeal genes (Dagan and Martin 2006; Esser et al. 2004). For example, in a comparison of human proteins to proteins from 224 prokaryotic genomes (24 archaea and 200 bacteria), only 5833 human proteins, about a quarter of the human protein-coding genes, have homologs in these prokaryotes. Of these 5833 proteins, 48% have homologs in bacteria only, 14% have homologs in archaea only, and 80% have greater sequence identity with bacterial homologs, whereas 15% are more similar to archaeal homologs (Dagan and Martin 2006). Consistently, Nasir and colleagues (Nasir, Kim, and Caetano-Anolles 2014) found that eukarya share many more protein domains, defined as fold families (FFs), with bacteria than with archaea (fig. 1A). These authors analyzed FFs from 420 organisms, including 48 archaea, 239 bacteria, and 133 eukarya. Of the 2397 FFs identified, 20% are found in all three domains of life, 23% in only two, and 57% in only one. Of the domain-specific FFs, 758 (31.6%) belong to eukarya, 522 (21.8%) to bacteria, and 89 (3.7%) to archaea. The FFs shared between eukarya and bacteria with the exclusion of archaea are ten times more than that between eukarya and archaea (414 vs 40).

Strikingly, even though the vast majority of the eukaryotic genes that have prokaryotic homologs either only have bacterial homologs, or are more similar to bacterial genes than to archaeal genes, the molecular machines for eukaryote information processing are much more similar to those of archaea than those of bacteria, although the archaea version is much simpler along with innovations unique to archaea (Allers and Mevarech 2005; Aves, Liu, and Richards 2012; Ishino and Ishino 2012; Raymann et al. 2014; Rivera et al. 1998). For instance, except for the universal ribosomal proteins that exist in all three domains of life, only archaea and eukarya share additional ribosomal proteins, but not between archaea and bacteria or between eukarya and bacteria (fig. 1B). Homologs for eukaryotic Orc (origin recognition complex) and helicase MCM (Mini-chromosome maintenance protein) in DNA replication have been identified in archaea but not in bacteria. The archaeal RNA polymerase (RNAP) has similar composition with that of eukarya (Huet et al. 1983), and like eukarya, archaea have many translation initiation factors (Allers and Mevarech 2005; Benelli, Maone, and Londei 2003). Thus, it is not surprising that about half of the FFs unique to eukarya and archaea are involved in information processing, including DNA replication, transcription, and translation (Nasir, Kim, and Caetano-Anolles 2014).

The higher similarities between archaea and eukaryote information processing molecules than that between bacteria and eukarya have prompted some to propose that archaea are a closer ancestor of eukarya than bacteria (Gribaldo et al. 2010). Therefore, a close examination of the archaeal information processing machinery will be very informative to our understanding of the molecular mechanisms in the three domains of life and the
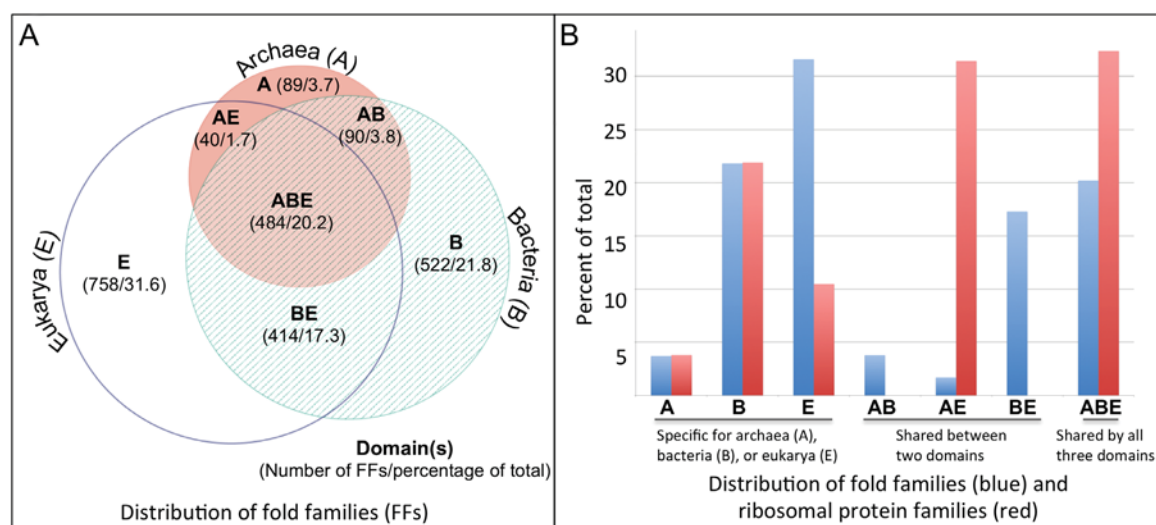

Fig. 1. A. A Venn diagram of protein FFs in the three domains of life, adapted from Fig. 1A of (Nasir, Kim, and Gaetano-Anolles 2014). **B**. A comparison of the distribution of fold families and numbers of ribosomal proteins in the three domains of life. FFs are the same as panel A, while ribosomal protein numbers are according to Lecompte et al. (2002) and Yutin et al. (2012). Abbreviations: A: archaea, B: bacteria, E: eukarya, AB: shared between archaea and bacteria, AE: shared between archaea and eukarya, BE: shared between bacteria and eukarya, ABE: shared by three domains.

origin of eukarya. Keep in mind that there are many diverse kinds of organisms in each domain, so that no molecular machine in particular is the same in all organisms and mosaics of systems and designs are often found within the same domain (Aves, Liu, and Richards 2012; Costa, Hood, and Berger 2013; Mardanov and Ravin 2012; Raymann et al. 2014; Sarmiento et al. 2014; Siddiqui, On, and Diffley 2013).

Archaea were recognized as a unique domain of life based on the sequence comparisons of ribosomal RNA (rRNA) of various organisms by Woese and colleagues (Woese and Fox 1977). Archaea differ from the other two domains of life—bacteria and eukarya—not only in their archaea-specific signatures in certain regions of rRNAs, but also in their cell membranes, which are composed of lipids made of ether, unlike bacteria or eukarya whose membrane lipids are made of ester (Gutell et al. 1985; Woese et al. 1983; Woese, Kandler, and Wheelis 1990). A simple comparison of the features found in archaea and bacteria and eukarya can be found in Table 1 (Aves, Liu, and Richards 2012; Cavicchioli 2011; Raymann et al. 2014). In fact, as stated by Woese and colleagues, "for every well characterized molecular system there exists a characteristic eubacterial, archaebacterial, and eukaryotic version" (Woese, Kandler, and Wheelis 1990).

Archaea are similar to bacteria in many aspects. Like bacteria, archaea do not have nuclei, and are thus prokaryotes. Archaea also lack other membrane-bound organelles, including mitochondria and chloroplasts. Archaeal genomes are small and circular like those of bacteria. No spliceosomal introns have been found in archaea. Like bacteria, archaea also lack the machinery to synthesize eukaryotic telomeres and to splice spliceosomal introns, two processes essential for the survival of eukarya. This shortage of higher level eukaryotic complexity does not hurt archaea in any way because they have no need of these systems. However, the lack of these

systems, including any transitional forms for them, creates an unbridgeable chasm between prokaryotes and eukarya in the grand evolutionary paradigm. In this article, we wish to show that one does not need to search far to find that there exists an unbridgeable chasm between prokaryotes, particularly archaea and eukarya, even in the places where they most resemble each other.

Archaeal Gene Replication, Transcription, and Translation and the Myth of Homologs
A. DNA replication in archaea

We will only compare some aspects of DNA replication initiation and elongation in archaea and eukarya since we have already said that we would ignore the fact that archaea has a systemic lack of the telomere synthesizing machinery, which is necessary to replicate the ends of linear eukaryotic chromosomes.

A.1 Initiation of DNA replication
Recognition of replication origins

The first step in DNA replication is the recognition of the origin of replication by origin recognition proteins. Archaea differ from eukarya both in the identification of the origin of replication, including the DNA sequence and/or the local DNA structure, and the recognition of the origin of replication. An archaeal genome can have one or several origins of replication. Like bacterial origins of replication, archaeal origins of replication have specifically defined sequences. In contrast, the eukaryotic origins of replication are numerous and mostly determined by the structure and context of their chromosomes. In bacteria, origins of replication are recognized by bacteria-specific DnaA. In eukarya, origins of replication are recognized by a heterohexamer Orc1-6. Each member of the hexamer is required for genomic DNA replication and the viability of yeast (a one-celled eukaryote)—none can be substituted with

Table 1. A comparison of a few traits of bacteria, archaea, and eukarya.

| Trait	Bacteria	Archaea	Eukarya
Carbon linkage of lipids	Ester	*Ether*	Ester
Phosphate backbone of lipids	Glycerol-3-phosphate	*Glycerol-1-phosphate*	Glycerol-3-phosphate
Metabolism	Bacterial	Bacterial-like	Eukaryotic
Nucleus	No	No	Yes
Organelles	No	No	Yes
Spliceosomal introns	No	No	Yes
Telomeres	No	No	Yes
Chromosome shape	Mostly circular	Circular	Linear
DNA replication	Bacterial	Eukaryotic-like	Eukaryotic
Transcription	Bacterial	Eukaryotic-like	Eukaryotic
Translation	Bacterial	Eukaryotic-like	Eukaryotic
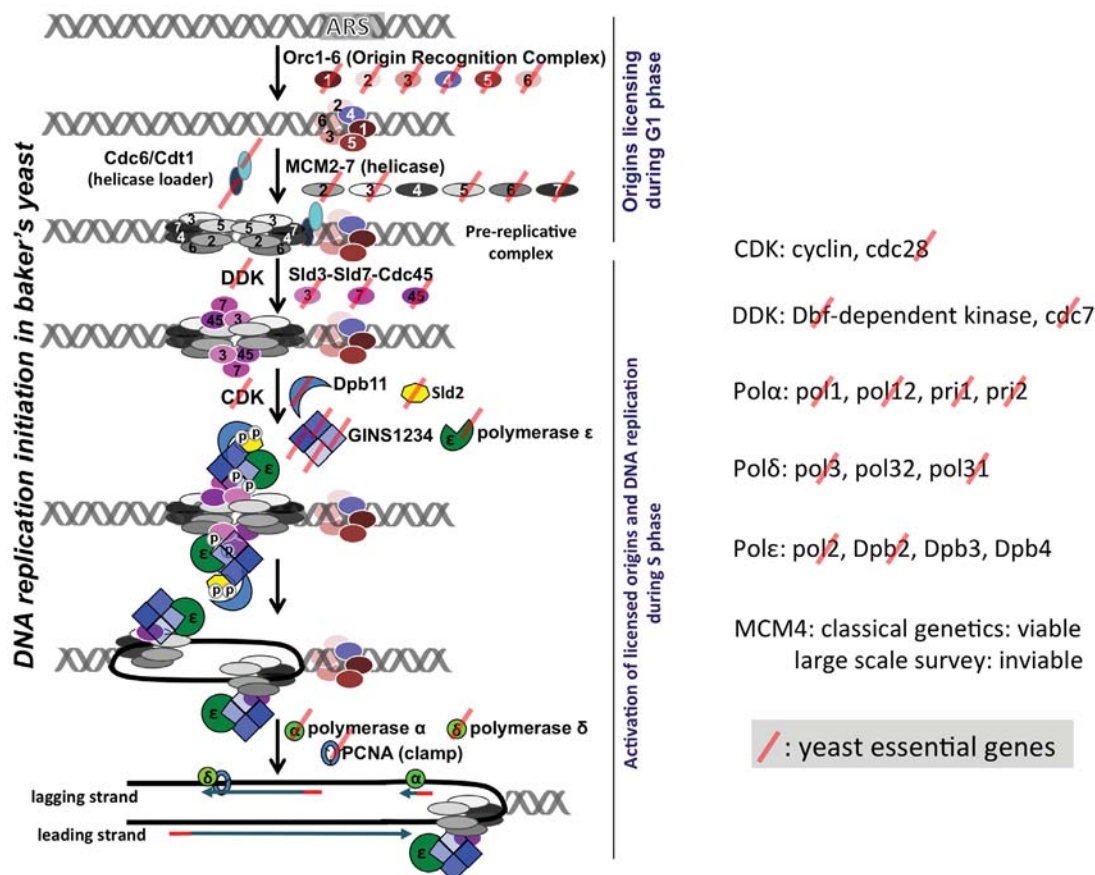
Fig. 2. Most yeast DNA replication genes are essential for the viability of the yeast cell. Left: A schematic view of the steps involved in DNA replication initiation in the budding yeast *Saccharomyces cerevisiae*. Right: Components of some of the multi-subunit factors of the replication pathway. The essential genes are marked with semi-translucent red slashes. The essentiality of genes are according to the yeast genome database http://www.yeastgenome.org.

another (http://www.yeastgenome.org) (fig. 2). Orc1-5 contain AAA+ (ATPases associated with various cellular activities) or AAA-like domains and winged-helix domains (WHDs). Orc6 is unrelated to Orc1-5 in sequence. Another AAA+ protein Cdc6 (cell division cycle 6) interacts with Orc1-6 and, together with Cdt1, recruits helicase MCM2-7 to the origin. Most archaea contain one to three genes that have some

sequence similarity to Cdc6 and the C-terminus of Orc1, thus are named Orc1/Cdc6 (fig. 3), though archaea lacking Orc1/Cdc6 also exist (Raymann et al. 2014; Sarmiento et al. 2014). Eukaryotic Orc1-6 bind their origins of replication as a heterohexamer, while archaeal Orc1/Cdc6 binds as a monomer or dimer (Dueber et al. 2007; Gaudier et al. 2007; Grainge et al. 2006; Ishino and Ishino 2012; Wigley 2009).
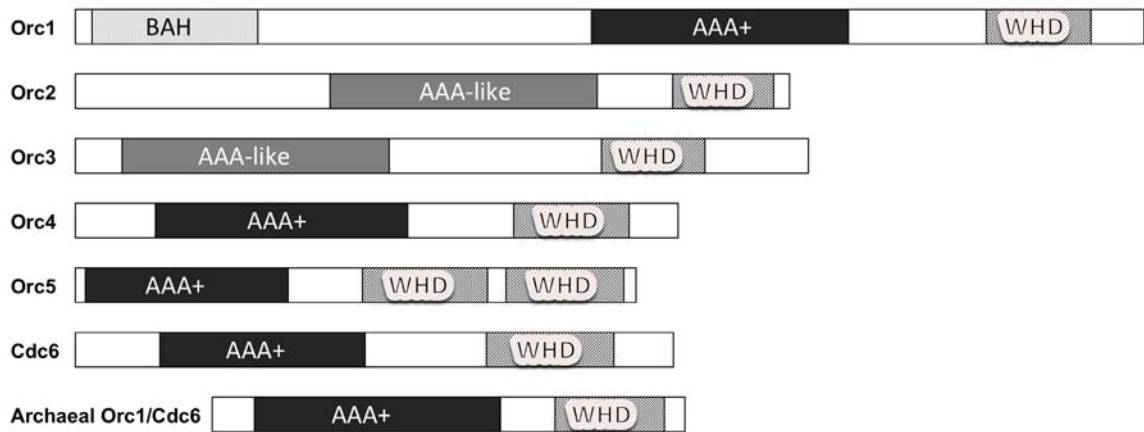

Fig. 3. A schematic drawing of origin recognition proteins of eukaryotes and archaea. AAA+: ATPases associated with various cellular activities domain, AAA-like: AAA+-like, WHD: winged-helix domain. Adapted from Fig. 3 of Costa, Hood, and Berger (2013).

Note that archaeal Orc1 does not have the eukaryotic specific Orc1 N-terminal extension (fig. 3). This extension contains a conserved bromo-adjacent homology (BAH) domain and is critical for binding multiple factors, including transcriptional silencing factors and histones in eukarya (Costa, Hood, and Berger 2013; Duncker, Chesnokov, and McConkey 2009). The BAH domain is important for loading the ORC complex onto chromatin in human cells and mutations in this region have been linked to primordial dwarfism (Costa, Hood, and Berger 2013; Kuo et al. 2012).

Strangely, Orc1 functions differently even in different eukaryotic organisms. For example, Orc1 is a potent transcription repressor in yeast while in the plant *Arabidopsis*, Orc1 functions as a transcription activator, probably because it contains a PHD domain that is absent in yeast and human Orc1 (Sanchez Mde and Gutierrez 2009). It is quite likely that the archaeal Orc1 also functions differently from its eukaryotic homologs (we use the term homolog in the sense that two proteins share some sequence similarity, not in the sense of two molecules sharing a common ancestor). Such organism-specific protein sequence extensions in homologs are not peculiar to Orc1, instead it is a very common phenomenon. Unfortunately, these extensions are simply ignored when people talk about the homologous relationships of molecules, and homologous proteins are often assumed to have shared a common ancestor in deep time. However, these extensions are either there or not, and their existence is critical for their interaction with organism-specific factors and vital for the life of that organism.

Since even an organism-specific protein extension may be crucial for the life of that organism, it is only logical that an organism-specific gene will be important for the organism containing it. And indeed there are many such examples. For example, the vast majority of the molecules involved in yeast DNA replication, many of which do not have archaeal homologs, are indispensable for the basic viability of yeast, including Orc1 to Orc6 mentioned above (fig. 2). On the other hand, there are also many times that deletion or loss-of-function mutation of a homolog appears to have no visible or detectable effect on its host organism or a specific cellular process. This may be the result of several reasons, including 1) the molecule does not play a role in the process analyzed but plays a role in a different process that is not known or not analyzed, 2) the molecule does play a role in the process analyzed but the analysis is not sensitive enough to detect its impact, 3) the molecule plays a role in the process analyzed but under different experimental conditions, 4) the molecule plays a role in the process analyzed but there exist a substitute (or substitutes) to compensate its absence. The last case is termed functional redundancy. A molecule that functions redundantly provides for a level of fault tolerance and robustness in the cell, a hallmark of designed systems. In light of this idea, the existence of two or more molecules that can function redundantly may point to the necessity of the specific function. Evolutionary theory, however, considers redundant molecules to largely be unnecessary and has no logical explanation for their existence since selective pressures upon the backup version to maintain its functional status would be low to none.

It is interesting that archaea have some homologs of the eukaryotic origin recognition complex and not of the bacterial type, setting archaea a dramatic step closer to eukarya in DNA replication. Simultaneously, we see that the archaeal DNA replication system is very different than that of eukarya, much more complexity is required in the latter. For example, each of the six subunits of the Orc complex is required in eukaryotic DNA replication, none can be substituted by any other, while one archaeal Orc1 may be enough for archaeal DNA replication and even this archaeal Orc1 lacks a functionally important eukaryote-specific extension and thus would be unable to perform the functions of the eukaryotic Orc1. In addition, many archaea do not have Orc1, and thus must use a different mechanism to replicate their genomic DNA.

DNA Helicase activity

DNA helicases are essential enzymes functioning in DNA replication by separating double-stranded DNA into single strands so that each strand can be copied. One of the important components of eukaryotic DNA helicase function is the MCM2-7 heterohexamer, which is made of six different proteins, each belonging to a distinct protein family (Bochman and Schwacha 2009). MCM stands for mini chromosome maintenance complex, which has a role in both the initiation and the elongation phases of DNA replication in the replication fork. Note that like the individual components of Orc1-6, each member of the MCM2-7 is essential for yeast DNA replication. Deletion of MCMs2, 3, 5, 6, 7 cause lethality in yeast (http://www.yeastgenome.org) (fig. 2).

Most archaea examined so far have one MCM gene, and only a homohexamer is formed during DNA replication in those organisms. Archaeal members of the *Methanococcales* group have two to eight copies of MCM. How MCMs in these organisms function is not clear. Nonetheless, all archaeal MCMs cluster into a unique family in a sequence comparison, which is almost equally related to, and distinct from, each of the six families of eukaryotic MCM2-7 (Bochman and Schwacha 2009).

Binding of the MCM at the origin is necessary but not sufficient for the formation of a replication bubble or unwinding of the double stranded DNA in eukarya. The process from origin recognition by Orc1-6 to the binding of MCM2-7 is called origin licensing in eukarya. Licensed replication origins need to be activated by cyclin-dependent kinases, resulting in the association of heterotetrameric GINS (go-ichi-ni-san) (made of Sld5, Psf1, 2, 3) and Cdc45 and the formation of the active helicase CMG (Cdc45-MCM-GINS). Archaea vary greatly in their activation of MCM, although no Cdc45 homolog has been identified in archaea and many archaea do not have a GINS homolog, some archaea have a gene GINS15, which is homologous to Sld5 and Psf1, and/or a GINS23 gene, which is homologous to Psf2 and Psf3. Note that not all archaeal GINS function the same in different species—GINS stimulates the MCM helicase in *Thermococcus kodakarensis* and *Pyrococcus furiosus* but not in *Sulfolobus solfataricus* (Sarmiento et al. 2014). Furthermore, like the situation with archaeal Orc1, there are many archaea species that do not have any homologs of GINS (Raymann et al. 2014; Sarmiento et al. 2014).

Most importantly, archaea lack proteins homologous to those eukaryotic proteins required to activate the licensed eukaryotic origins of replication, including cyclin-dependent kinase CDK and Dbf4 and Dbf4-dependent kinase DDK, all essential for the viability of yeast (Bochman and Schwacha 2009; Tye 2000) (fig. 2). Therefore, the archaeal helicases alone are not enough to open up the double-stranded DNA in eukarya. Even if they could license the eukaryotic origins of replication, the licensed origins could not be activated and no DNA replication would occur and no eukaryote could survive or propagate. Thus, this lack of genes to activate the licensed origins of replication constitutes an unbridgeable chasm between archaea and eukarya. Alternatively, archaea helicases are able to open up the double-stranded DNA in eukarya but the necessary cell-cycle dependent regulation of DNA replication cannot occur, resulting in eukaryotic cell death.

Primase

DNA replication starts with the synthesis of an RNA primer by DNA primase in all domains of life because DNA polymerases are incapable of de novo DNA synthesis. DNA primase is a type of RNA polymerase that creates an RNA primer that DNA polymerase uses to replicate single stranded DNA. The RNA primer is later removed by exo- and/or endonucleases.

The single subunit protein DnaG serves as the primase in bacteria, while eukaryotic primase is composed of two subunits PriS and PriL in a complex

with the eukaryotic-specific DNA polymerase α and its accessory B subunit. Homologues of PriS and PriL have been identified in archaea, but not DNA polymerase α. However, the Pol α is required in eukarya to synthesize a short DNA fragment (10~30 nucleotides) after the RNA primer and only after Pol α has done its job, other eukaryotic polymerases can then function and finish the process of DNA replication. Therefore, the lack of Pol α in archaea is a dramatic barrier for any archaeal cell to be able to evolve into a eukaryote.

Furthermore, archaeal homologues identified via sequence comparison function differently compared to their bacterial or eukaryotic counterparts. For example, eukaryotic PriS and PriL synthesize RNA primers. However, the archaeal homologue *Pyrococcus furiosus* PriS synthesizes long DNA fragments in vitro, PriL decreases its DNA polymerase activity and increases its RNA polymerase activity (Liu et al. 2001). Homologs of bacterial DnaG have also been found in archaea, but instead of functioning as a primase, they function in RNA degradation in *Thermococcus kodakarensis* and *Sulfolobus solfataricus* (Evguenieva-Hackenberg et al. 2003; Walter et al. 2006). Thus, judging the functions of a molecule based on sequence similarity alone can be misleading.

A.2 Elongation and DNA polymerases

Seven families of DNA polymerases have been identified based on sequence comparison: A, B, C, D, E, X, and Y (Ishino and Ishino 2012). Family C is unique for bacteria; family D and E are unique for archaea; family X is unique for eukarya. Variants of family Y members have been found in all kingdoms of life. The major replicative DNA polymerase in *E. coli* is a family C polymerase, Pol III, while in eukarya family B members (Pol α, Pol δ, Pol ε) are responsible for DNA replication. Family B polymerase members have been identified in all archaea, and were naturally proposed as the replicative DNA polymerase in archaea because DNA replication in archaea is more similar to that in eukarya than that in bacteria and archaea do not have Pol C, the principle DNA polymerase that bacteria use to replicate their genomes. Thus, it came as a surprise that DNA polymerase D, a family unique to archaea, were discovered as the enzyme used for DNA replication in archaea *Thermococcus kodakarensis* and *Methanococcus maripaludis* and Pol D, but not Pol B, is essential for the viability of *M. maripaludis* (Cann et al. 1998; Cubonova et al. 2013; Sarmiento, Mrazek, and Whitman 2013).

With the knowledge of DNA replication in archaea and bacteria, it defies evolution that eukarya use a special eukaryote-specific polymerase, Pol α, in their

DNA replication. While the two principle enzymes for synthesizing DNA during eukaryotic DNA replication are Pol δ and Pol ε, neither Pol δ nor Pol ε is capable of performing its task until Pol α has synthesized a short DNA fragment after the RNA primer synthesized by the primase. The fact that no homolog of Pol α has been identified in bacteria or archaea along with other missing eukaryotic-like enzymes, prompted Leipe and colleagues to propose that the foundational process of DNA replication had at least two independent origins (Leipe, Aravind, and Koonin 1999).

DNA clamps or sliding clamps work as part of the DNA polymerase complex to greatly increase the processivity and functionality of replication. Here is a rare case where the archaeal version is more complicated than that of eukarya. In eukarya, the sliding clamp or PCNA is a homotrimeric ring. In contrast, the majority of *Crenarchaeota* have multiple PCNA homologs, which are capable of forming heterotrimeric rings, although most *Euryarchaeota* possess only a single PCNA homolog (Ishino and Ishino 2012).

Taken together, archaeal DNA replication machinery is unique to the archaea domain. Not only is the archaeal machinery distinct from that of bacteria, making the DNA replication machines un-exchangeable between archaea and bacteria, but archaeal DNA replication machinery is insufficient as a natural evolutionary precursor to the eukaryotic DNA replication machinery. The archaeal DNA replication machinery would fail in every step of eukaryotic DNA replication, including initiation, elongation, and termination, due to the lack of eukaryote-specific proteins required. Furthermore, as indicated with the cases of DnaG and Pol B, functional prediction based on partial sequence homology alone can often be misleading. Finally, partial sequence homology comparison often excludes the organism-specific regions of the proteins, such as the N-terminus of Orc1. Yet, these regions are vital for the organism to replicate its DNA.

B. Transcription

Like DNA replication, some aspects of archaeal transcription are similar to that of bacteria while some are similar to that of eukarya. The composition of archaeal RNA polymerase (RNAP) is more similar to that of eukarya, consisting of more than twice the number of protein subunits than bacterial RNAP. On the other hand, like bacteria, archaea use one RNAP to synthesize all RNAs. In contrast, eukarya have three different types of RNAP. Many archaeal genes are also organized into operons, thus all genes in the same operon are transcribed together as a single unit, similar to bacterial genes. In addition,

archaeal transcripts are not capped at the 5′ end, polyadenylated at the 3′ end, or spliced like those in eukarya. Furthermore, archaeal regulation of transcription is also very bacteria-like, though many transcription factors are unique to archaea. Here we focus on the archaeal RNAP and general transcription factors, part of the archaeal transcription machinery that is most similar to that of eukarya. In so doing, it will become obvious that archaeal and eukaryotic transcription machineries cannot substitute for each other, even in the places where they most closely resemble each other.

B.1 RNA polymerase

One of the most advertised characteristics of archaeal RNAP is that it contains almost a one-to-one counterpart presence of the 12 subunits of the eukaryotic RNA pol II with the following "minor" exceptions (fig. 4 and appendix A). First, the archaeal subunit A′ is homologous to the N-terminus of Rpb1, while the subunit A″ is homologous to the C-terminus of Rpb1. Second, no homolog of the Rpb9 subunit of eukaryotic RNA pol II has been identified in archaea while homologs of Rpb8 have been identified in some archaea. Third, some archaeal RNAP contains an additional protein Rpo13 that does not have a eukaryotic homolog (Jun et al. 2011).

A sequence alignment shows that archaeal homologues are often shorter than their corresponding eukaryotic counterparts, with many deletions or truncations, and a few archaea-specific insertions (fig. 4). As the eukaryote-specific N-terminal extension of Orc1, the eukaryote-specific extensions in the various subunits of RNA pol II appear to be critical to the survival of eukarya.

The best studied eukaryotic-specific extension is the C-terminal domain (CTD) of the Rpb1 (Egloff, Dienstbier, and Murphy 2012; Heidemann et al. 2013; Hsin and Manley 2012; Hsin, Xiang, and Manley 2014; Meinhart et al. 2005; Napolitano, Lania, and Majello 2014; Yang and Stiller 2014). The CTD is made of repeats of a heptad. The number of repeats typically correlates with the complexity of the organism; yeast has 26 repeats, while humans have 52 (fig. 5A). The CTD is linked to the core of Rpb1 with a flexible linker and is long enough to reach anywhere on the RNA polymerase surface. Each conserved repeat is made of seven residues YSPTSPS (fig. 5B). The non-conserved repeats deviate from the conserved at positions 2, 3, 4, 5, and 7, mostly at position 7 (fig. 5B). Each of the residues of the repeats can be modified post-translationally in a transcription cycle-dependent manner (fig. 5C) (Heidemann et al. 2013). Without any phosphorylation of CTD, Pol II is inactive in initiating transcription, although the formation of the initiation complex can only
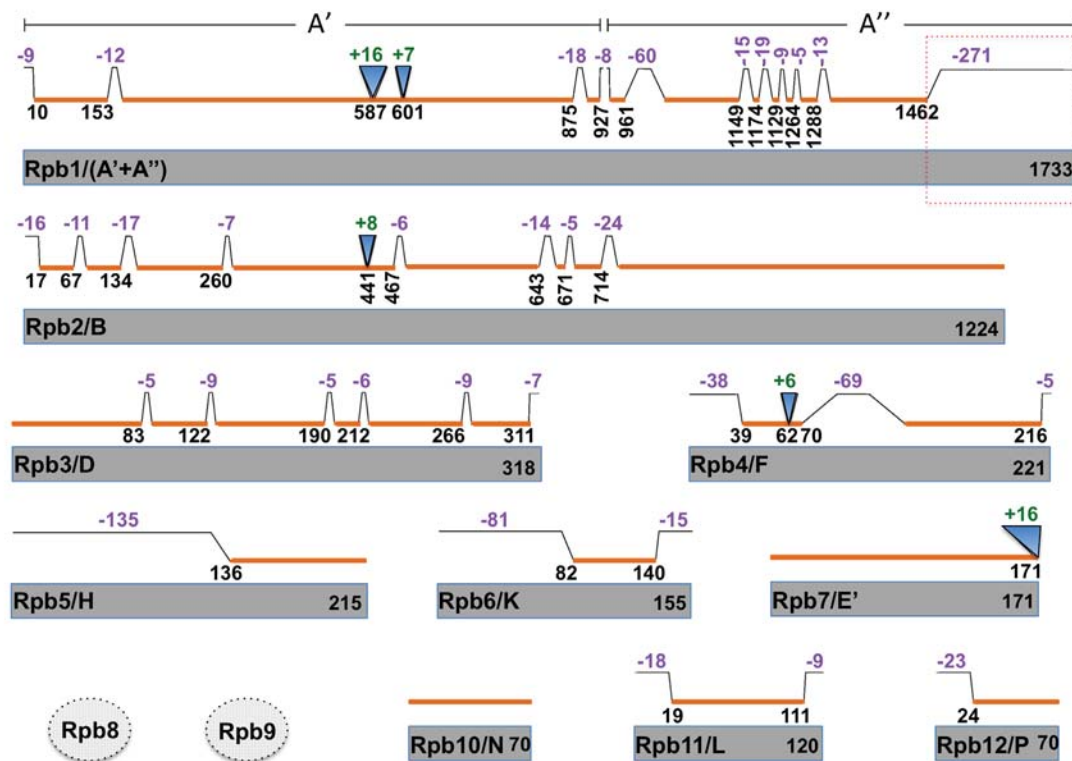
Alignments of *S. cerevisiae* RNA Pol II with *P. furiosus* RNA Polymerase

Fig. 4. Comparisons of the RNA polymerase II (RNA Pol II) of yeast *S. cerevisiae* with those of archaea *P. furiosus*. Regions of homology in each pair (names of archaeal homologs are after slashes) are indicated with orange bars above the yeast proteins, which are in gray. Insertions and deletions longer than four amino acid residues are marked with green and purple numbers, respectively. Positions are according to that of yeast proteins. *P. furiosus* does not have Rpb8 and Rpb9 homologs, though some other archaea contain Rpb8 homologs. The CTD of Rpb1 that is missing in archaea is boxed with dashed lines. Adapted from Fig. 2 of (Kusser et al. 2008).

occur when CTD is not phosphorylated. Different combinations of the post-translational modification, known as the CTD code, are crucial for the differential binding of hundreds of diverse factors, directly or indirectly (Egloff et al. 2012; Heidemann et al. 2013). These factors, a few listed in Fig. 5C, are required for eukaryotic transcription initiation, elongation, termination, and RNA processing (Egloff et al. 2012; Hsin and Manley 2012).

It is clear that without the CTD of Rpb1, no Pol II transcription would be possible, which includes all eukaryotic protein-coding genes, long noncoding RNAs, and many small nuclear RNAs. Thus, no eukaryote would survive without this CTD. Indeed, it has been found that deletion of the CTD is lethal in yeast, flies, and mice. Thus, the simple fact that eukaryotic transcription requires the CTD renders the archaeal transcription machinery unsuitable for transcribing eukaryotic genomes. Thus, the necessity of the CTD alone in eukarya creates an unbridgeable chasm between archaea and eukarya.

As noted above, people normally ignore the organism-specific deletions, additions, or extensions when talking about protein homologues among different organisms. Furthermore, it is also often

claimed that a region is functionally important because it is conserved across different organisms, while the regions not conserved are functionally dispensable—a concept that is clearly not true. This is demonstrated by the vital roles of the CTD of Rpb1 and the N-terminal of Orc1.

B.2 Transcription factors

Most archaea studied so far contain three general transcription factor genes, TBP, TFB, and TFE (fig. 6A and appendix B). TBP is homologous to the TATA-box binding protein (TBP) subunit of TFIID, which is made of TBP and 14 additional eukaryotic-specific TBP association factors (TAFS). TFE is homologous to the N-terminal domain of TFIIEα, one of the two subunits of TFIIE. Similar to the RNAP subunits, archaeal general transcription factors are also shorter than their corresponding eukaryotic proteins, with many deletions and truncations (fig. 6A and appendix B).

Note that no homologs have been identified in archaea for most eukaryotic general transcription factors, including TAFs, TFIIA, TFIIEß, TFIIF, TFIIH, and Mediator (fig. 6A). These eukaryotic general transcription factors are necessary either
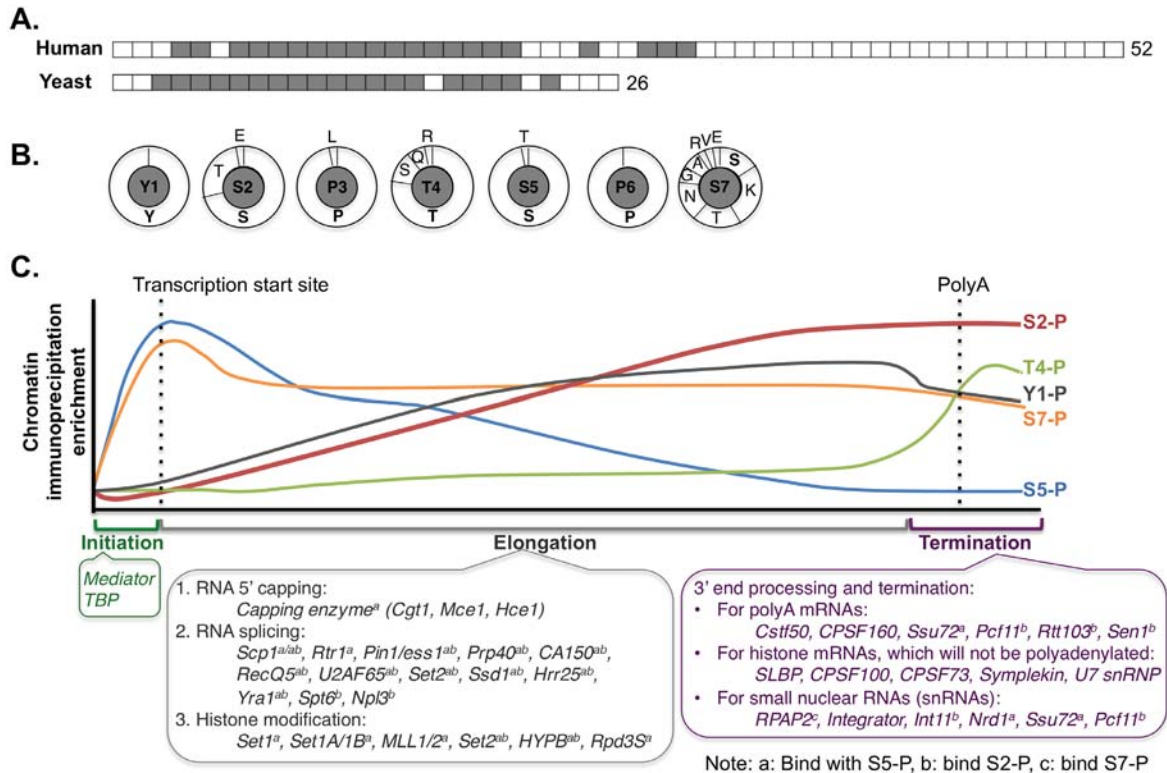
Fig. 5. C-terminal domain (CTD) of eukaryotic Rpb1. **A**. A schematic representation of CTD of human and yeast Rpb1 (gray: consensus repeats, white: non-consensus repeats). **B**. Amino acids in the consensus repeats (center) and the non-consensus repeats (outer circle) in human Rpb1. Numbers indicate the positions of the amino acid in each repeat. The size of each arch corresponds to the frequency of the specific amino acid found in that position in the non-consensus repeats. **C**. Top: Transcription progress-dependent phosphorylation of the serine (S), threonine (T), and tyrosine (Y) residues in the hepta-repeats of the CTD. Bottom: some proteins that interact with the CTD during translation initiation, elongation, or termination. Panels A and B are adapted from Fig. 2 of (Egloff, Dienstbier, and Murphy 2012) and panel C from Fig. 2 of (Heidemann et al. 2013).

to recruit Polymerase II to the promoter and/or to activate Polymerase II in a sequential manner (fig. 6B). Without these general transcription factors and Mediator, no eukaryotic transcription can occur and, thus, no eukarya can survive. The requirement of these eukaryote-specific factors constitute yet another unbridgeable chasm between archaea and eukarya.

C. Translation

Translation is one of the most, if not the most demanding of biological processes in the cell. It involves ribosomes, tRNAs, translation factors, and many other behind-the-scenes proteins and small RNAs. Translation in archaea has not been as well studied as in bacteria or eukarya. Many factors are proposed to play a role during archaeal translation because they share some sequence similarity to a bacterial or eukaryotic factor or a segment of it. It is foreseeable that many archaea-specific translation factors will be discovered in the future. Nonetheless, enough information on archaeal translation has been obtained to allow an informative conclusion that archaeal translation machinery cannot be interchanged with those of bacteria or eukarya—

another unbridgeable chasm between archaea and eukarya.

We will see that the archaeal translation machinery is neither bacterial, nor eukaryotic, but customized to the archaea. Indeed some parts of the archaeal translation machinery and those of bacteria or eukarya have similar sequence and/or structures since all life forms share the same task of decoding information carried by mRNA and translating the message into the amino acid sequences of proteins. However, the archaeal translation machinery can't be exchanged with those of bacteria or eukarya, including ribosomes, tRNAs, and translation factors. Thus, there exists an evolutionarily unbridgeable gap between archaea and eukarya and bacteria in translation, just as in DNA replication and transcription. We will demonstrate this by comparisons of ribosomes, tRNAs, and translation initiation factors.

C.1 Ribosome

Archaeal ribosomes distinguish themselves from those of bacteria and eukarya in many ways, of which only three will be discussed here:

A. Alignments of general transcription factors of *S. cerevisiae* and *P. furiosus*
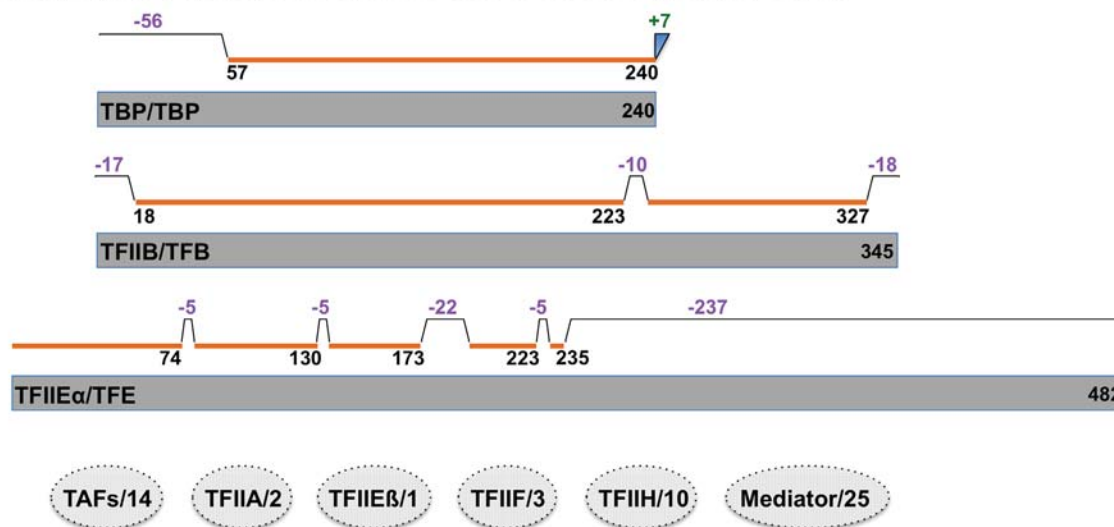
B. Formation of eukaryotic preinitiation complex

Fig. 6. Comparisons of general transcription factors of yeast *S. cerevisiae* with those of archaea *P. furiosus*. **A.** Comparisons of general transcription factors of *S. cerevisiae* and *P. furiosus*. Homologous regions, insertions, and deletions are labeled in the same way as Fig. 4. Factors that identified in yeast but not in archaea are shown as dashed discs in which the numbers after the slash lines indicate distinct protein subunits in t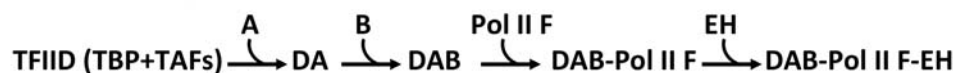he yeast transcription factor. Names of proteins before the slash lines are yeast proteins, those behind the slash lines archaeal. The alignments were generated using pairwise sequence alignments of proteins with Clustal Omega, a multiple sequence alignment program that uses seeded guide trees and hidden Markov model, on http://www.ebi.ac.uk/Tools/msa/clustalo/. The sequences of archaeal *P. furiosus* TBP, TFB, and TFE are downloaded from: http://bioinformatics.zj.cn/archaeatf/, those of yeast TBP, TFIIB, and TFIIEα from http://www.yeastgenome.org. Detailed amino acid alignments are shown in Appendix B. **B.** Formation of eukaryotic preinitiation complex. Eukaryotic transcription preinitiation complex is assembled from transcription factors and RNA polymerase in a sequential order, starting with the binding of transcription factor II D (TFIID, "TFII" are omitted from TFIIA, TFIIB, TFIIF, TFIIE, and TFIIH for simplicity) with DNA. Note that Pol II is recruited simultaneously with TFIIF.

Ribosomal RNAs

The unique signatures of rRNA sequences of archaea is what established archaea as a domain of life, separate from the domains of bacteria and eukarya (Gutell et al. 1985; Woese et al. 1983; Woese and Fox 1977; Woese, Kandler, and Wheelis 1990; Woese, Magrum, and Fox 1978).

Ribosomal proteins

One hundred and two families of ribosomal proteins have been identified in a large-scale analysis of 66 complete genome sequences (45 bacteria, 14 archaea, and seven eukarya) (Lecompte et al. 2002). Archaea contains 68 different families of ribosomal proteins, bacteria 57, eukarya 78, and 32 of the 102 families are considered universal because they are found in all the genomes analyzed. Archaea and eukarya share another 33 families that do not exist in bacteria. Archaea has one family that is not found in either bacteria or eukarya, but it lacks the 23 families unique for bacteria and the 11 families

unique for eukarya. A similar study carried out later by Yutin and colleagues with more species (995 bacteria, 87 archaea, and ten eukarya) confirmed the prior conclusions (Yutin et al. 2012), except three more archaea-specific ribosomal proteins were added that were identified via a proteomic study (Marquez et al. 2011).

Eukarya also contain many (more than 100) mitochondrial-specific ribosomal proteins, which do not have any homologs in bacteria or archaea (Amunts et al. 2014; Desmond et al. 2011; Greber et al. 2014; Rackham and Filipovska 2014; Zikova et al. 2008). In fact, it has recently been found that mitochondrial ribosomes are very different from those of bacteria or archaea or eukarya (Amunts et al. 2014; Greber et al. 2014). This and many other peculiar mitochondrial-specific molecules and processes have enlarged the gap between eukarya and prokaryotes (Zimorski et al. 2014) and have made the endosymbiotic origin theory of mitochondria more questionable than ever.

It is worth pointing out that when one wishes to study the differences of genes using sequences of genomic DNA, RNA transcripts, or proteins, the protein sequences are the least informative. Protein sequence comparisons minimize the differences between the compared genes because it ignores the differences in promoters, 5′ untranslated regions, 3′ untranslated regions, and introns—all integral parts of eukaryotic genes and often much longer in total sequence than the protein coding regions. For example, the small ribosomal protein S12, although it is the most conserved universal ribosomal protein and is assumed to be the oldest ribosomal protein (Harish and Caetano-Anolles 2012), the archaeal, bacterial, and eukaryotic S12 DNA regions differ greatly, especially the eukaryotic S12 gene (fig. 7) (Nakao, Yoshihama, and Kenmochi 2014). The eukaryotic S12 contains spliceosomal introns (one intron for *S. cerevisiae* S12 and three for human S12) whose splicing requires at least over a hundred different proteins that only eukarya, but not archaea or bacteria, have (Behzadnia et al. 2007). No functional S12 proteins can be generated unless those introns are correctly spliced out. Furthermore, accumulating data have shown that even the so-called "silent" or synonymous mutations—those that do not change amino acid sequence of a protein, can have a profound effect on the function of a gene. These mutations can alter the structure and stability of mRNA, the translation rate, and/or the folding and post-translational modifications of proteins (Shabalina, Spiridonov, and Kashina 2013). It is well known that many proteins have different functions depending on their posttranslational modifications. Therefore, one may be misled if he/she only considers differences in the coding regions of genes, which are the most conserved segments of genes between different types of organisms.

Ribosome biogenesis

Ribosome biogenesis includes the processing of rRNA, modification of rRNA, and assembly of ribosomes. Each of these steps requires domain-specific factors. Ribosomal DNA are transcribed as a single unit or as separate units and the regions (5′, 3′, and/or spacers between different RNA species) that are not included in the mature ribosomes are cleaved out by domain-specific endo- and exo-nucleases (Connolly and Culver 2009; Goto, Muto, and Himeno 2013; Henras et al. 2008; Karbstein 2011; Kressler, Hurt, and Bassler 2010; Panse and Johnson 2010; Strunk and Karbstein 2009; Yip, Vincent, and Baserga 2013). The rRNAs are also post-transcriptionally modified in functionally-important bases and many of the modifications are domain-specific (Yip, Vincent, and Baserga 2013). Furthermore, the rRNAs and ribosomal proteins are assembled together into functional ribosomes. This assembling is probably the biggest hurdle of ribosomal biogenesis—hundreds of different proteins and dozens of small RNAs are required to assemble the eukaryote ribosomes (Andersen et al. 2002, 2005; Coute et al. 2006; Moss et al. 2007; Scherl et al. 2002). Very few of these "behind-the-scenes" eukaryotic assembly factors have archaeal homologs (Blombach, Brouns, and van der Oost 2011). Thus, with all archaea have, they can't assemble even a single eukaryotic ribosome, not to mention carrying out eukaryotic translation—one more unbridgeable chasm between archaea and eukarya. On top of that, one would have to face the chicken and the egg problem—all these ribosomal proteins and the ribosomal assembly proteins themselves need to be synthesized by ribosomes.
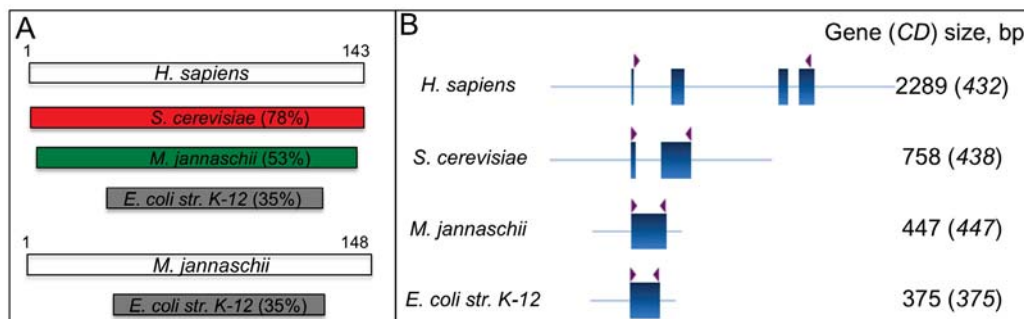
Fig. 7. Comparisons of S12 from eukaryotes *Homo sapiens*, *S. cerevisiae*, bacterium *E. coli*, and archaea *Methanocaldococcus jannaschii*. **A.** Comparison of S12 proteins. Top four: comparisons of *S. cerevisiae*, *E. coli*, *M. jannaschii* S12 with *H. sapiens* S12. Bottom two: a comparison of *E. coli* S12 with *M. jannaschii* S12. Percentages of identity are according to pairwise sequence alignments of proteins using BLASTP with default settings at http://blast.ncbi.nlm.nih.gov/. **B.** Comparison of S12 DNA. Exons are shown as boxes, introns and the upstream and the downstream sequences as lines. The translation start and stop sites are marked with arrowheads. The gene annotations are downloaded from http://ribosome.med.miyazaki-u.ac.jp. *M. jannaschii* is a thermophilic methanogenic archaea in the class *Methanococci* and is the first archaea to have its complete genome sequenced.
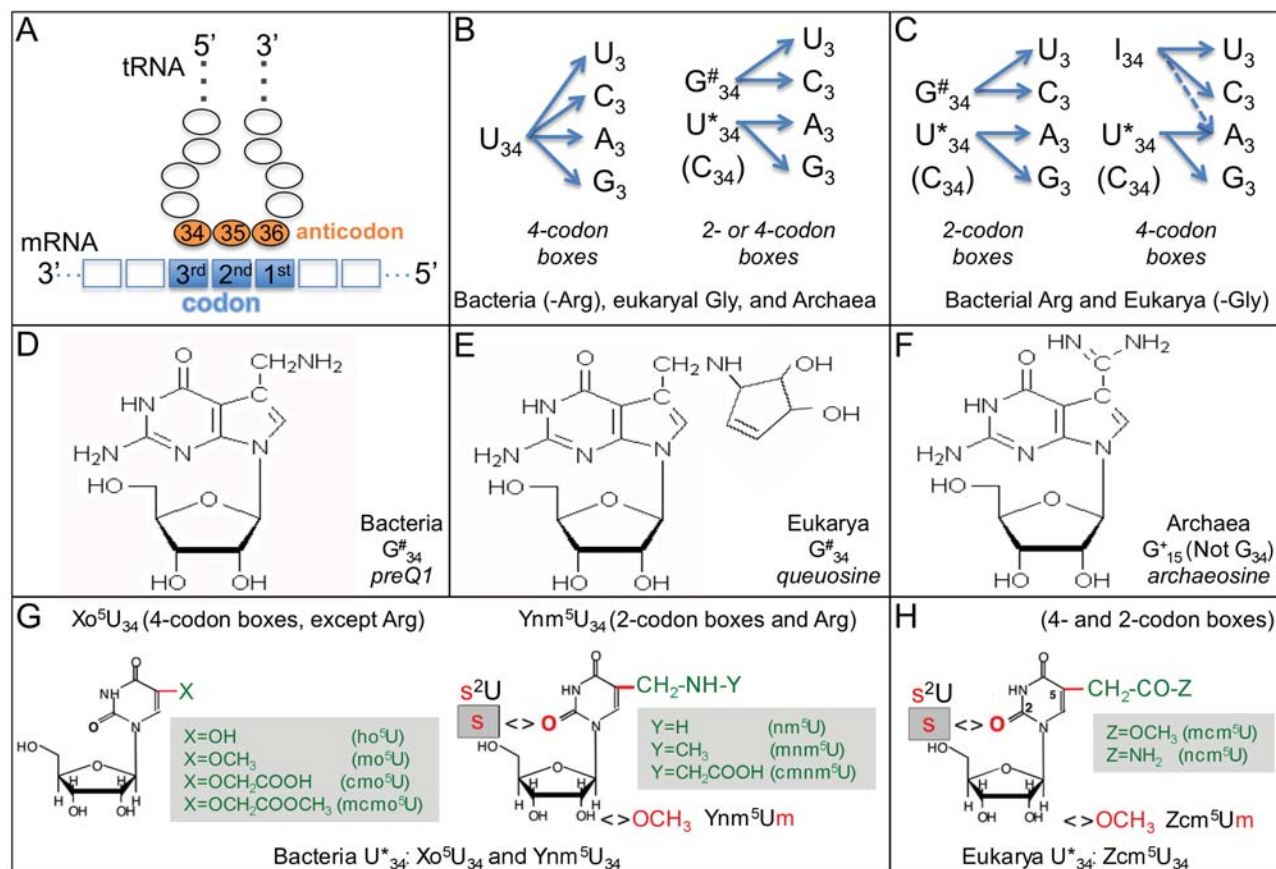
Fig. 8. Transfer RNAs of the three domains. **A**. Base-pairing of the codons in mRNA and the anticodons in the tRNA. **B**. Decoding in bacteria (except the arginine codons CGN) and archaea. N is any base. Only the first base of the anticodons, base 34, and the third base of the codons are shown. **C**. Decoding in eukaryotes (except the glycine codons GGN). **D**. Structure of preQ1, a modification of bacteria G_{34}. **E**. Structure of queuosine, a modification of eukaryotic G_{34}. **F**. Structure of archaeosine, a modification of archaeal G_{15}, a base in the D loop of archaeal tRNA. Note that the archaeal anticodon base G_{34} is not modified. **G**. Structure of modified U_{34} in bacterial tRNAs. Xo^5U_{34} for decoding 4-codon boxes, except Arg, while Ynm^5U_{34} for decoding 2-codon boxes and the 4-codon Arg box. **H**. Structure of modified U_{34} in eukaryotic tRNAs. Panels B, C, G, and H are adapted from Fig. 5 of (Grosjean, de Crecy-Lagard, and Marck 2010). The structures of preQ1, queuosine, and archaeosine are adapted from the RNA Modification Database http://mods.rna.albany.edu.

C.2 Transfer RNAs

No translation of mRNA into proteins would occur without the translators, called transfer RNAs or tRNAs. There are 64 different codons that code for 20 amino acids (22 in some organisms). Thus, most amino acids have multiple codons, a phenomenon known as the codon degeneracy. In addition, although one might expect that 61 amino acid coding codons would require 61 anticodons, and thus 61 kinds of tRNAs (in most organisms, three codons are stop codons that do not code for any amino acids but serve as signals to terminate translation). In reality, almost all archaea and some eukarya use only 46 kinds of tRNAs (Grosjean, de Crecy-Lagard, and Marck 2010), though each tRNA may have multiple copies. This is accomplished by using either a G_{34} or an A_{34}-anticodon containing tRNA to decode both the U_3 and C_3-ending codons (codon bases 1, 2, and 3 base-pair with and are read by anticodon base 36, 35, and 34 as shown in fig. 8A). Strikingly, archaea and bacteria codons

NNU_3 and NNC_3, except bacterial arginine codons, are almost always decoded by a tRNA with a G_{34} anticodon (fig. 8B). On the other hand, eukarya use G_{34}-containing anticodon tRNAs to decode only two codon box codons (fig. 8C). For the 4-codon box codons, eukarya use tRNAs with A_{34}-anticodons. The A_{34} in the anticodons are normally deaminated into inosine both in bacteria tRNA[Arg] and in eukaryotic tRNAs, though through different enzymes—homodimeric tadA in bacteria and heteromeric tad2 and tad3 in eukarya. No archaeal homologues of *tadA/tad2/tad3* genes have been found (Grosjean, de Crecy-Lagard, and Marck 2010). Some bacteria use a single U_{34}-anticodon containing tRNA to decode all members of a 4-codon box (fig. 8B). Thus, archaea, bacteria, and eukarya deal with the degeneracy issue in their own ways, mostly with eukarya on one side and bacteria and archaea on the other (El Yacoubi, Bailly, and de Crecy-Lagard 2012; Grosjean, de Crecy-Lagard, and Marck 2010). See Table 2 for a listing of anticodon

Table 2: Anticodon usage and modifications of base 34 of anticodons in bacteria, archaea, and eukarya.

aa	cdn	acd	Anticodon Base 34			Anticodon Usages					
			Bacteria	Archaea	Eukaryotes	Bacteria		Archaea		Eukaryotes	
Ala	GCU	AGC	–	–	I	0[$]	0.00[$$]	0[$]	0.00[$$]	830[$]	0.44[$$]
	GCC	GGC	G	G		559	0.24	42	0.33	43	0.02
	GCG	CGC	C	C	C	212	0.09	35	0.27	312	0.16
	GCA	UGC	xo^5U		ncm^5U	1573	0.67	52	0.40	715	0.38
Arg	CGU	ACG	I		I	1080	0.61	0	0.00	627	0.32
	CGC	GCG		G		46	0.03	44	0.37	45	0.02
	CGG	CCG	C	C	C	510	0.29	32	0.27	211	0.11
	CGA	UCG	mnm^5U		mcm^5U	128	0.07	42	0.36	1091	0.55
	AGG	CCU	C	C	C	487	0.42	35	0.45	470	0.43
	AGA	UCU	ynm^5U	$?U$	zcm^5U	678	0.58	42	0.55	621	0.57
Asn	AAU	AUU	–	–		0	0.00	0	0.00	92	0.05
	AAC	GUU	Q0	G	Q#	1234	1.00	43	1.00	1859	0.95
Asp	GAU	AUC	–	–		0	0.00	0	0.00	50	0.05
	GAC	GUC	Q0	G	Q#	1324	1.00	52	1.00	956	0.95
Cys	UGU	ACA	–	–		0	0.00	0	0.00	176	0.15
	UGC	GCA	G**	G**	G**	704	1.00	55	1.00	1010	0.85
Glu	GAG	CUC	C	C	C	206	0.13	35	0.42	1012	0.47
	GAA	UUC	ynm^5s^2U	$?U$	zcm^5s^2U	1436	0.87	49	0.58	1159	0.53
Gln	CAG	CUG	C	C	C	264	0.22	34	0.45	692	0.40
	CAA	UUG	ynm^5s^2U	$?U$	zcm^5s^2U	952	0.78	42	0.55	1046	0.60
Gly	GGU	ACC	–	–		0	0.00	0	0.00	54	0.01
	GGC	GCC	G	G	G	1329	0.53	45	0.37	1515	0.35
	GGG	CCC	C	C	C	337	0.13	35	0.29	1780	0.42
	GGA	UCC	ynm^5Um	$?U$	scm^5Um	855	0.34	42	0.34	935	0.22
His	CAU	AUG		–		1	0.00	0	0.00	65	0.07
	CAC	GUG	Q0	G	Q#	716	1.00	42	1.00	886	0.93
Ile	AUU	AAU		–	I	1	0.00	0	0.00	953	0.74
	AUC	GAU	G	G		1400	0.99	42	0.98	42	0.03
	AUA	UAU	k^2C^*	agm^2C^*	Ψ	10	0.01	1	0.02	300	0.23
Leu	CUU	AAG			I	68	0.03	1	0.01	723	0.41
	CUC	GAG	G	G		573	0.28	43	0.36	22	0.01
	CUG	CAG	C	C	C	625	0.30	34	0.28	550	0.31
	CUA	UAG	xo^5U		U	793	0.39	42	0.35	462	0.26
	UUG	CAA	C	C	C	599	0.45	34	0.45	460	0.63
	UUA	UAA	ynm^5U	$?U$	zcm^5U	745	0.55	42	0.55	271	0.37
Lys	AAG	CUU	C	C	C	404	0.24	35	0.42	2832	0.70
	AAA	UUU	ynm^5s^2U	$?U$	zcm^5s^2U	1287	0.76	49	0.58	1225	0.30
Met	AUG	CAU	ac^4C	Cm	Cm	2672		130		1486	
Phe	UUU	AAA		–		1	0.00	0	0.00	31	0.04
	UUC	GAA	Gm	Gm	Gm	912	1.00	43	1.00	692	0.96
Pro	CCU	AGG		–	I	1	0.00	0	0.00	703	0.38
	CCC	GGG	G	G		415	0.26	41	0.35	25	0.01
	CCG	CGG	C	C	C	324	0.20	33	0.28	359	0.19
	CCA	UGG	xo^5U		ncm^5U	849	0.53	42	0.36	769	0.41
Ser	UCU	AGA		–	I	1	0.00	0	0.00	779	0.50
	UCC	GGA	G	G		636	0.34	42	0.36	68	0.04
	UCG	CGA	C	C	C	377	0.20	32	0.27	259	0.17
	UCA	UGA	xo^5U		ncm^5U	840	0.45	43	0.37	444	0.29
	AGU	ACU	–	–		0	0.00	0	0.00	26	0.03
	AGC	GCU	G**	G**	G**	677	1.00	43	1.00	778	0.97
Thr	ACU	AGU		–	I	22	0.01	0	0.00	842	0.50
	ACC	GGU	G	G		644	0.32	43	0.36	57	0.03
	ACG	CGU	C	C	C	399	0.20	35	0.29	223	0.13
	ACA	UGU	xo^5U		ncm^5U	927	0.47	43	0.36	569	0.34
Trp	UGG	CCA	Cm	Cm	Cm	664		41		593	
Tyr	UAU	AUA	–	–		0	0.00	0	0.00	32	0.04
	UAC	GUA	Q0	G	Q#	933	1.00	43	1.00	777	0.96
Val	GUU	AAC	–	–	I	0	0.00	0	0.00	758	0.41
	GUC	GAC	G	G		610	0.28	43	0.35	52	0.03
	GUG	CAC	C	C	C	235	0.11	36	0.30	611	0.33
	GUA	UAC	xo^5U		ncm^5U	1300	0.61	43	0.35	432	0.23

Highlighted backgrounds: light blue: four codon boxes, gray: C_{34} modified or unmodified, green: mainly or exclusively G_{34} anticodon-containing tRNAs, brown: mainly A_{34} anticodon-containing tRNAs. Highlighted letters: decoding of Arg, which deviates from the norm manner of decoding of bacteria and archaea, and Gly, which deviates from the norm manner of decoding of eukaryotes. Regarding to the anticodon usages, for each domain, on the left (marked with [$]) is the numbers of tRNAs with the anticodon of the specific row, while on the right (marked with [$$]) is the percentage of that anticodon in the corresponding four codon box or two codon box (for Ile, it is a three codon box). The tRNA numbers are according to the genomic tRNA database http://gtrnadb.ucsc.edu. Anticodon base 34 modifications are based on El Yacoubi, Bailly, and de Crecy-Lagard (2012); Grosjean et al. (2010); Jackman and Alfonzo (2013); Phillips and de Crecy-Lagard (2011); Vinayak and Pathak (2010). Abbreviations: aa: amino acid, cdn: codon, acd: anticodon.

usage and modifications of base 34 of anticodons in bacteria, archaea, and eukarya.

Another character of tRNAs is that they are the most modified RNA species, with many of their composing bases modified. Remarkably, the two most often modified bases are the anti-codon base 34 and an anti-codon-next-door-neighbor base, base 37 (El Yacoubi, Bailly, and de Crecy-Lagard 2012; Jackman and Alfonzo 2013; Paris, Fleming, and Alfonzo 2012). Some of the modification requires sequential actions of multiple enzymes (El Yacoubi, Bailly, and de Crecy-Lagard 2012). Localized in the very business center of the tRNA, modifications on bases 34 and 37 are essential for correctly decoding the information encoded in a gene. In addition, they are critical for the structure, stability, and function of the tRNAs and their interaction with other molecules critical for translation, including the aminoacyl-tRNA synthetases, which charge tRNA with the correct amino acids, and translation elongation factors (El Yacoubi, Bailly, and de Crecy-Lagard 2012). Though the effect of missing single modifications may be small on translation, missing multiple tRNA modifications is lethal (Alexandrov et al. 2006).

Strikingly, many of the tRNA modifications are domain-specific (El Yacoubi, Bailly, and de Crecy-Lagard 2012; Jackman and Alfonzo 2013). For example, G_{34} is often modified into preQ1 in bacteria, queuosine or its derivatives in eukarya, but not modified in archaea (figs. 8D and 8E). Archaeal G_{15} is modified into archaeosine, a modification that has not been identified in bacteria or eukarya (fig. 8F). U_{34} is modified into Xo^5U_{34}, derivatives of hydroxyuridine, or Ynm^5U_{34}, derivatives of aminomethyluridine, depending on whether it is a 4-codon box or 2-codon box (except Arg tRNA), in bacteria (fig. 8G). In contrast, eukaryotic tRNA U_{34} are modified into derivatives of carbamoylmethyluridine, regardless of the codon types (fig. 8H). Modification of archaea tRNA U_{34} is not clear, and is possibly eukaryote-like (Selvadurai et al. 2014).

Furthermore, even the same modification, for example, the deamination of A_{34} mentioned above, is normally accomplished by different enzymes in bacteria, archaea, and eukarya (Grosjean, de Crecy-Lagard, and Marck 2010; Jackman and Alfonzo 2013). This makes it impossible to predict the presence of a specific modification in an organism based on the presence in its genome of an enzyme that is homologous to an experimentally characterized enzyme in another organism (El Yacoubi, Bailly, and de Crecy-Lagard 2012; Jackman and Alfonzo 2013).

Thus, archaea differ from eukarya in decoding strategies. Archaea group together with bacteria in the choice of types of tRNAs or use of anticodons. The tRNA modifications and especially the enzymes carrying out these modifications are mostly domain specific.

C.3 Translation initiation factors

Translation initiation differs greatly among different life forms, a reflection of the differences of mRNA structures and translation regulations. Several different translation initiation mechanisms have been identified, including Shine Dalgarno (SD)-dependent and SD-independent mechanisms in bacteria and archaea, cap-dependent and cap-independent (including internal ribosomal entry site-dependent) scanning mechanisms in eukarya, and various leaderless mechanisms (Christian and Spremulli 2010; Gabel et al. 2013; Malys and McCarthy 2011; Shatsky et al. 2010; Vesper et al. 2011). The best-studied mechanisms are the SD-dependent mechanism in bacteria and the scanning mechanism in eukarya, and hence we will limit our discussions to these two mechanisms.

Multiple initiation factors (IFs) are required to initiate translation. Unfortunately, the names of the initiation factors are very confusing because similar names may not be related, either in sequence or in function. SD-dependent translation in bacteria uses three initiation factors, IF1, IF2, and IF3. The cap-dependent scanning translation in eukarya requires more than 30 different proteins organized into more than ten eukaryotic initiation factors (eIFs) (Table 3): eIF1, eIF1A, eIF2 (made of eIF2α, eIF2β, eIF2γ), eIF2B (made of five different proteins), eIF3 (made of six to 13 different proteins, depending on the organisms), eIF4B, eIF4F (a complex of eIF4A, eIF4E, and eIF4G), eIF4H, eIF5, eIF5B, and eIF6. Bacteria IF1 and eukaryotic eIF1A, as well IF2 and eIF5B, share some sequence similarity and thus these two groups are considered universal initiation factors, though their functions are not equivalent. The eukaryotic-specific initiation factor eIF4F is the principle factor for the binding of the 5′ cap of eukaryotic mRNA and the recruitment of the mRNA to ribosome. The eukaryotic-specific initiation factor eIF3 interacts with almost all other initiation factors via its multiple subunits and is essential for the interaction of mRNA and ribosome and cooperation of other initiation factors.

Several archaeal translation initiation factors have been predicted based on sequence homology (named according to their eukaryotic homologs with a prefix "a") (Table 3): homologues of the universal initiation factors aIF1, aIF1A, and aIF5B; three subunits of aIF2 (aIF2α, aIF2β, aIF2γ), two subunits of aIF2B (aIF2Bα and aIF2Bδ), aIF5A, and aIF6. Whether these proteins function as bona fide initiation factors has been questioned (Gabel et al. 2013). In addition, some of the similarity is very limited (Table 3). For example, only less than ten amino acids of aIF5′ 124 amino acids aligned with some amino acids in eIF5A. Yet they are claimed to be homologues. Nonetheless,

Table 3. Comparison of translation initiation factors in three domains of life.

Eukaryote	E. coli	H. volcanii	S. cerevisiae	Coverage	Identities	Gaps
eIF1	–	aIF1-1 (HVO_1946)*	Sui1*	74%	26/77	6
		aIF1-2 (HVO_B0284)	Sui1*	65%	20/70	6
		alignment of HVO_1946 and HVO_B0284		*91%*	*44/89*	*0*
eIF1A	IF1	aIF1A-1 (HVO_0136)**	Tif11*	96%	31/93	0
		aIF1A-2 (HVO_A0637)**	Tif11*	92%	25/87	0
		alignment of HVO_A0637 and IF1		*65%*	*19/63*	*1*
		alignment of Tif11 and IF1		*37%*	*15/58*	*1*
		alignment of HVO_A0637 and HVO_0136		*97%*	*63/92*	*0*
eIF2α	–	aIF2α (HVO_0699)	Sui2*	93%	71/258	16
eIF2β	–	aIF2β-1 (HVO_1678)**	Sui3*	83%	29/115	2
		aIF2β-2 (HVO_2242)**	Sui3*	64%	44/140	9
		alignment of HVO_2242 and HVO_1678		*69%*	*53/132*	*1*
eIF2γ	–	aIF2γ (HVO_1901)*	Gcd11*	98%	174/422	17
eIF2Bα	–	aIF2Bα (HVO_1934)	Gcn3	32%	40/134	10
eIF2Bβ	–	–	GCD7*			
eIF2Bγ	–	–	GCD1*			
eIF2Bδ	–	aIF2Bδ (HVO_2706)	Gcd2*	83%	54/224	15
eIF2Bε	–	–	GCD6*			
eIF3a	–	–	Tif32*			
eIF3b	–	–	Prt1*			
eIF3c	–	–	Nip1*			
eIF3g	–	–	Tif35*			
eIF3i	–	–	Tif34*			
eIF3j	–	–	Hcr1			
eIF3 d-f, h, k-m	–	–	–			
eIF4A1	–	aIF4A (HVO_1333)	Tif1**	42%	78/318	21
eIF4A2	–	–	Tif2**			
eIF4B	–	–	Tif3			
eIF4E	–	–	CDC33/TIF45*			
eIF4G1	–	–	TIF4631**			
eIF4G2	–	–	TIF4632**			
eIF4H	–	–	–			
eIF5	–	aIF5A (HVO_2300) (41-48)*	IF5/TIF5	6%	5/8	0
		alignment of HVO_2300 (107-124) with eIF2β (Sui3)		*14%*	*8/26*	*8*
		alignment of HVO_2300 (90-103) with eIF2Bβ (GCD7)		*11%*	*5/14*	*0*
		alignment of HVO_2300 (108-116) with eIF2Bγ (GCD1)		*7%*	*6/9*	*0*
		alignment of HVO_2300 (84-106) with eIF2Bε (GCD6)		*18%*	*8/23*	*0*
eIF5B	IF2	aIF5B (HVO_1963)*	Fun12*	90%	219/568	29
		alignment of aIF5B (HVO_1963) and IF2		*90%*	*164/568*	*107*
		alignment of eIF5B (Fun12) and IF2		*62%*	*155/159*	*103*
eIF6	–	aIF6 (HVO_0117)*	Tif6*	100%	76/223	4
–	IF3	–	–			

*(colored brown): Single mutations are lethal for *H. volcanii* or *S. cerevisiae*. **(shaded gray): Double mutations are lethal for *H. volcanii* (*HVO_A0637* and *HVO_0136*, as well as *HVO_2242* and *HVO_1678*) or *S. cerevisiae* (Tif1 and Tif2, as well as TIF4631 and TIF4632). Lethality of the genes are according to (Gabel et al. 2013) and the yeast genome database http://www.yeastgenome.org. The coverage, identities, and gaps are according to pairwise sequence alignments of proteins using BLASTP with default settings at http://blast.ncbi. nlm.nih.gov/. All protein sequences are obtained from http://www.ncbi.nlm.nih.gov/protein/. The names of the archaeal homologs include the names according to their eukaryotic homologs and the original *H. volcanii* gene names (in brackets). The numbers underlined after the *H. volcanii* gene names indicate the portions of the corresponding proteins that aligned with their eukaryotic homologs.

even with such low similarity cutoff, most of the eukaryote initiation factors, such as eIF3 and eIF4F, do not have homologues in bacteria or archaea, yet are essential for the survival of the eukarya—another unbridgeable chasm between archaea and eukarya.

As repeatedly mentioned above, to claim a gene has certain function based only on sequence similarity can be misleading. Several archaeal homologues of eIFs, including aIF1, aIF2α, aIF2Bα, aIF2Bδ, and eIF4A homolog, can be deleted in *Haloferax volcanii*, thus probably are not always essential for *H. volcanii* translation, although their eukaryotic counterparts are important for translation initiation in eukarya (Gabel et al. 2013). Another example is the Shine Dalgarno sequence, which has been shown to be critical for translation initiation for many *E. coli*

genes and base pairs with the anti-SD sequence in the 16S rRNA. SD and anti-SD have been identified in *H. volcanii* but an intensive deletion study shows that it plays no role at all in the translation initiation of its host gene (Kramer et al. 2014).

Conclusion

The above comparisons of a few molecules involved in the information processing in the three domains of life reveals several interesting phenomena: 1) Molecular machines are employed as modules, that is, a process is either bacterial-like or eukaryote-like. 2) Each machine is a molecular mosaic of modules that is fine-tuned to meet the unique need of an organism. 3) The machines for DNA replication, transcription, and translation in bacteria, archaea, and eukarya are unique and specific for each domain of life, and thus, can't be exchanged. 4) Functional annotations of genes based on sequence homology comparisons can be misleading because they only take into account isolated parts of proteins, not the entire gene. 5) Organism-specific protein extensions, such as the CTD of eukaryotic Rpb1, can be the determinant factor of life vs. death for the specific organism.

Therefore, the molecules involved in DNA replication, transcription, and translation in bacteria, archaea, and eukarya are crying out loudly that eukarya did not evolve from archaea. These complex systems when analyzed independently and in detail, reveal the impossibility of them being able to have evolved through the alleged naturalistic processes of selection upon accumulated mutations. There are many uncrossable gaps between archaea, bacteria, and eukarya, and any one such gap makes it impossible for any archaea or bacteria to evolve into a eukaryote.

References

Alexandrov, A., I. Chernyakov, W. Gu, S.L. Hiley, T.R Hughes, E.J. Grayhack, and E.M. Phizicky. 2006. Rapid tRNA decay can result from lack of nonessential modifications. *Molecular Cell* 21, no. 1:87–96. doi: 10.1016/j. molcel.2005.10.036.

Allers, T., and M. Mevarech. 2005. Archaeal genetics–the third way. *Nature Reviews: Genetics* 6, no. 1:58–73. doi: 10.1038/ nrg1504.

Amunts, A., A. Brown, X.C. Bai, J.L. Llacer, et al. 2014. Structure of the yeast mitochondrial large ribosomal subunit. *Science* 343, no. 6178:1485–1489. doi: 10.1126/ science.1249410.

Andersen, J.S., Y.W. Lam, A.K. Leung, S.E. Ong, C.E. Lyon, A.I. Lamond, and M. Mann. 2005. Nucleolar proteome dynamics. *Nature* 433, no. 7021:77–83. doi: 10.1038/ nature03207.

Andersen, J.S., C.E. Lyon, A.H. Fox, A.K. Leung, Y.W. Lam, H. Steen, M. Mann, and A.I. Lamond. 2002. Directed proteomic analysis of the human nucleolus. *Current Biology* 12, no. 1:1–11.

Aves, S. J., Y. Liu, and T. A. Richards. 2012. Evolutionary diversification of eukaryotic DNA replication machinery. *Subcellular Biochemistry* 62:19–35. doi: 10.1007/978-94- 007-4572-8_2.

Behzadnia, N., M.M. Golas, K. Hartmuth, B. Sander, B. Kastner, J. Deckert, P. Dube. et al. 2007. Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO Journal* 26, no. 6:1737–1748. doi: 10.1038/sj.emboj.7601631.

Benelli, D., E. Maone, and P. Londei. 2003. Two different mechanisms for ribosome/mRNA interaction in archaeal translation initiation. *Molecular Microbiology* 50, no. 2:635– 643.

Blombach, F., S.J. Brouns, and J. van der Oost. 2011. Assembling the archaeal ribosome: roles for translation- factor-related GTPases. *Biochemical Society Transactions* 39, no. 1:45–50. doi: 10.1042/BST0390045.

Bochman, M.L., and A. Schwacha. 2009. The Mcm complex: unwinding the mechanism of a replicative helicase. *Microbiology and Molecular Biology Reviews* 73, no. 4:652– 683. doi: 10.1128/MMBR.00019-09.

Cann, I.K., K. Komori, H. Toh, S. Kanai, and Y. Ishino. 1998. A heterodimeric DNA polymerase: evidence that members of Euryarchaeota possess a distinct DNA polymerase. *Proceedings of the National Academy of Sciences USA* 95, no. 24:14250–14255.

Cavicchioli, R. 2011. Archaea—timeline of the third domain. *Nature Reviews: Microbiology* 9, no. 1:51–61. doi: 10.1038/ nrmicro2482.

Christian, B.E., and L.L. Spremulli. 2010. Preferential selection of the 5′-terminal start codon on leaderless mRNAs by mammalian mitochondrial ribosomes. *Journal of Biological Chemistry* 285, no. 36:28379–28386. doi: 10.1074/jbc.M110.149054.

Connolly, K., and G. Culver. 2009. Deconstructing ribosome construction. *Trends in Biochemical Sciences* 34, no. 5:256– 263. doi: 10.1016/j.tibs.2009.01.011.

Costa, A., I.V. Hood, and J.M. Berger. 2013. Mechanisms for initiating cellular DNA replication. *Annual Review of Biochemistry* 82:25–54. doi: 10.1146/annurev- biochem-052610-094414.

Couté, Y., J.A. Burgess, J.J. Diaz, C. Chichester, F. Lisacek, A. Greco, and J.C. Sanchez. 2006. Deciphering the human nucleolar proteome. *Mass Spectrometry Reviews* 25, no. 2:215–234. doi: 10.1002/mas.20067.

Čuboňová, L., T. Richardson, B.W. Burkhart, Z. Kelman, B.A. Connolly, J N. Reeve, T. J. Santangelo. 2013. Archaeal DNA polymerase D but not DNA polymerase B is required for genome replication in *Thermococcus kodakarensis*. *Journal of Bacteriology* 195, no. 10:2322–2328. doi: 10.1128/ JB.02037-12.

Dagan, T., and W. Martin. 2006. The tree of one percent. *Genome Biology* 7, no. 10:118. doi: 10.1186/gb-2006-7-10-118.

Desmond, E., C. Brochier-Armanet, P. Forterre, and S. Gribaldo. 2011. On the last common ancestor and early evolution of eukarya: reconstructing the history of mitochondrial ribosomes. *Research in Microbiology* 162, no. 1:53–70. doi: 10.1016/j.resmic.2010.10.004.

Dueber, E.L., J.E. Corn, S.D. Bell, and J.M. Berger. 2007. Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex. *Science* 317 no. 5842:1210-1213. doi: 10.1126/science.1143690.

Duncker, B.P., I.N. Chesnokov, and B.J. McConkey. 2009. The origin recognition complex protein family. *Genome Biology* 10, no. 3:214. doi: 10.1186/gb-2009-10-3-214.

Egloff, S., M. Dienstbier, and S. Murphy. 2012. Updating the RNA polymerase CTD code: adding gene-specific layers. *Trends in Genetics* 28, no. 7:333–341. doi: 10.1016/j.tig.2012.03.007.

El Yacoubi, B., M. Bailly, and V. de Crecy-Lagard. 2012. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annual Review of Genetics* 46:69–95. doi: 10.1146/annurev-genet-110711-155641.

Esser, C., N. Ahmadinejad, C. Wiegand, C. Rotte, et al. 2004. A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Molecular Biology and Evolution* 21, no. 9:1643–1660. doi: 10.1093/molbev/msh160.

Evguenieva-Hackenberg, E., P. Walter, E. Hochleitner, F. Lottspeich, and G. Klug. 2003. An exosome-like complex in *Sulfolobus solfataricus*. *EMBO Reports* 4, no. 9:889–893. doi: 10.1038/sj.embor.embor929.

Gabel, K., J. Schmitt, S. Schulz, D.J. Nather, and J. Soppa. 2013. A comprehensive analysis of the importance of translation initiation factors for *Haloferax volcanii* applying deletion and conditional depletion mutants. *PLoS One* 8, no. 11:e77188. doi: 10.1371/journal.pone.0077188.

Galperin, M.Y. 2007. Linear chromosomes in bacteria: no straight edge advantage? *Environmental Microbiology* 9, no. 6:1357–1362.

Gaudier, M., B.S. Schuwirth, S.L. Westcott, and D.B. Wigley. 2007. Structural basis of DNA replication origin recognition by an ORC protein. *Science* 317, no. 5842:1213–1216. doi: 10.1126/science.1143664.

Goto, S., A. Muto, and H. Himeno. 2013. GTPases involved in bacterial ribosome maturation. *Journal of Biochemistry* 153, no. 5:403–414. doi: 10.1093/jb/mvt022.

Grainge, I., M. Gaudier, B.S. Schuwirth, S.L. Westcott, J. Sandall, N. Atanassova, and D.B. Wigley. 2006. Biochemical analysis of a DNA replication origin in the archaeon *Aeropyrum pernix*. *Journal of Molecular Biology* 363, no. 2:355–369. doi: 10.1016/j.jmb.2006.07.076.

Greber, B.J., D. Boehringer, M. Leibundgut, P. Bieri, A. Leitner, N. Schmitz, R. Aebersold, and N. Ban. 2014. The complete structure of the large subunit of the mammalian mitochondrial ribosome. *Nature* 515, no. 7526:283–286. doi: 10.1038/nature13895.

Gribaldo, S., A.M. Poole, V. Daubin, P. Forterre, and C. Brochier-Armanet. 2010. The origin of eukaryotes and their relationship with the Archaea: Are we at a phylogenomic impasse? *Nature Reviews: Microbiology* 8, no. 10:743–752. doi: 10.1038/nrmicro2426.

Grosjean, H., V. de Crecy-Lagard, and C. Marck. 2010. Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. *FEBS Letters* 584, no. 2:252–264. doi: 10.1016/j.febslet.2009.11.052.

Gutell, R.R., B. Weiser, C.R. Woese, and H.F. Noller. 1985. Comparative anatomy of 16-S-like ribosomal RNA. *Progress in Nucleic Acid Research and Molecular Biology* 32:155–216.

Harish, A., and G. Caetano-Anollés. 2012. Ribosomal history reveals origins of modern protein synthesis. *PLoS One* 7, no. 3:e32776. doi: 10.1371/journal.pone.0032776PONE.

Heidemann, M., C. Hintermair, K. Voss, and D. Eick. 2013. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochimica et Biophysica Acta—Gene Regulatory Mechanisms* 1829, no. 1:55–62. doi: 10.1016/j.bbagrm.2012.08.013.

Henras, A.K., J. Soudet, M. Gerus, S. Lebaron, M. Caizerques-Ferrer, A. Mougin, and Y. Henry. 2008. The post-transcriptional steps of eukaryotic ribosome biogenesis. *Cellular and Molecular Life Sciences* 65, no. 15:2334–2359. doi: 10.1007/s00018-008-8027-0.

Hsin, J.P., and J.L. Manley. 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes and Development* 26, no. 19:2119–2137. doi: 10.1101/gad.200303.112.

Hsin, J.P., K. Xiang, and J.L. Manley. 2014. Function and control of RNA polymerase II C-terminal domain phosphorylation in vertebrate transcription and RNA processing. *Molecular and Cellular Biology* 34, no. 10:2488–2498. doi: 10.1128/MCB.00181-14.

Huet, J., R. Schnabel, A. Sentenac, and W. Zillig. 1983. Archaebacteria and eukarya possess DNA-dependent RNA polymerases of a common type. *EMBO Journal* 2, no. 8:1291–1294.

Ishino, Y., and S. Ishino. 2012. Rapid progress of DNA replication studies in Archaea, the third domain of life. *Science China: Life Sciences* 55, no. 5:386–403. doi: 10.1007/s11427-012-4324-9.

Jackman, J.E., and J.D. Alfonzo. 2013. Transfer RNA modifications: nature's combinatorial chemistry playground. *Wiley Interdisciplinary Reviews: RNA* 4, no. 1:35–48. doi: 10.1002/wrna.1144.

Jun, S.H., M.J. Reichlen, M. Tajiri, and K.S. Murakami. 2011. Archaeal RNA polymerase and transcription regulation. *Critical Reviews in Biochemistry and Molecular Biology* 46, no. 1:27–40. doi: 10.3109/10409238.2010.538662.

Karbstein, K. 2011. Inside the 40S ribosome assembly machinery. *Current Opinion in Chemical Biology* 15, no. 5:657–663. doi: 10.1016/j.cbpa.2011.07.023.

Knoll, A.H., E.J. Javaux, D. Hewitt, and P. Cohen. 2006. Eukaryotic organisms in Proterozoic oceans. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 361, no. 1470:1023–1038. doi: 10.1098/rstb.2006.1843.

Kramer, P., K. Gabel, F. Pfeiffer, and J. Soppa. 2014. *Haloferax volcanii,* a prokaryotic species that does not use the Shine Dalgarno mechanism for translation initiation at 5′-UTRs. *PLoS One* 9, no. 4:e94979. doi: 10.1371/journal. pone.0094979.

Kressler, D., E. Hurt, and J. Bassler. 2010. Driving ribosome assembly. *Biochimica and Biophysica Acta: Molecular Cell Research* 1803, no. 6:673–683. doi: 10.1016/j.bbamcr.2009.10.009.

Kuo, A.J., J. Song, P. Cheung, S. Ishibe-Murakami, S. Yamazoe, J.K. Chen, D.J. Patel, and O. Gozani. 2012. The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. *Nature* 484, no. 7392:115–119. doi: 10.1038/nature10956.

Kusser, A.G., M.G. Bertero, S. Naji, T. Becker, M. Thomm, R. Beckmann, and P. Cramer. 2008. Structure of an archaeal RNA polymerase. *Journal of Molecular Biology* 376, no. 2:303–307. doi: 10.1016/j.jmb.2007.08.066.

Lake, J.A., R.G. Skophammer, C.W. Herbold, and J.A. Servin. 2009. Genome beginnings: rooting the tree of life. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 364, no. 1527:2177–2185. doi: 10.1098/rstb.2009.0035.

Lecompte, O., R. Ripp, J.-C. Thierry, D. Moras, and O. Poch. 2002. Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Research* 30, no. 24:5382–5390.

Leipe, D.D., L. Aravind, and E.V. Koonin. 1999. Did DNA replication evolve twice independently? *Nucleic Acids Research* 27, no. 17:3389–3401.

Liu, L., K. Komori, S. Ishino, A.A. Bocquier, I.K.O. Cann, D. Kohda, and Y. Ishino. 2001. The archaeal DNA primase: Biochemical characterization of the p41-p46 complex from *Pyrococcus furiosus*. *Journal of Biological Chemistry* 276 no. 48:45484–45490. doi: 10.1074/jbc.M106391200.

Malys, N., and J.E. McCarthy. 2011. Translation initiation: Variations in the mechanism can be anticipated. *Cellular and Molecular Life Sciences* 68, no. 6:991–1003. doi: 10.1007/s00018-010-0588-z.

Mardanov, A.V., and N.V. Ravin. 2012. The impact of genomics on research in diversity and evolution of archaea. *Biochemistry (Moscow)* 77, no. 8:799–812. doi: 10.1134/S0006297912080019.

Márquez, V., T. Fröhlich, J.P. Armache, D. Sohmen, A. Dönhöfer, A. Mikolajka, O. Berninghausen, et al. 2011. Proteomic characterization of archaeal ribosomes reveals the presence of novel archaeal-specific ribosomal proteins. *Journal of Molecular Biology* 405, no. 5:1215–1232. doi: 10.1016/j.jmb.2010.11.055.

Meinhart, A., T. Kamenski, S. Hoeppner, S. Baumli, and P. Cramer. 2005. A structural perspective of CTD function. *Genes and Development* 19, no. 12:1401–1415. doi: 10.1101/gad.1318105.

Moss, T., F. Langlois, T. Gagnon-Kugler, and V. Stefanovsky. 2007. A housekeeper with power of attorney: the rRNA genes in ribosome biogenesis. *Cellular and Molecular Life Sciences* 64, no. 1:29–49. doi: 10.1007/s00018-006-6278-1.

Nakao, A., M. Yoshihama, and N. Kenmochi. 2014. RPG (Ribosomal Protein Gene Database). Accessed October 2014. http://ribosome.med.miyazaki-u.ac.jp.

Napolitano, G., L. Lania, and B. Majello. 2014. RNA polymerase II CTD modifications: how many tales from a single tail. *Journal of Cellular Physiology* 229, no. 5:538–544. doi: 10.1002/jcp.24483.

Nasir, A., K.M. Kim, and G. Caetano-Anollés. 2014. Global patterns of protein domain gain and loss in superkingdoms. *PLoS Computational Biology* 10, no. 1:e1003452. doi: 10.1371/journal.pcbi.1003452.

O'Malley, M.A., and E.V. Koonin. 2011. How stands the Tree of Life a century and a half after *The Origin*? *Biology Direct* 6:32. doi: 10.1186/1745-6150-6-32.

Panse, V.G., and A.W. Johnson. 2010. Maturation of eukaryotic ribosomes: acquisition of functionality. *Trends in Biochemical Sciences* 35, no. 5:260–266. doi: 10.1016/j.tibs.2010.01.001.

Paris, Z., I.M. Fleming, and J.D. Alfonzo. 2012. Determinants of tRNA editing and modification: avoiding conundrums, affecting function. *Seminars in Cell Development and Biology* 23, no. 3:269–274. doi: 10.1016/j.semcdb.2011.10.009.

Phillips, G., and V. de Crecy-Lagard. 2011. Biosynthesis and function of tRNA modifications in Archaea. *Current Opinion in Microbiology* 14, no. 3:335–341.

Rackham, O., and A. Filipovska. 2014. Supernumerary proteins of mitochondrial ribosomes. *Biochimica et Biophysica Acta* 1840, no. 4:1227–1232. doi: 10.1016/j.bbagen.2013.08.010.

Raymann, K., P. Forterre, C. Brochier-Armanet, and S. Gribaldo. 2014. Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in Archaea. *Genome Biology and Evolution* 6, no. 1:192–212. doi: 10.1093/gbe/evu004.

Rivera, M.C., R. Jain, J.E. Moore, and J.A. Lake. 1998. Genomic evidence for two functionally distinct gene classes. *Proceedings of the National Academy Sciences USA* 95, no. 11:6239–6244.

Sanchez Mde, L., and C. Gutierrez. 2009. Novel insights into the plant histone code: lessons from ORC1. *Epigenetics* 4, no. 4:205–208.

Sarmiento, F., F. Long, I. Cann, and W.B. Whitman. 2014. Diversity of the DNA replication system in the *Archaea* domain. *Archaea* 2014:675946. doi: 10.1155/2014/675946.

Sarmiento, F., J. Mrázek, and W.B. Whitman. 2013. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *Proceedings of the National Academy Sciences USA* 110, no. 12:4726–4731. doi: 10.1073/pnas.1220225110.

Scherl, A., Y. Couté, C. Déon, A. Callé, K. Kindbeiter, J.-C. Sanchez, A. Greco, D. Hochstrasser, and J.-J. Diaz. 2002. Functional proteomic analysis of human nucleolus. *Molecular Biology of the Cell* 13. no. 11:4100-4109. doi: 10.1091/mbc.E02-05-0271.

Selvadurai, K., P. Wang, J. Seimetz, and R.H. Huang. 2014. Archaeal Elp3 catalyzes tRNA wobble uridine modification at C5 via a radical mechanism. *Nature: Chemical Biology* 10, no. 10:810–812. doi: 10.1038/nchembio.1610.

Shabalina, S.A., N.A. Spiridonov, and A. Kashina. 2013. Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity. *Nucleic Acids Research* 41, no. 4:2073–2094. doi: 10.1093/nar/gks1205.

Shatsky, I.N., S.E. Dmitriev, I.M. Terenin, and D.E. Andreev. 2010. Cap- and IRES-independent scanning mechanism of translation initiation as an alternative to the concept of cellular IRESs. *Molecules and Cells* 30, no. 4:285–293. doi: 10.1007/s10059-010-0149-1.

Siddiqui, K., K.F. On, and J.F. Diffley. 2013. Regulating DNA replication in eukarya. *Cold Spring Harbor Perspectives in Biology* 5, no. 9. doi: 10.1101/cshperspect.a012930.

Strunk, B.S., and K. Karbstein. 2009. Powering through ribosome assembly. *RNA* 15, no. 12:2083–2104. doi: 10.1261/rna.1792109.

Tye, B.K. 2000. Insights into DNA replication from the third domain of life. *Proceedings of the National Academy of Sciences USA* 97, no. 6:2399–2401.

Vesper, O., S. Amitai, M. Belitsky, K. Byrgazov, A.C. Kaberdina, H. Engelberg-Kulka, and I. Moll. 2011. Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in *Escherichia coli*. *Cell* 147, no. 1:147–157. doi: 10.1016/j.cell.2011.07.047.

Vinayak, M., and C. Pathak. 2010. Queuosine modification of tRNA: its divergent role in cellular machinery. *Bioscience Reports* 30, no. 2:135–148.

Walter, P., F. Klein, E. Lorentzen, A. Ilchmann, G. Klug, and E. Evguenieva-Hackenberg. 2006. Characterization of native and reconstituted exosome complexes from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Molecular Microbiology* 62, no. 4:1076–1089. doi: 10.1111/j.1365-2958.2006.05393.x.

Wigley, D.B. 2009. ORC proteins: marking the start. *Current Opinion in Structural Biology* 19, no. 1:72–78. doi: 10.1016/j.sbi.2008.12.010.

Woese, C. R., and G. E. Fox. 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy Sciences USA* 74, no. 11:5088–5090.

Woese, C. R., R. Gutell, R. Gupta, and H. F. Noller. 1983. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiological Reviews* 47, no. 4:621–669.

Woese, C. R., O. Kandler, and M. L. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy Sciences USA* 87, no. 12:4576–4579.

Woese, C. R., L. J. Magrum, and G. E. Fox. 1978. Archaebacteria. *Journal of Molecular Evolution* 11, no. 3:245–251.

Yang, C., and J. W. Stiller. 2014. Evolutionary diversity and taxon-specific modifications of the RNA polymerase II C-terminal domain. *Proceedings of the National Academy Sciences USA* 111, no. 16:5920–5925. doi: 10.1073/pnas.1323616111.

Yip, W. S., N. G. Vincent, and S. J. Baserga. 2013. Ribonucleoproteins in archaeal pre-rRNA processing and modification. *Archaea* 2013:614735. doi: 10.1155/2013/614735.

Yutin, N., P. Puigbò, E. V. Koonin, and Y. I. Wolf. 2012. Phylogenomics of prokaryotic ribosomal proteins. *PLoS One* 7, no. 5:e36972. doi: 10.1371/journal.pone.0036972.

Ziková, A., A. K. Panigrahi, R. A. Dalley, N. Acestor, A. Anupama, Y. Ogata, P. J. Myler, and K. Stuart. 2008. *Trypanosoma brucei* mitochondrial ribosomes: Affinity purification and component identification by mass spectrometry. *Molecular and Cellular Proteomics* 7, no. 7:1286–1296. doi: 10.1074/mcp.M700490-MCP200.

Zimmer, C. 2009. On the origin of eukarya. *Science* 325 no. 5941:666–668. doi: 10.1126/science.325_666.

Zimorski, V., C. Ku, W. F. Martin, and S. B. Gould. 2014. Endosymbiotic theory for organelle origins. *Current Opinion in Microbiology* 22C:38–48. doi: 10.1016/j.mib.2014.09.008.

Appendix A

Alignment of RNA polymerase subunits, adapted from Figure S1 of (Kusser et al. 2008):

1. *S. cerevisiae* RPB4 vs *P. furiosus* F

```
RBP4    MNVSTSTFQTRRRRLKKVEEEENAATLQLGQEFQLKQINHQGEEEELIAL 50
F       ------------------------------------MIGRGKLGEKYI 12


RPB4    NLSEARLVIKEA------LVERR-RAFKRSQKKHKKKHLKHENANDETTA 93
F       TIAEAKELLLKRREEEVKAGIEEPLY---------------------- 38


RPB4    VEDEDDDLDEDDVNADDDDFMHSETREKELESIDVLLEQTTGGNNKDLKN 143
F       -------------------------------------------YEARL 43


RPB4    TMQYLTNFSRFRDQETVGAVIQLLKS--TGLHPFEVAQLGSLACDTADEA 191
F       ALEHAERFAKL-PADKAKEAVEELMNAFEWMSDRIACKIVDIMPEDSMDL 92


RPB4    KTLIPSLNNK---ISDDELERILKELSNLETLY 221
F       RVIFAKEEYQPTQEEMKQILDILDKYRE----- 120
```

2. *S. cerevisiae* RPB5 vs *P. furiosus* H

```
RPB5    MDQENERNISRLWRAFRTVKEMVKDRGYFITQEEVELPLEDFKAKYCDSM 50
H       -------------------------------------------------


RPB5    GRPQRKMMSFQANPTEESISKFPDMGSLWVEFCDEPSVGVKTMKTFVIHI 100
H       ------------------------------------------------- 2


RPB5    QEKNFQTGIFVYQNNITPSAMKLVPSIPPATIETFNEAALVVNITHHELV 150
H       -------------------------------MAGKKEFSIFDHVLV 15


RPB5    PKHIRLSSDEKRELLKRYRLKESQLPRIQRADPVALYLGLKRGEVVKIIR 200
H       PEHRVLSEEEKKALLEKYKITLAQLPQIKASDPAVKALGAKPGDVIEIKR 65


RPB5    KSETSGRYASYRICM-- 215
H       KSPTAGVYYYYRVVVED 82
```

3. *S. cerevisiae* RPB6 vs *P. furiosus* K

```
RPB6   MSDYEEAFNDGNENFEDFDVEHFSDEETYEEKPQFKDGETTDANGKTIVT 50
K      -------------------------------------------------- 3


RPB6   GGNGPEDFQQHEQIRRKTLKEKAIPKDQRATTPYMTKYERARILGTRALQ 100
K      -----------------------------MFKYTRFEKARIIGARALQ 19


RPB6   ISMNAPVFVDLEGETDPLRIAMKELAEKKIPLVIRRYLPDGSFEDWSVEE 150
K      ISMGAPVLIDVPPGITPLEAAILEFEKGVIPITVIRPS------------ 57


RPB6   LIVDL 155
K      -----
```


Appendix B

Alignments of general transcription factors:

1. *S. cerevisiae* TBP YER148W vs *P. furiosus* TBP PF1295

```
         1                                                  50
YER148W  MADEERLKEF KEANKIVFDP NTRQVWENQN RDGTKPATTF QSEEDIKRAA
PF1295   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~


         51                                                 100
YER148W  PESEKDTSAT SGIVPTLQNI VATVTLGCRL DLKTVALHAR NAEYNPKRFA
PF1295   ~~~~~~MVDM SKVKLRIENI VASVDLFAQL DLEKVLDLCP NSKYNPEEFP


         101                                                150
YER148W  AVIMRIREPK TTALIFASGK MVVTGAKSED DSKLASRKYA RIIQKIGFAA
PF1295   GIICHLDDPK VALLIFSSGK LVVTGAKSVQ DIERAVAKLA QKLKSIGVKF


         151                                                200
YER148W  KF.TDFKIQN IVGSCDVKFP IRLEGLAFSH GTFSSYEPEL FPGLIYRMVK
PF1295   KRAPQIDVQN MVFSGDIGRE FNLDVVALT. LPNCEYEPEQ FPGVIYRVKE


         201                                      248
YER148W  PKIVLLIFVS GKIVLTGAKQ REEIYQAFEA IYPVLSEFRK M~~~~~~~
PF1295   PKSVILLFSS GKIVCSGAKS EADAWEAVRK LLRELDKYGL LEEEEEEL
```


2. *S. cerevisiae* TFIIB, YPR086W vs *P. furiosus* TFIIB PF1377

```
         1                                                  50
YPR086W  MMTRESIDKR AGRRGPNLNI VLTCPECKVY PPKIVERFSE GDVVCALCGL
PF1377   ~~~~~~~~~~ ~~~~~~~MNK QKVCPACESA ..ELIYDPER GEIVCAKCGY


         51                                                 100
YPR086W  VLSDKLVDTR SEWRTFSNDD HNGDDPSRVG EASNPLLDGN NLSTRIGKGE
PF1377   VIEENIIDMG PEWRAFDASQ RE..RRSRTG APESILLHDK GLSTEIGIDR


         101                                                150
YPR086W  TTDMRFTKEL NK...AQ.GK NVMDKKDNEV QAAFAKITML CDAAELPKIV
PF1377   SLSGLMREKM YRLRKWQSRL RVSDAAERNL AFALSELDRI TAQLKLPRHV
```

```
        151                                                        200
YPR086W KDCAKEAYKL CHDEKTLKGK SMESIMAASI LIGCRRAEVA RTFKEIQSLI
PF1377  EEEAARLYRE AVRKGLIRGR SIESVMAACV YAACRLLKVP RTLDEIADIA

        201                                                        250
YPR086W HVKTKEFGKT LNIMKNILRG KSEDGFLKID TDNMSGAQNL TYIPRFCSHL
PF1377  RVDKKEIGRS YRFIARNLNL TPKKLFV... .......KPT DYVNKFADEL

        251                                                        300
YPR086W GLPMQVTTSA EYTAKKCKEI KEIAGKSPIT IAVVSIYLNI LLFQIPITAA
PF1377  GLSEKVRRRA IEILDEAYKR GLTSGKSPAG LVAAALYIAS LLEGEKRTQR

        301                                                   349
YPR086W KVGQTLQVTE GTIKSGYKIL YEHRDKLVDP QLIANGVVSL DNLPGVEKK
PF1377  EVAEVARVTE VTVRNRYKEL VEKLKIKVPI A~~~~~~~~~ ~~~~~~~~~~
```

3. *S. cerevisiae TFIIE large subunit,* YKL028W vs *P. furiosus* TFIIE PF0491

```
===      1                                                          50
YKL028W MDRPIDDIVK NLLKFVVRGF YGGSFVLVLD AILFHSVLAE DDLKQLLSIN
PF0491  MGRDKK...N TALLDIARDI GGDEAVEVVK ALEKKGEATD EELAELTGVR

        51                                                         100
YKL028W KTELGPLIAR LRSDRLISIH KQREYPPNSK SVERVYYYVK YP..HAIDAI
PF0491  VNTVRKILYA LYDAKLATFR RVRD.....D ETGWYYYYWR IDTKRLPEVI

        101                                                        150
YKL028W KWKVHQVVQR LKDDLDKNSE PNGYMCPICL TKYTQLEAVQ LLNFDRTEFL
PF0491  RTRKLQELEK LKQMLQEETS ETYYHCGTPG HP.....KLT FDEAFEYGFQ

        151                                                        200
YKL028W CSLCDEPLVE DDSGKKNKEK QDKLNRLMDQ IQPIIDSLKK IDDSRIEENT
PF0491  CPICGEILYE YDNSKIIEEL KKRIE..... .......... .......E..

        201                                                        250
YKL028W FEIALARLIP PQNQSHAAYT YNPKKGSTMF RPGDSAPLPN LMGTALGNDS
PF0491  LEIELGLRSP PKEEKPKKAT RRKKS..... RSGKKKK~~~ ~~~~~~~~~~

        251                                                        300
YKL028W SRRAGANSQA TLHINITTAS DEVAQRELQE RQAEEKRKQN AVPEWHKQST
PF0491  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

        301                                                        350
YKL028W IGKTALGRLD NEEEFDPVVT ASAMDSINPD NEPAQETSYQ NNRTLTEQEM
PF0491  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

        351                                                        400
YKL028W EERENEKTLN DYYAALAKKQ AKLNKEEEEE EEEEEDEEEE EEEEMEDVMD
PF0491  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

        401                                                        450
YKL028W DNDETARENA LEDEFEDVTD TAGTAKTESN TSNDVKQESI NDKTEDAVNA
PF0491  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

        451                                   484
YKL028W TATASGPSAN AKPNDGDDDD DDDDDEMDIE FEDV
PF0491  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~
```