



Article

The Ribosome as a Missing Link in Prebiotic Evolution III: Over-Representation of tRNA- and rRNA-Like Sequences and Plieofunctionality of Ribosome-Related Molecules Argues for the Evolution of Primitive Genomes from Ribosomal RNA Modules

Robert Root-Bernstein ¹ and Meredith Root-Bernstein ^{2,*}

¹ Department of Physiology, Michigan State University, 567 Wilson Road, Room 2201, East Lansing, MI 48824, USA; rootbern@msu.edu

² UMR SAD-APT, INRA, AgroParisTech, Université Paris-Saclay, 78850 Grignon, France

* Correspondence: meredithrootbernstein@gmail.com

Received: 11 December 2018; Accepted: 23 December 2018; Published: 2 January 2019



Abstract: We propose that ribosomal RNA (rRNA) formed the basis of the first cellular genomes, and provide evidence from a review of relevant literature and proteomic tests. We have proposed previously that the ribosome may represent the vestige of the first self-replicating entity in which rRNAs also functioned as genes that were transcribed into functional messenger RNAs (mRNAs) encoding ribosomal proteins. rRNAs also encoded polymerases to replicate itself and a full complement of the transfer RNAs (tRNAs) required to translate its genes. We explore here a further prediction of our “ribosome-first” theory: the ribosomal genome provided the basis for the first cellular genomes. Modern genomes should therefore contain an unexpectedly large percentage of tRNA- and rRNA-like modules derived from both sense and antisense reading frames, and these should encode non-ribosomal proteins, as well as ribosomal ones with key cell functions. Ribosomal proteins should also have been co-opted by cellular evolution to play extra-ribosomal functions. We review existing literature supporting these predictions. We provide additional, new data demonstrating that rRNA-like sequences occur at significantly higher frequencies than predicted on the basis of mRNA duplications or randomized RNA sequences. These data support our “ribosome-first” theory of cellular evolution.

Keywords: ribosome; rRNA; mRNA; tRNA; SINE; LINE; genome; ribosome-binding protein; aminoacyl tRNA synthetase; extra-ribosomal; non-canonical; paralog

1. Introduction: The “Ribosome-First” Theory Predicts that tRNA- and rRNA-Like Genes should be Over-Represented in Cellular Genomes

Orthologous genes conserved across all species constitute what is called the Universal Gene Set of Life (UGSL), and consists of less than 100 genes [1–3]. Not surprisingly, the UGSL is dominated by translation-related genes. These ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), and ribosomal proteins make up the most abundant macromolecular species in all living organisms [4,5]. More surprisingly, as we will demonstrate here, macromolecules based on rRNA-like, tRNA-like and ribosomal-protein-like modules also dominate the rest of the genome, playing roles far removed from translation. We argue that this over-abundance of ribosome-like sequences can only be explained if ribosomes evolved prior to cells providing, through an rRNA-based genome, the basis from which

subsequent cellular genomes were elaborated. We call this ribosomal origin of the genome the “ribosome-first” theory of the origins of cellular life.

This “ribosome first” theory of the origin of cellular genomes follows directly from our previous work, in which we have suggested that ribosomes were the first self-replicating entities from which modern cells evolved. The “ribosome-first” hypothesis makes a series of surprising predictions, including the necessity for rRNA to encode genetic information containing ribosomal-protein-related functions; to function as messenger RNA (mRNA) as well as rRNA; and to encode the tRNAs that are required to implement translation. In previous papers, we demonstrated that rRNA does appear to encode, in one or more reading frames, the genetic information required to produce many of the active-site protein segments that are associated with its own functions (including ribosomal proteins, polymerases, ligases, amino acid transferases, and phosphatases). The rRNA also encodes at least two complete sets of tRNAs to translate its own sequence into these protein segments. Also, rRNA appears to have the properties of messenger RNAs (mRNAs). In some cases, rRNA has retained this mRNA function by encoding proteins that are translated in modern organisms. In addition, modern ribosome-binding proteins, as well as many amino acid transferases (synthetases) have the unique property of autologously regulating their own synthesis by binding to their own mRNA. We attribute this autologous control feature to the evolution of a self-sufficient regulatory network of molecular interactions within the ribosomal self-replication schema [6–8]. These observations alter the basic conception of the ribosome from a specialized molecular machine that evolved, following an established RNA-world to one in which the ribosome is a vestigial remnant of a prebiotic self-replicating entity that contained genes, maintained a limited metabolism, and evolved into the first cells.

A logical prediction of our “ribosome-first” theory of the origins of cellular life, which we explore for the first time here, is that the ribosomal genome preceded the evolution of the cellular genome and gave rise to it. Divergent evolution of the ribosomal genes through duplication and modification would have produced the fundamental core of modern cellular genomes. Ribosomal RNAs themselves appear to have evolved from the ligation of multiple tRNA-like modules [6,9–11]. Modern genomes should therefore contain an unexpectedly large percentage of tRNA- and rRNA-like modules derived from both sense and antisense reading frames. These molecules should therefore have been adapted to play multiple roles in primitive cells and should therefore be plieofunctional. Moreover, divergent evolution should have dispersed these tRNA- and rRNA-like genetic elements so that they should be found to be embedded within non-ribosomal proteins with key functions within the cell. In this paper, we review existing literature that provides evidence for all of these predictions of the ribosome-first hypothesis, and provide a series of proteomic tests of this hypothesis that distinguish its predictions from those derived from other possible hypotheses.

Testing the ribosome-first theory of genome evolution (or any other theory) requires not only marshalling evidence for it, but equally importantly, distinguishing between the predictions made from the theory being tested, and the predictions that stem from other possible explanations of cellular evolution. In general, origins of life theories fall into three categories: (1) RNA- or more broadly, “genome-first” theories; (2) protein- or “metabolism-first” theories; and (3) co-evolution theories, in which genome, proteins and metabolism emerge synergistically. The “ribosome-first” theory of cellular evolution is one version of possible co-evolution theories, since it posits an RNA genome encoding the proteins necessary to make a functional ribosome that could carry out replication, transcription, and translation [6,12].

The critical differences between these general theories are as follows. In a “genome-first” theory, the ribosome evolves as a late addition to a pre-existing, functional genome that evolved from an RNA- or RNA–DNA-world. The pre-existing genome should be the basis for the emerging cellular genome, from which protein metabolism would emerge, with ribosome-related genes representing a small, specialized, late supplement to that genome. Extra-ribosomal functions for rRNAs, tRNAs, and ribosomal proteins should have evolved only rarely, and rRNAs should not contain any genetic

information of their own. In this case, randomly evolved sequences should have been the basis for the first genomes with selection for functionality following from selection on the diversity of sequences.

In the “metabolism-first” theory, the cell, presumably including its diverse protein components, evolves prior to the genome or the ribosome. The genome then evolves to encode these pre-adapted proteins, such as RNA and/or DNA genes. The ribosome would, in a metabolism-first scenario, play a subsidiary role as a mere translation mechanism, and therefore, once again, be encoded by a small number of late-evolving genes, and rRNA should not contain any genetic information of their own. rRNA and ribosomal protein-related sequences should, once again, be rarely encountered in the wider genome and ribosome-related proteins should only rarely have evolved extra-ribosomal functions. Instead, protein-encoding genes should be found to be over-represented and modified throughout the genome.

However, in a “ribosome-first” theory, the first self-replicating organism is the ribosome itself, and the first genome that of the ribosome. Thus, both protein metabolism and a core genome pre-date the evolution of cells. In the case of the ribosome-first theory, this integration specifically arises within the ribosome complex itself, such that rRNA-encoded genes and ribosome-related proteins and their metabolic functions emerge as a tightly knit network or interactome. This interactome should be the basis from which many cellular functions evolve, and therefore the cellular genome would have evolved from this small set of ribosome-related genes. Moreover, rRNAs, tRNAs, and ribosomal proteins would have been co-opted to carry out key non-translation related functions during cellular evolution. In short, the “ribosome-first” theory predicts that ribosome-like genes will be significantly over-represented in cellular genomes, and that rRNA, tRNA and ribosomal proteins should all be pleiofunctional.

The purpose of this article is to aggregate and review various types of evidence acquired by many investigations of the nature of genome organization in light of the genes-first, metabolism-first and ribosome-first theories, and to provide proteomic tests of whether ribosome-related genes are significantly over-represented in cellular genomes, as compared with what one might expect from random evolution of sequences or from a protein-first origin.

2. Results

2.1. Eukaryotic and Prokaryotic Genomes Contain Unexpectedly Large Numbers of rRNA-Like Sequences

All cellular genomes contain unexpectedly large numbers of rRNA-like sequences, an observation that was first reported by Mauro and Edelman in 1997 for eukaryotes [13]: “Many eukaryotic mRNAs contain sequences that resemble segments of 28S and 18S rRNAs, and these rRNA-like sequences are present in both the sense and antisense orientations. Some are similar to highly conserved regions of the rRNAs, whereas others have sequence similarities to expansion segments. In particular, four 18S rRNA-like sequences are found in several hundred different genes, and the location of these four sequences within the various genes is not random. One of these rRNA-like sequences is preferentially located within protein coding regions immediately upstream of the termination codon of a number of genes We consider the hypotheses that rRNA-like sequences may have spread throughout eukaryotic genomes and that their presence in primary transcripts may differentially affect gene expression.”

Several lines of evidence supported Mauro and Edelman’s conjecture. Mauro and Edelman themselves demonstrated that regions within mRNAs that were complementary to rRNA bound to rRNAs in the ribosome resulting in translational control of gene products. The Gtx homeodomain protein is one well-characterized example [14–16]. Several investigators have also demonstrated the presence of rRNA-like sequences in a wide array of poly-adenylated mRNAs. Poly-adenylation is a process in which adenines are added to the end of mRNA during transcription, to aid in the transport of the mRNA to the ribosome, and to identify the RNA as carrying a translatable message. Thus, the

presence of rRNA-like sequences in polyadenylated RNAs provides further evidence of a role for rRNA sequences in identifying functional proteins.

Mauro and Edelman, for example, found that [13]: “Northern blot analysis of poly(A) RNA from different vertebrates (chicken, cattle, rat, mouse, and human) revealed that a large number of discrete RNA molecules hybridize at high stringency to cloned probes prepared from the 28S or 18S rRNA sequences that were found to match those in mRNAs. Inhibition of polymerase II activity, which prevents the synthesis of most mRNAs, abolished most of the hybridization to the rRNA probes.” Kong, et al. [17] verified that rRNA-like sequences are unusually abundant in poly-adenylated mRNA. Kong, et al. [17] also identified two different types of such poly-adenylated rRNA-containing transcripts (PART). Type I PART contain sequences of tens to several hundreds of rRNA-associated nucleotides within the transcript itself, while Type II PART contain very long, sometimes almost complete, rRNA sequences with a 5' cap and a 3' polyadenylation sequence. PART I include tRNA synthetases, RNA-binding proteins, insulin-like growth factor binding proteins, tumor necrosis factor, homeobox genes, heat shock 86 protein, connexins, glycine methyltransferase, nuclear respiratory factor 1 (NRF 1), and many other classes of proteins that are essential to cell functions. PART II include mitochondrial leucyl-tRNA synthetase, which is 98% identical to a 396 nucleotide stretch of the 28S rRNA; phosphocan short isoform (RPTP-beta gene), which is 98% of 1537 nucleotides of the 28S rRNA; superoxide dismutase 1 (SOD1) upregulated mRNA 1, which is identical to the first 1700 of 1879 nucleotides of 18S rRNA; human humanin (HN1), which is identical to 1553 nucleotides of the 28S rRNA (see also [18]); human serine/threonine protein kinase Kp78, CTAK75a, which is 99% identical to 905 nucleotides of the 18S rRNA; mouse ETS-related transcription factor ERF (Erf1), which is 95% identical to a 705 nucleotide stretch of the 18S rRNA; mouse phosphocan short isoform (RPTP-beta gene), which is 98% identical to 1573 nucleotides of the 28S rRNA; human succinate dehydrogenase complex, subunit A, which is 99% identical to 643 nucleotides of the 18S rRNA [17]. Kong, et al., conclude that, “It appears that rRNA segments are dispersed throughout the genome” [17].

Additional proteins that share significant portions of rRNA sequences have been identified by other investigators. One is the rDNA-encoded mitochondrial protein Tar1p (transcript antisense to ribosomal RNA) in *Saccharomyces cerevisiae* [19,20], which regulates glucose-related respiration. Coelho et al. argue that the existence of Tar1p may demonstrate, “that rDNA transcription and mitochondrial function are coordinately regulated in eukaryotic cells” [19]. Lee, et al. [18] add that Tar1p is likely to be only one of many such mitochondria-derived rRNA transcripts: “Many of the mRNA species identified from the mitochondria are discrete smaller length ones that do not map to the traditional mitochondrial protein-encoding genes. In particular, multiple such mRNAs are observed from the 16S rRNA region, including the site of the humanin ORF Using parallel analysis of RNA ends (PARE) to map the transcript cleavage sites, a plethora of expected and unexpected cleavage sites have been discovered for the mitochondria [21]. The majority of tRNAs and mRNAs have distinct dominant cleavage sites at the 5' termini, but intragenic cleavage sites are especially abundant in rRNAs. The latter stages of precursor modification include 3' polyadenylation of mRNAs, with the exception of the mitochondrial ND6 gene, and rRNAs, in agreement with previous reports [21,22]. In addition, mitochondrial rRNAs are transported and found in the cytoplasm with significant biological roles in, for instance, *Drosophila* germ cell establishment [23,24], and 16S rRNA was found to be localized in the nucleus of human spermatogenic cells [25]. It is possible that the polyadenylated humanin transcript is exported to the cytoplasm where it can be translated.”

Mignone and Pesole [26] have also reported very large numbers of rRNA-like sequences in the human genome, but in contradistinction to the papers just cited, have argued that the vast majority (especially of the longer, type II PART) are artifacts of homology searching, and that only a small subset of short, type I PART stand up to critical analysis of the data. Their own data analysis, however, argues against their conclusions, at least from a ribosome-first perspective. They note that while “the distribution of low-similarity matches was similar for both real and simulated rRNA sequences, . . . interestingly, . . . the long/highly similar matches can be found only with the actual rRNA sequences”

([26] p. 150). Oddly, then, they discarded all of the long sequence (type II PART) homologies as artifacts on the basis that, “The great majority of rRNA-like tracts found in human mRNAs were not supported by counterparts in genomic or EST [expressed sequence tag] sequences” ([26] p. 151)—a statement in direct contradiction to the results of Mauro and Edelman [13] and Kong [17]. In any event, Mignone’s and Pesole’s requirement that rRNA-like sequences actually be transcribed is not required by the ribosome-first theory, which demands only that rRNA-like sequences be found in unexpectedly high numbers compared to the numbers of genes that are required to encode ribosome production. It is quite possible that organismal genomes evolved from duplicated rRNAs and subsequently transferred protein-encoding functions to more specialized exon regions. We believe that this is, in fact, exactly what Mignone and Pesole found. Additionally, our own analysis of the occurrence of rRNA-like sequences in the genome, which we provide below, reveals a possible flaw in the method by which Mignone and Pesole analyzed their results.

A third set of evidence also argues for the presence of rRNA-derived sequences in cellular genomes at unusually high rates. Gene sequences that are complementary to rRNA genes also populate the genome in unexpectedly large numbers, perhaps representing the incorporation of transcription products of rRNA “genes” into the genome. Examples include complementarity between ferritin H mRNA and 28S ribosomal RNA [27] between the avian myeloblastosis oncogene and eukaryotic 28S ribosomal RNA [28], and murine 18S rRNA with a wide range of mRNAs [29]. Such complementary sequences are not predicted by the genome-first or metabolism-first theories of cellular evolution, but they follow directly from the ribosome-first theory in which all six reading frames of rRNA are utilized to encode protein sequences.

Other data also take on different interpretations, depending on the theoretical framework from which they are interpreted. Genome-wide sequencing studies have repeatedly run into the problem of “discovering” new genes and new gene families within or across overlapping genome regions that encode rRNA. The ribosome-first theory of cellular evolution explicitly proposes that such overlaps not only should, but they must exist, since the rRNAs must contain the genes encoding for their own ribosomal proteins. Thus, the protein-encoding regions must be contained within, or overlap, the rRNA sequences [6–8]. No such prediction follows from the genome-first or the metabolism-first theories of cellular evolution. In a genome-first theory, the ribosome would have evolved after the core genome in order to make possible protein production, in which case protein-encoding regions that overlapped with rRNAs would interfere with pre-existing gene function. Similarly, if the ribosome evolved to optimize pre-existing protein metabolism, then key protein sequences should be the basis for rRNA evolution rather than rRNAs serving as the basis for protein evolution.

In fact, a robust literature clearly demonstrates that rRNAs do overlap both known and possible protein-encoding sequences, and there is no evidence that specific protein-encoding sequences are the origins of rRNAs. For example, metagenomic studies of RNA expression generally yield a predominance of rRNA-like sequences that can approach 90% of the total RNA sequenced [30–35]. Tripp, et al. [36] have provided a good summary of the type of data that are treated as artefactual: “In the course of analyzing 9,522,746 pyrosequencing reads from 23 stations in the Southwestern Pacific and equatorial Atlantic oceans, it came to our attention that misannotations of rRNA as proteins is now so widespread that the rate of false-positive matching of rRNA pyrosequencing reads to the National Center for Biotechnology Information (NCBI) non-redundant protein database approaches 90%. One conserved portion of 23S rRNA was consistently misannotated often enough to prompt curators at Pfam to create a spurious protein family” ([36], p. 8792). This “problem” of pseudo-proteins being encoded within rRNA sequences, or overlapping them, is so widespread that programs have been written to aid researchers in eliminating them during genome-wide sequencing experiments (e.g., [30–35]). Without challenging whether proteins are actively encoded in rRNA-encoding sequences in prokaryotes and archaea, the ribosome-first theory does force us to consider the possibility that such apparent encodings are fossils of the ribosomal origins of the genome. At the same time, the widespread existence of such inclusions and overlaps makes no sense from other genome origins

theories. From an evolutionary perspective, we must confront the question of why so many “spurious” rRNA-like sequences exist.

It is also important to point out that the desire to eliminate “false” protein families from genome-wide sequencing studies is based on three key points. One is the assumption that rRNA-like sequences should not appear ubiquitously in the genome. The ribosome-first theory undercuts that assumption. The second point is that the argument for eliminating rRNA-like sequences is based on a lack of evidence that they are proper annotations. Lack of evidence is not evidence against their possible validity. Yet, this is precisely the current reasoning. As Tripp, et al. point out, “rRNA operons in Bacteria and Archaea are not known to contain naturally expressed protein coding regions that also code for rRNA” ([36] p. 8792). Tripp, et al., therefore consider that, “Annotations of Bacteria and Archaea proteins embedded in rRNA operons and overlapping with rRNA coding regions within those operons have been *rightly* [emphasis added] presumed to be misannotations [37] and should continue to be, until hard evidence to the contrary emerges” ([36] p. 8792). From a ribosome-first perspective, we would strongly urge a search for such “hard evidence”, since it seems unlikely that dual-use, overlapping protein-rRNA sequences evolved ubiquitously without some function, either past or present. Moreover, there is no rationale other than a ribosome-first origin, for the genome to explain why rRNA-like sequences would be both ubiquitous and meaningless, and that rationale would hold only if functions originally associated with rRNA genes were off-loaded onto other genes during evolution.

In fact, many dual-use genes encoding both proteins and rRNAs are known in eukaryotes [13,20,24, 37–39]. Recently, Mumtaz and Couso [40] have used Ribo-Seq techniques to examine ribosome-bound mRNAs, and they discovered large numbers of rRNA-like sequences bound to the ribosomes that were apparently translated, and which had previously been ignored as “non-coding” sequences: “Earlier bioinformatic approaches had shown the presence of hundreds of thousands of putative small ORFs (smORFs) in eukaryotic genomes, but they had been largely ignored, due to their large numbers and the difficulty in determining their translation and function. Ribo-Seq has revealed that hundreds of putative smORFs within previously assumed long non-coding RNAs (lncRNAs) and UTRs of canonical mRNAs are associated with ribosomes, appearing to be translated” [40].

One must consider an additional hypothesis, which is that dual-use sequences may have been very common during the emergence of cellular life, but that selection pressures have largely separated rRNA-encoding sequences from protein-encoding sequences, so as to avoid regulatory confusion. The evolutionary problem that must be faced, regardless of which perspective is taken then, is to explain the origins of ubiquitous rRNA-protein overlaps, whether they are non-functional fossils or functional overlaps. This problem currently appears to be intractable from anything other than the ribosome-first theory. Indeed, this evolutionary problem exists for the origins of all rRNA-like genes within genomes. To quote Kong, et al. again: “For type II PART, there is a possibility that some of these RNA may derive from rRNA genes or rRNA. The origins of these types of RNA remain to be elucidated” [17].

2.2. Genomes Contain Unexpectedly Large Numbers of rRNA- and tRNA-Derived Transposon Sequences

According to the ribosome-first theory, not only should rRNA-like sequences be unusually prevalent in all cellular genomes, so should tRNA-like sequences. This prediction follows from the fact that tRNA-like sequences appear in multiple reading frames within rRNAs [6–11]. Indeed, recent research has revealed that prokaryotic and eukaryotic genomes contain large numbers of short interspersed repetitive elements (SINE) derived from fragments of rRNAs and tRNAs (SINE1) or 7SL (Alu) (SINE 2), as well as mammalian-wide interspersed repeats (MIRs). These sequences may provide evidence for how rRNA (or tRNA)-derived ribosome-encoded genes multiplied to form cellular genomes.

Alu (named after the *Arthrobacter luteus* restriction endonuclease), or SINE2, are abundant transposable elements within genomes. While the genes encoding ribosomal RNA, tRNAs, and ribosomal proteins (rDNA) comprise only 0.01% of the human genome, and exons comprise only

1.5% of the human genome [41], tRNA-like Alu and SINES make up about 11% to 15% of mammalian genomes, and are largely located in intron regions [41,42]. MIRs make up another 1–2% of most organismal genomes [43].

Investigators have also found increasing numbers of new types of SINE derived from 5S rRNA (SINE3) and 28S rRNA (SINE 28). Longo, et al. [44] provide a good summary of the new findings: “Complex eukaryotic genomes are riddled with repeated sequences whose derivation does not coincide with phylogenetic history, and thus is often unknown. Among such sequences, the capacity for transcriptional activity, coupled with the adaptive use of reverse transcription, can lead to a diverse group of genomic elements across taxa, otherwise known as selfish elements or mobile elements. Short interspersed nuclear elements (SINEs) are nonautonomous mobile elements found in eukaryotic genomes, typically derived from cellular RNAs such as tRNAs, 7SL, or 5S rRNA. Here, we identify and characterize a previously unknown SINE derived from the 3'-end of the large ribosomal subunit (LSU or 28S rDNA) and transcribed via RNA polymerase III. This new element, SINE28, is represented in low-copy numbers in the human reference genome assembly, wherein we have identified 27 discrete loci” [44]. Additional sequences derived from the 28S ribosome can be found scattered throughout the human genome in varying copy numbers [45,46].

Highly conserved 5S rRNA-derived SINEs (SINE3) have been reported in human and other mammals that Nishihara, et al. [47] have called AmnSINE1. “AmnSINE1 has a chimeric structure of a 5S rRNA and a tRNA-derived SINE, and is related to five tRNA-derived SINE families that we characterized here in the coelacanth, dogfish shark, hagfish, and amphioxus genomes. All of the newly described SINE families have a common central domain that is also shared by zebrafish SINE3, and we collectively name them the DeuSINE (Deuterostomia SINE) superfamily. Notably, of the approximately 1000 still identifiable copies of AmnSINE1 in the human genome, 105 correspond to loci phylogenetically highly conserved among mammalian orthologs.” This high degree of conservation strongly suggests evolutionary selection for functionality. Other examples of SINE3 have been found in fruit bats [48], fungi [49], and zebrafish [50] among other organisms [51]. In zebrafish, Kapitonov and Jurka report that, “SINE3s are transcribed from the type 1 internal pol III promoter. Approximately 10,000 copies of SINE3 elements are present in the zebrafish genome; they constitute approximately 0.4% of the genomic DNA” [50]. The 1000 or so SINE3 found in the human genome constitute an equivalent percentage.

In light of the fact that SINES and MIRs make up 15% to 20% of many genomes, Frenkel, et al. [52] and Schmitz [42] have proposed that SINEs may be the drivers of genome evolution. If that hypothesis is correct, then the further possibility must be entertained that the tRNA and rRNA origins of SINEs and MIRs are molecular “fossils” of the mechanism by which the original cellular genomes were formed through self-replicating, or insertion-amplifying rRNA and tRNA sequences. Such a mechanism would be entirely consistent with a ribosome-first origin of cellular genomes, but once again, this is difficult to explain from other theoretical perspectives.

2.3. tRNA and rRNA Control of Non-Ribosomal Cellular Functions

A further unique, testable implication of the ribosome-first theory of genome origins is that ribosomal proteins, rRNAs and tRNAs should be pleiofunctional, playing roles in cellular functions beyond translation. This prediction follows from the assumption that the ribosomal molecules were the foundation from which other core cellular molecules and their functions evolved. The interactivity of the ribosomal complex itself would have assured a high degree of interactivity among the genes and proteins that evolved from it, creating a stable interactome. Therefore ribosomally-derived molecules would have been under significant selection pressure to be retained in primordial cells, and to become adapted to the broader range of functions that the cell required. Evidence for such a scenario is abundant.

Wegrzyn and Wegrzyn [53] and Katz, et al. [54] have reviewed studies demonstrating that tRNA has multiple cellular functions beyond its canonical role in protein translation (Table 1). Various

retroplasmids and retroviruses encode a reverse transcriptase that can use a tRNA sequence as the basis for synthesizing a cDNA copies. DNA replication itself is regulated by the accumulation of uncharged tRNAs that bind to, and inactivate, several repressors of DNA synthesis. The proportion of charged to uncharged tRNAs functions as a critical sensor of amino acid deficiency in both prokaryotes and eukaryotes (though through different mechanisms), mediating transcription and translation to improve the metabolic efficiency of amino acid synthesis. Also, uncharged tRNAs can act as ribozymes for the self-deletion of introns, again fostering increased metabolic efficiency under stress conditions. In addition, tRNA transcript fragments averaging about 30 to 35 nucleotides in length have been demonstrated to play diverse cellular roles, mainly in controlling transcription and translation in both prokaryotes and eukaryotes [55]. These tRNA fragments are widespread and numerous in prokaryotic [56] (and eukaryotic cells [57] under stress (see also [58–62]). tRNA fragments not only interact with other short RNA to regulate their activity, but also sequester RNA-binding proteins, buffering the translation system. Some tRNA-like repeats encode proteins that function specifically as Rho-dependent transcriptional regulators [55]. Finally, eukaryotic tRNA genes sometimes function as “chromatin insulators”, separating active chromatin domains from silenced ones [58]. Thus, tRNA and its fragments play important metabolic roles in replication, transcription and translation beyond simply recognizing specific codons on mRNAs and providing amino acids for ligation into proteins.

Table 1. Non-canonical functions of transfer RNAs (tRNAs), their fragments, and mimics. Table based mainly on references [53–62]. See text for additional references. miRNA = microRNA; siRNA = small interfering RNA.

Non-Canonical tRNA Properties	Molecules Involved	Functions
Modification of Macromolecules		
Lipid aminoacylation	Phosphatidylglycerol synthetases	Protects against some antibiotics
Enzymatic modification of peptidoglycans	MurM, Fem aminoacyl transferases	Modifies peptidoglycan synthesis in cell wall; alters protein binding
Enzymatic modification of amino terminus of proteins	Leucyl/Phenylalanyl transferases; Arg-tRNA transferase	Targets proteins for degradation
Enzymatic modification of peptides	AlbC, PacB, VlmA, DhpH-C peptidyltransferase	Intracellular signaling; Metal chelation; Antibiotic formation
Synthesis of Small Metabolites		
Dipeptide synthesis	Diketopiperazines	Metal chelation
Tetrapyrrole synthesis	Heme; Chlorophyll	Energy transduction
Aminoacylation of tRNA	Gln; Asn; fMet; Cys; Sec	Makes up for missing tRNA
Regulatory Processes		
Bind directly to ribosome to modify activity	Deacylated tRNA	Stringent response: Form alarmones, ppGpp and pppGpp
DNA replication initiation	ColE1-type plasmids	Control of plasmid replication
Riboswitch control	Operons ; T-box	Transcription attenuation
Metabolism	Siderophores; Virulence factors	Stress responses; Aerobic/respiratory metabolism; Gene expression
Stress responses	Proteins and polysomes involved in siRNA binding & miRNA binding	Translation regulation
tRNA Mimicry		
Translation regulation	tmRNA (small transfer-messenger RNA; SsrA or 10Sa RNA) & tmRNA-specific binding protein (SmpB or small protein B).	Rescues stalled ribosomes
“t-elements” (tRNA-like elements) in mRNAs	3' and 5' ends of mitochondrial mRNA	Sites for RNAase P and RNAase Z activation of mRNA?
Short Interspersed nuclear elements (SINEs)	Genome (non-coding DNA)	Retrotransposons
Viral tRNA mimics	Viral mRNA	Hijacking translation machinery
tRNA Fragments		
tRNA halves	Outer membrane vesicles	Stress responses? Bacterial signaling?
tRNA-derived fragments (tRFs)	Ribosomal proteins; mRNAs; RNA-binding proteins	Interferes w/with peptidyl transferase; Act as miRNAs or siRNAs in post-transcriptional gene silencing; Buffer regulatory RNAs; Sequester RNA-binding proteins

tRNA-derived small RNA (tsRNA) or tRNA fragments (tRF) have also been identified as novel regulatory RNAs controlling a wide range of physiological and pathological processes in organisms

from Archaea through plants and animals [63–66]. tsRNA and tRF functions that have been identified include transcription regulation, cell proliferation, tumor genesis, stress response, intergenerational epigenetic inheritance, mRNA binding to ribosomes, amino acylation of tRNAs by amino acyl transferases, ribosomal protein function, and protection against retrotransposons such as SINEs [67–70].

Finally, tRNA participate in non-canonical forms of peptide synthesis independent of the ribosome, in conjunction with non-ribosomal peptide synthases. To quote Giessen and Marahiel, tRNA, “are also used as substrates and amino acid donors in a variety of other important cellular processes, ranging from bacterial cell wall biosynthesis, and lipid modification to protein turnover and secondary metabolite assembly” [71]. For example, tRNA^{Gly} [72] is involved in peptidoglycan synthesis, while tRNA^{Lys} participates in remodeling lipids [73]. Ribosome-free production of cyclic dipeptides (diketopiperazines), for example, results in molecules that are essential for iron chelation or other types of siderophore activity; quorum sensing in bacteria; anti-viral, anti-bacterial, or anti-fungal activity, thereby protecting organisms from their pathogens [74,75].

Notably, tRNA may also have provided the material from which rRNA evolved. 5S rRNA has significant homologies with tRNA, and probably evolved directly from one or more tRNA precursors [9–11,76]. Like tRNAs, the 5S rRNA has also evolved to take on broader cellular functions. The 5S rRNA has not only undergone an alternative splicing event that has exonized it within the gene of its own transcription regulator, TFIIIA (transcription factor for polymerase III A) in all land plants [77,78], but has also mutated into an RNA mimic that regulates the alternative splicing of TFIIIA [79]. The 5S rRNA can also exert extra-ribosomal function by binding to ribosomal proteins prior to their integration into functional ribosome complexes. One classic example is the control of the oncogene p53 by a 5S rRNA–RPL5–RPL11 complex that inhibits the Hdm2 protein that normally blocks p53 activity [80,81].

Like tRNA fragments, rRNA fragments (rRF) have also been found to exert a range of extra-ribosomal functions (Table 2). For example, Arkov, et al. [82] and Chernyaeva, et al. [83], “identified a fragment of the *Escherichia coli* 23S rRNA (nucleotides 74 to 136) whose expression caused readthrough of UGA nonsense mutations (but not UAA or UAG) in certain codon contexts in vivo. The antisense RNA fragment produced a similar effect Since termination at UGA in *E. coli* specifically requires release factor 2 (RF2), our data suggest that the fragments interfere with RF2-dependent termination” [83]. Similarly, Feagen, et al. [84] reported high levels of rRNA fragments generated as functionally active transcripts from rDNA in *Plasmodium*. Chen, et al., [85] report that, “Two series of rRFs (rRF5 and rRF3) were precisely aligned to the 5' and 3' ends of the 5.8S and 28S rRNA genes. The rRF5 and rRF3 series were significantly more highly expressed than the rRFs located in the body of the rRNA genes. These series contained perfectly aligned reads, the lengths of which varied progressively, with 1-bp differences. The rRF5 and rRF3 series in the same expression pattern exist ubiquitously from ticks to human. The cellular experiments showed the RNAi knockdown of one 20-nt rRF3 induced the cell apoptosis and inhibited the cell proliferation.” Shi, et al. [86] have developed a program called “SPORTS1.0”, for annotating and profiling additional non-coding RNAs optimized for rRNA- and tRNA-derived small RNAs.

rRNA fragments can also function as interference RNAs. Hadjiargyrou and Delihias [87] have found that a large proportion of non-coding RNAs with regulatory functions derive from SINE elements, which are themselves often derived from tRNA-like sequences. Thus, tRNA and rRNA fragments appear to have fundamental regulatory roles in cellular metabolism that predate the origins of cellular life and have been incorporated as essential components of cellular life. Cerutti and Casas-Mollano [88] suggest that interference functions for RNA were already in place prior to the origins of cellular life, making rRF and tRNA-derived small RNA fundamental aspects of ribosome function and evolution: “We have examined the taxonomic distribution and the phylogenetic relationship of key components of the RNA interference (RNAi) machinery in members of five eukaryotic supergroups. On the basis of the parsimony principle, our analyses suggest that a relatively complex RNAi machinery was already present in the last common ancestor of eukaryotes

and consisted, at a minimum, of one Argonaute-like polypeptide, one Piwi-like protein, one Dicer, and one RNA-dependent RNA polymerase. As proposed before, the ancestral (but non-essential) role of these components may have been in defense responses against genomic parasites such as transposable elements and viruses. From a mechanistic perspective, the RNAi machinery in the eukaryotic ancestor may have been capable of both small-RNA-guided transcript degradation as well as transcriptional repression, most likely through histone modifications.”

Table 2. Non-canonical functions of ribosomal RNA (rRNA) [82–88].

Non-Canonical rRNA Properties	Molecules Involved	Functions
Transcription regulation	5S rRNA acting on transcription factor for polymerase factor IIIA	Inhibits activation of RNA polymerase III-dependent transcription and related genes
Translation regulation	Ribosomal protein fragments	siRNA and RNAi inhibiting mRNA translation
Permits read-through of termination codons	23S rRNA fragment	Inhibits release factor 2 (RF2)-dependent translation termination
Protein activity control	5S rRNA-RPL5-RPL11 complex	Inhibits the Hdm2 protein which in turn regulates p53
Sources of novel transcripts	5.8S and 28S rRNA fragments	Translated into proteins

2.4. Non-Ribosomal Functions of Ribosomal Proteins

As surprising as it may be that tRNAs and rRNAs and their fragments participate directly in cellular functions beyond their canonical roles, one of the strongest lines of evidence that tRNA and rRNA provide the basis for cellular genomes comes from studies of widespread non-ribosomal function for ribosomal proteins and genes. As can be seen in the next four Tables, almost all ribosomal proteins exhibit extra-ribosomal functions. We have previously reviewed evidence for autologous regulation of most ribosomal proteins of their own translation by means of binding directly to their own mRNA [7], and they include only a summary of that information in Tables 3–6. The majority of ribosomal proteins, both in prokaryotes and eukaryotes, play additional roles in cellular metabolism, ranging from DNA repair, control of transcription and replication, and histone binding to the regulation of various enzymes and mRNA splicing as summarized in Tables 3–6 [81,89–112]. The degree to which these extra-ribosomal functions are integrated into cellular functions is exemplified by the participation of riboprotein L13a into the GAIT complex in eukaryotes. The GAIT complex is produced in response to interferon gamma activation and consists of five components: riboprotein L13A binding to a conserved region on mRNA encoding ceruloplasmin or related inflammation response proteins, along with glutamyl-prolyl-tRNA synthetase, NS1-associated protein-1, and glyceraldehyde 3-phosphate dehydrogenase. GAIT prevents the 43S ribosome complex from integrating into a completed ribosome [81,89]. We will discuss further roles for tRNA synthetases in the next section, only noting here that it is important to realize that the artificial division that we have adopted here of summarizing extra-ribosomal function by the class of the molecule hides important interactions among these classes that are of equal importance to understanding the self-replicating ribosome as a functional interactome.

Table 3. Extra-ribosomal functions of ribosomal proteins of the small subunit of prokaryotic ribosomes (PROK RP). Auto Reg = autologous regulation of ribosomal protein by binding to its own mRNA.

Prok RP	Auto REG	Other Extra-Ribosomal Functions of RP	References
S1	YES	Stimulates RNA polymerase; Phage Q β replicase	[7,89–91]
S2	YES	Regulates rpsB and tsf genes	[7,89]
S3	YES	DNA repair; NF- κ B signaling pathway; Apoptosis	[7,92]
S4	YES	Regulates rpsD gene and α operons; Transcription anti-termination; NUS transcription elongation complex	[7,89,90]
S5	YES		[7]
S6	YES		[7]
S7	YES		[7]
S8	YES	Repressor of spc operon	[7,89]
S9	YES	SOS DNA repair system	[7,89,91]
S10	YES	Transcription anti-termination; NUS transcription elongation complex	[7,89–91]
S11	YES		[7]
S12	YES	T4 phage RNA chaperone and intron splicing	[7,89,91]
S13	YES		[7]
S15	YES		[89]
S18	YES		[7]
S19	YES		[7]
S20	YES	Inhibition of ornithine and arginine decarboxylase enzymes	[7,89]
S21	YES		[7]
S26	YES		[7]
S28	YES		[7]

Table 4. Extra-ribosomal functions of ribosomal proteins of the large subunit of prokaryotic ribosomes (PROK RP). Auto Reg = autologous regulation of ribosomal protein by binding to its own mRNA.

Prok RP	Auto REG	Other Extra-Ribosomal Functions of RP	References
L1	YES		[7,89]
L2	YES		[7]
L3		Transcription anti-termination	[7,89]
L4	YES	Transcription anti-termination; inhibits S10 transcription; regulates RNAase E	[7,89,90]
L6	YES		[7]
L7/12	YES		[7]
L10	YES		[7,89]
L11	YES		[7]
L13		Transcription anti-termination	[7,89]

Table 4. Cont.

Prok RP	Auto REG	Other Extra-Ribosomal Functions of RP	References
L14		Phage replication helicase	[7,91]
L19	YES		[7]
L20	YES		[7]
L24	YES		[7]
L25	YES		[7]
L30	YES		[7]
L31	YES		[7]
L32	YES		[7]
L35	YES		[7]

Table 5. Extra-ribosomal functions of ribosomal proteins of the small subunit of eukaryotic ribosomes (EUK RP). Auto Reg = autologous regulation of ribosomal protein by binding to its own mRNA.

Euk RP	Auto REG	Extra-Ribosomal Functions of RP	Reference
SA		67 kD Laminin receptor	[93]
S0+S21		Promotes 18S rRNA maturation	[89]
S2		Methylation of Rps2p; Laminin receptor	[95,96]
S3		Identical to apurinic-apyrimidinic DNA endonuclease III; functions as DNA β -lyase; binds NFkB; Activates p53 tumor suppression gene; Cdk2- mediated cell cycle regulation; Inhibits RECQL4 ATPase/helicase; Substrate of Akt mediating neuronal apoptosis and DNA repair; Binds HSP 90/70; Binds the transcription factor CHOP	[89–91,95,97–101]
S4		Cysteine protease	[102]
S6	YES	Translation regulator; Replication regulator; Enhances virus replication	[7,95,97]
S7		Binds MDM2; Activates p53 tumor suppression gene; Regulates MAPK; Binds to BCCIP β	[97,101,103]
S8		Identical to p27, an nuclear oocyte peripheral membrane protein	[91]
S9	YES	Translation regulator	[7,95]
S12	YES	RNA splicing and modification; Karyopherin Kap121-dependent sequestration of ribonucleotide reductase	[7,89,90,104]
S13	YES	Regulates own intron and mRNA splicing	[7,89]
S14	YES	Inhibits own mRNA splicing; Regulates own transcription; Represses CRY2 expression; Activates p53 & Tap73 tumor suppression genes; Inactivates c-Myc	[7,89,90,95,97]
S15		Activates p53 tumor suppression gene	[97]
S19		Apoptosis; Inhibits MIF, ERK, NF- κ B; Erythropoiesis via C5a receptor binding	[93,94,97]
S20		Regulates Pol III transcription; Apoptosis; Activates p53 tumor suppression gene	[89,93,95,97]
S26	YES	Developmental processes	[7,89,95]
S27		Activates p53 tumor suppression gene	[97]
S28	YES		[7,90,95]
S30	YES		[7,89]

Table 6. Extra-ribosomal functions of ribosomal proteins of the large subunit of eukaryotic ribosomes (EUK RP). Auto Reg = autologous regulation of ribosomal protein by binding to its own mRNA.

Euk RP	Auto REG	Extra-Ribosomal Functions of RP	Reference
L1	YES	Regulates gene expression	[7,95]
L2	YES		[7,90,95]
L3		Cell cycle regulation; Apoptosis	[105]
L4	YES	Regulates nucleolar RNA helicase 2 (Guα); Inhibits virus production	[7,89,90,97]
L5		Binds to MDM2; Activates p53 & Tap73 tumor suppression genes; Inactivates c-Myc; Stimulates aminoacyl tRNA synthases	[91,93,97]
L6		Regulates Pol III gene transcription; Activates p53 tumor suppression gene	[95,97]
L7	YES	Cell arrest; Apoptosis; Co-regulates VDR-RXR nuclear receptor complex	[7,93,101]
L10	YES	Translation regulator; Replication regulator; Binds c-Jun gene; Antiviral activity	[7,90,95]
L11		Regulates PPARα gene transcription; Activates p53 & Tap73 tumor suppression genes; Inactivates c-Myc	[89,97]
L12	YES		[7,90]
L13a		Regulates inflammatory gene expression and ceruloplasmin mRNA translation (GAIT complex); Anti-viral immunity; Binds to glyceraldehyde-3-phosphate dehydrogenase	[89,97,106,107]
L21		Apoptosis	[93]
L22		Binds histone H1; Binds EBV small RNA; modulates splicing of the pre-mRNA encoding smad2 (morphogenesis); Hematopoiesis; Binds to casein kinase 2α	[90,93,108–110]
L23		Activates p53 tumor suppression gene; Binds MDM2	[97]
L26		Activates p53 tumor suppression gene; Binds MDM2	[97]
L27		Apoptosis; Developmental control	[93,111]
L30	YES		[7,90]
L31		Apoptosis	[93]
L32	YES		[7,95]
L33A		Inhibits GCN4 translation	[95]
L37		Activates p53 tumor suppression gene	[97]
Asc1		Cell-wall integrity; Iron homeostasis; Energy metabolism	[95]
P-prots		Regulates phosphate transporters and phosphatases	[95]
P0/LP0		Endonuclease; DNA repair	[89,91,93]
P2		Iron-binding protein	[91]
RACK1		Receptor of activated C kinase (signal transduction)	[90]

As extensive as the currently documented extra-ribosomal functions of ribosomal proteins are, new functions continue to be discovered, which has led Warner and McIntosh to ask whether the fact that we have not found even more such functions is, “due to a lack of imaginative evolution by cells and viruses, or to a lack of imaginative experiments by molecular biologists?” [90]. Only further research can determine the answer. We predict that there are many more extra-translational ribosomal-protein functions to be found.

2.5. Non-Canonical Functions of Aminoacyl tRNA Synthetases (aaRS)

As noted in the previous section, yet another class of ribosome-related proteins that have important additional cellular functions are the aminoacyl tRNA synthetases. The function usually ascribed to these proteins is to conjugate the appropriate amino acid to the correct tRNA. We have previously demonstrated that key functional elements of tRNA synthetases are encoded in all six reading frames of rRNAs, so that ribosomes may have had the ability to produce both tRNAs and the core enzyme structures that are required to load them with the correct amino acids [6,8].

Caprara, et al. [113] and Mohr, et al. [114] reported that the *Neurospora crassa* mitochondrial (mt) tyrosyl-tRNA synthetase (CYT-18 protein) functions in splicing group I introns, in addition to aminoacylating tRNA(Tyr). Leucyl-tRNA synthetase has a similar function [115,116]. Lysyl-tRNA synthetase catalyzes the synthesis of dinucleoside polyphosphates in *E. coli* [117,118] while the threonyl-tRNA synthetase from the same organism exerts autogenous control on its own mRNA, to self-regulate its translation [119]. Other documented functions include the regulation of transcription, translation, splicing, inflammation (sometimes by mimicking cytokine activity) [120–122], angiogenesis and apoptosis [123], and cell signaling [124,125]. Some aaRSs interact specifically with other proteins [126]. For example, many of these functions are mediated in higher eukaryotes by an aaRS complex involving eight different aaRS and three aminoacyl-tRNA synthetase-interacting multifunctional proteins (AIMS). The complex is formed of ArgRS, GlnRS, IleRS, LeuRS, LysRS, MetRS, AspRS, and Glu-ProRS around a core of AIMP1, AIMP2, and AIMP3, also known as p43, p38, and p18. The release of individual proteins from the complex triggers specific cell signaling activities (see TABLE) [125,127]. aaRS can also catalyze the synthesis of dinucleoside oligophosphates, and thus indirectly influence many other cell functions such as metal chelation and cell wall metabolism.

Evolutionary divergence from aa-synthetases has provided additional functions for “paralogs”, which resemble the aa-synthetases in sequence and structure, and sometimes also in function. These paralog functions have been reviewed by Pang, et al. [127] and Schimmel, et al., [128]. “For instance, the *E. coli* YadB protein resembles GluRS and attaches glutamate to a queuosine base (position 34) at the first anticodon position of tRNA^{Asp} [129]. This modification expands and alters the tRNA decoding of codons [129,130]. HisZ, a paralog of HisRS, is one component of the enzyme catalyzing the first step of histidine biosynthesis (Sissler, et al., 1999” [131]. Like some of the tRNA synthetases, paralogs also carry out cyclodipeptide synthase activities [132,133].

Paralogs also play critical roles in translation by replacing missing amino acid tRNA synthetases. It is a common misconception is that all living organisms have a complete complement of 20 aaRS corresponding to each of the essential amino acids. In fact, 59% of organisms have less than the required 20, and 22% have more than the required 20, leaving only 19% with the exact complement that is often stated in textbooks. In organisms with both too few and too many aaRS, function is supplied by paralogs that are often able to modify one amino acid into another [134]. Many of these paralogs have taken on additional, non-ribosomal functions that are summarized in Table 7, which include amino acid synthesis, metabolic synthesis, polymerase activity, and stress and immunological functions [128,135–138]. Notably, the human 50S ribosomal protein L2 and *Mycobacterium* lysyl-tRNA synthetase are very similar [139], suggesting that ribosomal proteins and tRNA synthetases may have common origins—another observation that is consistent with a ribosome-first origin of cellular function, but which could also emerge from a protein-first theory as well.

Table 7. Non-canonical functions of tRNA synthetases and their paralogs. Table based on [117–139].

Non-Canonical Properties of Aminoacyl tRNA Synthetases (aa-RS)	Molecules Involved	Functions
Metabolism		
Lysyl-tRNA synthetase	Nucleosides	Polyphosphate synthesis
Regulation		
Alanyl-tRNA synthetase	Ala-tRNA synthetase gene operon	Autogenous regulation of transcription
Asparaginyl-tRNA synthetase	FGF-2 induced osteoblast growth	Regulates anti-apoptotic APK/Atk signalling
Leucyl-tRNA synthetase	Group I introns, mRNA	Splices group 1 introns
Threonyl-tRNA synthetase	Thr-tRNA synthetase mRNA	Autogenous regulation of translation
Tyrosyl-tRNA synthetase	Group I introns, mRNA	Splices group 1 introns

Table 7. Cont.

Non-Canonical Properties of Aminoacyl tRNA Synthetases (aa-RS)	Molecules Involved	Functions
Signaling		
Asparaginyl-tRNA synthetase	CXCR1, CXCR2, ERK1, etc.	Activates CCR5 & IL-8 receptors CXCR1 and CXCR2; MAP kinases
Glycyl-tRNA synthetase	Cadherin (CDH6)	Binding to CDH6 (cadherin) releases suppressed phosphatase 2A (PP2A) that dephosphorylates activated ERK
Histidyl-tRNA synthetase	CCR5	Chemokine: CCR5 activation
Tryptophyl-tRNA synthetase	DNA-PKcs, PARP-1	Bridges DNA-PKcs to PARP-1 to link IFN- γ and p53 signaling
Release of following from MM-aaRS complex:	Macromolecular aa-RS Complex (MM-aaRS):	Results in the following activity:
LysRS-II	Complex of ArgRS, GlnRS, IleRS, LeuRS, LysRS, MetRS, AspRS, Glu-ProRS, AIMP1, AIMP2, AIMP3	Inflammatory cytokine (Extracellular) Viral assembly (Plasma membrane) Transcriptional control (Nuclear)
Glu-ProRS	Ditto above	Translational silencing, e.g., VEGF
GlnRS	Ditto above	Anti-apoptosis
MetRS	Ditto above	rRNA transcription
IleRS	Ditto above	Autoimmune response
AIMP1 (aminoacyl-tRNA synthetase interacting multifunctional protein 1) (p43)	Ditto above	Cytokine activation of macrophages
AIMP2 (p38)	Ditto above	Degradation of FBP (transcriptional activator of <i>c-myc</i>)
AIMP3 (p18)	Ditto above	p53 induction and DNA repair by activating nuclear ATM/ATR
Fragments		
Lysyl-tRNA synthetase	Proteolytic internal fragment of Lys-RS	Cytokine-like domain that functions in chemotaxis
Tryptophyl-RS	Trp-RS C-terminal	Up-regulates TNF-alpha
Tryptophyl-RS	Trp-RS N-terminal	Mimics interleukin-8 functions
Tyrosyl-RS	Tyr-RS C-terminal	Up-regulates TNF-alpha
Tyrosyl-RS	Tyr-RS N-terminal	Mimics interleukin-8 functions
Paralogs		
AlaX, paralog of Ala-RS	tRNA	Trans-editing of mischarged tRNA
AlbC, paralog of TyrRS & TrpRS	Phenylalanine, Serine	Cyclodipeptide synthase
AsnA, paralog of Asn-RS	Asparagine	Asparagine biosynthesis
Arc1p, paralog of Met-RS	mRNA	RNA editing
BirA, paralog of Ser-RS	Biotin	Biotin repressor
DNA pol γ , paralog of Gly-RS	Polynucleotides	DNA polymerase
EMAP II, paralog of Tyr-RS	Cytokine receptors	Cytokine
GCN2, paralog of His-RS	Proteins, Histidine	Protein kinase, Histidine biosynthesis
HlsZ, a paralog of His-RS	Histidine	Histidine biosynthesis
LysK, paralog of LysRS	Lysine	Lysine biosynthesis
PheX, paralog of Phe-RS	?	Undetermined
Poly β , paralog of GlyRS	Polynucleotides	DNA polymerase
PoxA, paralog of LysRS	Elongation factor P	Modifies EF-P
PoxA/GenX, paralog of Lys-RS	Pyruvate	Pyruvate oxidase
ProX, paralog of Pro-RS	tRNA	Trans-editing of mischarged tRNA
Slimp, paralog of mito SerRS	RNA binding	Regulates rRNA, COX2 & 3 mRNA
Trbp111, paralog of Met-RS	mRNA	RNA editing
YadB, or GluX, a paralog of Glu-RS	Adds glutamate to first anticodon position of tRNA-Asp	Stress response
Ybak, paralog of ProRS	Cys-tRNA(Pro)/Cys-tRNA(Cys) deacylase	Trans-editing of mischarged tRNA
YvmC-Blic, class Ic aaRS paralog	Cyclodileucine	Non-ribosomal peptide synthesis

2.6. Non-Canonical Functions of RNA Polymerases

Finally, it is worth noting that the ribosome-first theory of cellular evolution requires that the rRNA be able to replicate, and therefore, that products of rRNA genes must include RNA polymerases. We have demonstrated that key active site sequences of RNA polymerases do appear in various rRNA reading frames in a previous paper [6]. In keeping with the predictions made above that ribosome-related proteins should play extra-ribosomal roles in the formation of the first genomes, it is important to remember that RNA- and DNA-polymerases evolved from common ancestral enzymes, containing a single highly conserved peptide at their catalytic core and which seems to be able to polymerize both types of polynucleotides [140–142]. Therefore, a self-replicating ribosome encoding an RNA-polymerase catalytic core could have given rise to the first DNA polymerases, thereby facilitating the transition to a DNA-encoded genome. Even beyond that necessary evolutionary step, however, RNA polymerases have been found to play at least two additional roles in modern cells. One is gene silencing and transcription regulation [143,144], and the other is in myeloproliferation and hematopoiesis [145]. Given the wide range of functions taken on by other ribosome-related molecules, we predict that the polymerases will also be found to have additional functions.

2.7. Theory Testing: How Well Do These Data Fit Other Possible Explanations of Genome Evolution?

Obviously, the data summarized above argues strongly for rRNA-containing genetic information that is expressed as translated proteins that are unrelated to ribosome function, and that rRNA-like and tRNA-like genetic sequences occur far more frequently than is either necessary or expected if rRNA and tRNA contain only structural, but no genetic, information. These findings raise the question of how much more frequent is such rRNA-like genetic information than one would expect. Here we attempt to address this question by comparing the frequency with which rRNA-like sequences appear in the *E. coli* K12 genome, as compared with the frequency that one would expect on the basis of a set of randomly generated rRNA-like sequences, and on the basis of a set of protein-encoding mRNAs of equivalent lengths.

The key data for testing these three possibilities are Figures 1 and 2, showing the distributions of rRNA-encoded proteins from all reading frames that are present in the *E. coli* K12 genome. These figures show that, on average, the *E. coli* K12 genome contains hundreds of copies of some version (translated in the six possible reading frames of the RNAs) of each of the 5S, 16S and 23S rRNAs. This observation is confirmed in the Tables below as well, and the data are consistent with the findings of Mauro and Edelman (1999) and many other studies of the ubiquity of rRNA-like sequences in the expressed proteins reviewed above. Four notable features of the locations of the rRNA homologies emerge from these plots. One is that the homology locations differ quite obviously between the 5S, 16S, and 23S homologies, indicating that they probably have independent origins (Figure 1). Secondly, the locations also differ from one reading frame to another (Figure 2), demonstrating that the homology search program is not merely picking up different versions of the same sequence. Third, the locations of the homologies differ even between closely related strains of *E. coli* K12 (Figures 1 and 2), indicating that they are subject to rearrangement processes. Finally, the homology locations, especially when averaged over all six possible reading frames, are more or less randomly distributed across the entire *E. coli* genome so that any region of 10,000 or so base pairs is very likely to have several sequences homologous to at least one of the six reading frames of the 5S, 16S, or 23S rRNAs.

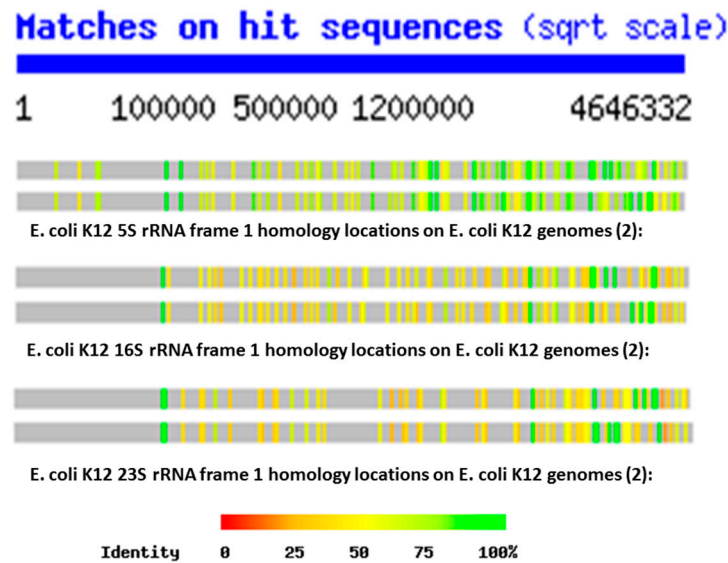


Figure 1. Locations of 5S rRNA, 16S rRNA, and 23S rRNA frame 1 homologies within two strains of *E. coli* K12 genomes. Note that the scale is non-linear. Homology locations are not identical between the 5S, 16S, and 23S rRNAs, demonstrating that the sequences probably have independent origins. Similarly, the homology locations are not identical even between two closely related *E. coli* K12 strains, suggesting that these sequences are subject to reorganization processes within these genomes.

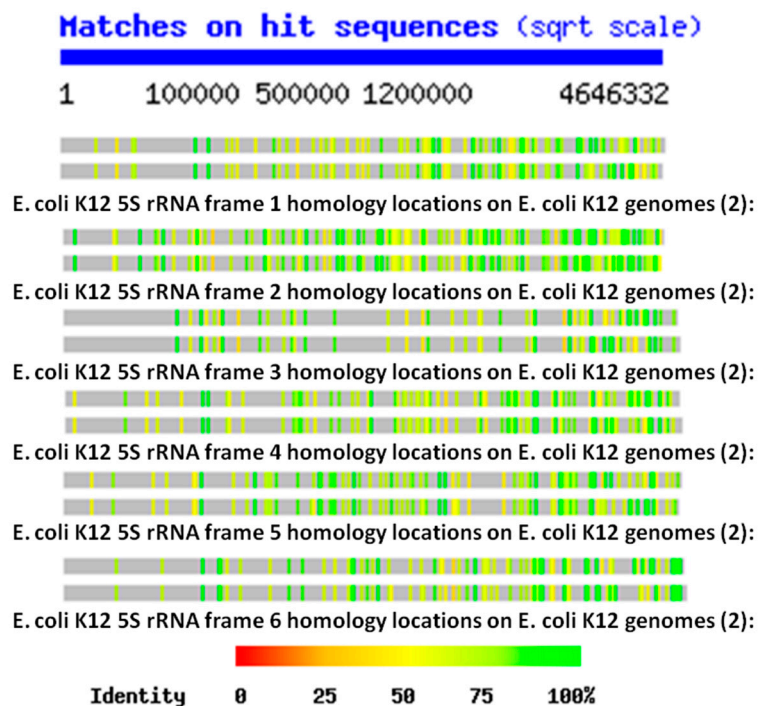


Figure 2. Locations of the six frames of 5S rRNA homologies within two strains of *E. coli* K12. Note that the scale is non-linear. As in the previous Figure, these homology locations differ between the two strains indicating that they are subject to reorganization processes. Notably, the locations also differ considerably from one reading frame to another, demonstrating that the search algorithm is identifying non-overlapping sequences.

The data resulting from a comparison of the frequency and qualities of rRNA in the *E. coli* genome confirm that ribosomal RNA-like sequences appear in the genome at far higher frequencies with much longer and more similar matches than do matches to randomized RNA sequences of equal length, or

matches from *E. coli* protein mRNA of equal length or randomized RNA sequences. Tables 8–10 and Figure 3 summarize the data comparing the frequency and quality of 5S rRNA homologies within the *E. coli* K12 genome with data from a randomly selected set of protein mRNA of equivalent length and a set of randomly generated RNA sequences. Tables 11–13 and Figure 4 summarize the data comparing the frequency and quality of 16S rRNA homologies within the *E. coli* K12 genome with data from a randomly selected set of protein mRNA of equivalent length, and a set of randomly generated RNA sequences. Tables 14–16 and Figure 5 summarize the data comparing the frequency and quality of 23S rRNA homologies within the *E. coli* K12 genome with data from a randomly selected set of protein mRNA of equivalent length and a set of randomly generated RNA sequences.

Table 8. 5S rRNA protein control sequences utilized to test the frequency and quality of homologous sequences in the *E. coli* K12 genome. Sample numbers identify the mRNA of control proteins of the *E. coli* K12 genome accessed from the U. C. Santa Cruz microbial genome list: http://microbes.ucsc.edu/lists/eschColi_K12/refSeq-list.html. See text for definitions of “identities”, “similarities”, etc.

Protein 5S Control	Count Data Total, Sum of Six Reading Frames		Sums across All Matches			Means across All Matches		
	Sample	N°. Matches	N°. Perfect Identities	Total Length of Matches	Total Number of Similarities	Total Number of Identities	Mean Match Length	Mean Similarity Length
b4626	83	2	632	568	550	7.6	7	6.6
b0466	241	1	3325	2326	1886	13.8	9.6	7.8
b4631	83	1	948	791	672	11.4	9.5	8.1
b4509	127	3	1545	1234	1101	12.1	9.7	8.7
b4515	127	1	1428	1150	990	11.2	9.1	7.8
mean	132.2	1.6	1575.6	1213.8	1039.8	11.2	8.9	7.8
SD	64.7	0.9	1044.9	677.7	523.8	2.3	1.1	0.8

Table 9. 5S rRNA random control sequences utilized to test the frequency and quality of similar sequences in the *E. coli* K12 genome. See text for definitions of “identities”, “similarities”, etc.

Random Sequence 5S Control	Count Data Total, Sum 6 Reading Frames		Sums across All Matches			Means across All Matches		
	Sample	N°. Matches	N°. Perfect Identities	Total Length of Matches	Total Number of Similarities	Total Number of Identities	Mean Match Length	Mean Similarity Length
120 bp 1	130	0	1882	1251	959	14.5	9.7	7.3
120 bp 2	141	0	1432	1107	926	10.3	7.9	6.6
120 bp 3	151	0	1715	1287	1040	11.4	8.5	6.9
mean	140.7	0	1676.3	1215	975	12.1	8.7	6.9
SD	10.5	0	227.5	95.2	58.7	2.2	0.9	0.4

Table 10. Frequency and quality of 5S rRNA homologies in the *E. coli* K12 genome. See text for definitions of “identity”, “similarity”, etc.

5S rRNA	Count Data Total, Sum 6 Reading Frames		Sums across All Matches			Means across All Matches		
	Sample	N°. Matches	N°. Perfect Identities	Total Length of Matches	Total Number of Similarities	Total Number of Identities	Mean Match Length	Mean Similarity Length
5S rRNA	480	158	5696	4580	4136	11.5	9.5	8.6

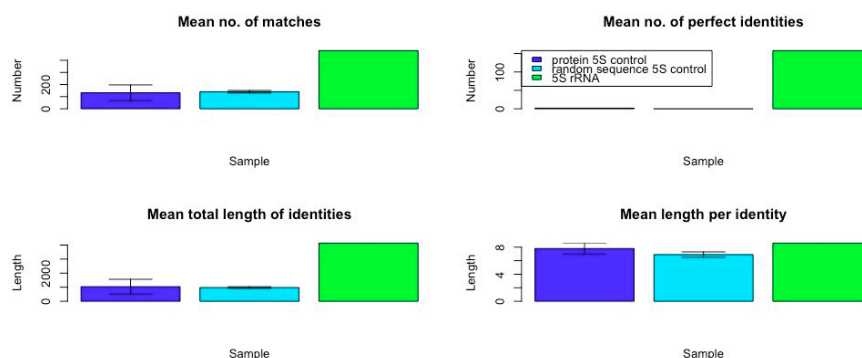


Figure 3. Comparison of 5S rRNA variables with matched protein mRNA and randomized RNA controls (original data in Tables above). Error bars show SD. While the mean length of the matches found in each group was very similar, the number of total matches, the number of identical amino acids found within those matches, and the number of exact matches found to the 5S rRNA was very significantly greater than those found for the proteins’ control mRNA and randomized RNA sequences tested.

Table 11. 16S rRNA protein control sequences utilized to test the frequency and quality of homologous sequences in the *E. coli* K12 genome. Sample numbers identify mRNA of control proteins of the *E. coli* K12 genome accessed from the U. C. Santa Cruz microbial genome list: http://microbes.ucsc.edu/lists/eschColi_K12/refSeq-list.html. See text for definitions of “identities”, “similarities”, etc.

Protein 16S Control	Count Data, across All Six Reading Frames		Sums across All Matches			Means across All Matches			
	Sample	N°. Matches	N°. Perfect Identities	Total Length of Matches	Total Number of Similarities	Total Number of Identities	Mean Match Length	Mean Similarity Length	Mean Identity Length
b0135		389	1	10,999	7041	5647	23.8	18.1	14.5
b0070		414	1	9725	3923	5378	27.2	14.8	10.8
b0376		625	1	15,008	8069	6029	24.0	12.9	9.6
b0014		289	2	10,260	7043	5852	35.5	24.4	20.2
b0091		247	1	7924	5385	4649	32.1	21.8	18.8
mean		392.8	0.8	10,783.2	6292.2	5511	28.5	18.4	14.8
SD		147.0	0.8	2620.3	1636.9	539.4	5.2	4.8	4.7

Table 12. 16S rRNA random control sequences utilized to test the frequency and quality of similar sequences in the *E. coli* K12 genome. See text for definitions of “identities”, “similarities”, etc.

Random Sequence6s Control	Count Data, across All Six Reading Frames		Sums across All Matches			Means across All Matches			
	Sample	N°. Matches	N°. Perfect Identities	Total Length of Matches	Total Number of Similarities	Total Number of Identities	Mean Match Length	Mean Similarity Length	Mean Identity Length
1500 bp 1		129	0	2668	1658	1347	20.7	12.9	10.4
1500 bp 2		61	0	1816	998	704	29.8	16.4	11.5
1500 bp 3		70	0	2100	1124	910	30	16.1	13
1500 bp 4		82	0	2273	1269	889	27.7	10.8	14.5
1500 bp 5		68	0	2090	1113	792	30.7	16.4	11.7
mean		82	0	2189.4	1232.4	928.4	27.8	15.4	11.5
SD		27.3	0	313.6	256.6	248	4.1	1.5	1.0

Table 13. Frequency and quality of 16S rRNA homologies in the *E. coli* K12 genome. See text for definitions of “identity”, “similarity”, etc.

16S rRNA	Count Data, across All Reading Frames		Sums across All Matches			Means across All Matches		
Sample	N°. Matches	N°. Perfect Identities	Total Length of Matches	Total Number of Similarities	Total Number of Identities	Mean Match Length	Mean Similarity Length	Mean Identity Length
16S rRNA	651	49	36,488	28,126	25,848	56.1	43.2	39.7

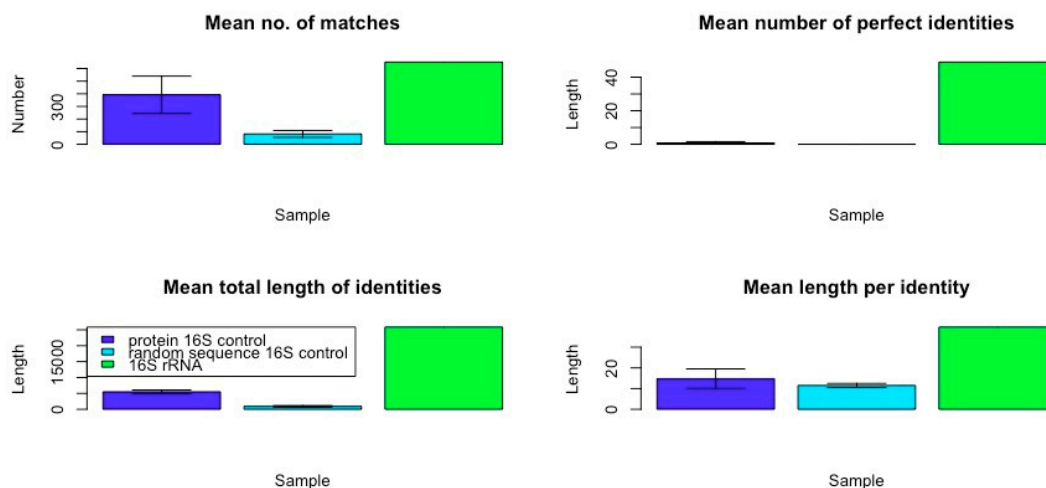


Figure 4. Comparison of 16S rRNA variables with matched protein mRNA and randomized RNA controls (original data in Tables above). Error bars show SD. The mean number of matches found within the *E. coli* genome was greatest for the 16S rRNA, as was the mean number of perfect identities, length of the sequences with identical sequences and the number of identical amino acid matches found within each match was significantly greater for the 23S rRNA than for the mRNA of the proteins controls or the randomized RNA sequences.

Table 14. 23S rRNA protein control sequences utilized to test the frequency and quality of homologous sequences in the *E. coli* K12 genome. Sample numbers identify mRNA of control proteins of the *E. coli* K12 genome accessed from the U. C. Santa Cruz microbial genome list: http://microbes.ucsc.edu/lists/eschColi_K12/refSeq-list.html. See text for definitions of “identities”, “similarities”, etc.

Protein 23S Control	Count Data, across All Six Reading Frames		Sums across All Matches			Means across All Matches		
Sample	N°. Matches	N°. Perfect Identities	Total Length of Matches	Total Number of Similarities	Total Number of Identities	Mean Match Length	Mean Similarity Length	Mean Identity Length
b0397	374	1	12,537	8947	7894	39	24	21.1
b0993	404	1	14,283	9795	8257	35.3	24.2	20.4
mean	389	1	13,405	9371	8075.5	37.2	24.1	20.8
SD	21.1	0	1234.6	599.6	256.7	2.6	0.1	0.5

Table 15. 23S rRNA random control sequences utilized to test the frequency and quality of similar sequences in the *E. coli* K12 genome. See text for definitions of “identities”, “similarities”, etc.

Random Sequence/23S Control	Count Data, across All Six Reading Frames		Sums across All Matches			Means across All Matches		
Sample	N°. Matches	N°. Perfect Identities	Total Length of Matches	Total Number of Similarities	Total Number of Identities	Mean Match Length	Mean Similarity Length	Mean Identity Length
2900 bp 1	130	0	1882	1251	959	14.5	9.7	7.4
2900 bp 2	141	0	1432	1107	926	10.3	7.9	6.6
2900 bp 3	151	0	1715	1287	1040	11.3	8.5	6.9
2900 bp 4	117	0	1192	934	761	10.2	8.0	6.5
2900 bp 5	120	0	1235	958	805	10.3	8.0	6.7
mean	131.8	0	1491.2	1107.4	898.2	11.3	8.4	6.8
SD	14.3	0	300.6	162.2	114.1	1.8	0.7	0.4

Table 16. Frequency and quality of 23S rRNA homologies in the *E. coli* K12 genome. See text for definitions of “identity”, “similarity”, etc.

23S rRNA	Count Data, across All Reading Frames		Sums across All Matches			Means across All Matches		
Sample	N°. Matches	N°. Perfect Identities	Total Length of Matches	Total Number of Similarities	Total Number of Identities	Mean Match Length	Mean Similarity Length	Mean Identity Length
23S rRNA	195	18	41,096	38,529	38,119	210.7	197.6	192.9

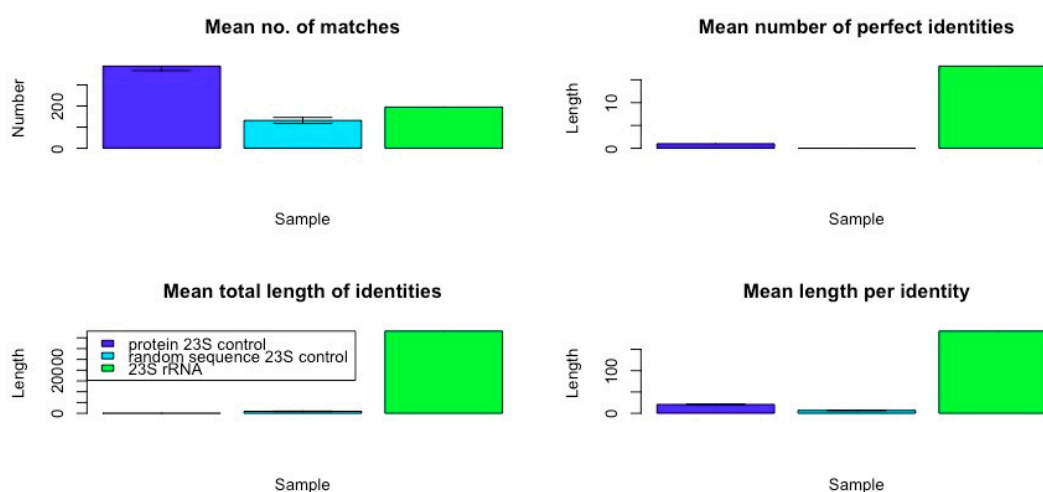


Figure 5. Comparison of 23S rRNA variables with matched protein mRNA and randomized RNA controls (original data in Tables above). Error bars show SD. Although the mean number of matches found within the *E. coli* genome was greatest for the protein controls, the mean number of perfect identities, length of the sequences with identical sequences and the number of identical amino acid matches found within each match was significantly greater for the 23S rRNA than for the mRNA of the protein controls or the randomized RNA sequences.

In general (with the exception of the 23S protein matches) the number of rRNA homologies was significantly greater than the number of mRNA and random RNA matches. The number of matches differed by sample type (protein control, random control, or rRNA) for the 5S rRNA (ANOVA, $F_{2,7} = 18.7, p = 0.002$), for the 16S (ANOVA, $F_{2,9} = 17.76, p = 0.001$), and for the 23S (ANOVA, $F_{2,6} = 187, p = 2 \times 10^{-5}$). However, for the 23S samples, it was the protein control that had the greatest number of matches (see Figure 5). The random RNA sequences had no identical matches with the *E. coli* genome,

where “identity” is defined as at least 90% identical amino acids over a sequence comprising at least 90% of the length of the RNA used as the search string. Thus, a 120 base pair (5S rRNA) would have to match a continuous sequence of 108 base pairs, with at least 90% identical amino acids in the *E. coli* genome, to satisfy the “identity” classification. Each protein-derived mRNA sequence had one (and occasionally two) identities with the genome, as would be expected from the fact that these proteins are encoded in the genome. The rRNA sequences, in contrast, had dozens of identical matches scattered throughout the genome, as shown in Figures 1 and 2. These data argue against a random or protein origin of the genome while supporting an rRNA origin.

Additional evidence for an rRNA origin was found in the quality measures derived from the homology searches. With the exception of the 5S rRNA matches, which were no longer on average than those derived from mRNAs or random RNA sequences, rRNA matches were significantly longer than the controls (Figures 3–5). Mean identity length did not differ across sample type (protein control, random control, or rRNA) for the 5S rRNA (ANOVA, $F_{2,7} = 2.938$, $p = 0.129$). Mean identity length differed significantly by sample type for 16S rRNA (ANOVA, $F_{2,9} = 26.09$, $p = 0.0003$), and for 23S (ANOVA, $F_{2,6} = 97,253$, $p = 3.35 \times 10^{-12}$).

Moreover, the number of identical amino acids found within these longer matches was also significantly greater, so that these longer matches were also better matches (Figures 3–5). The number of perfect identities differed by sample type (protein control, random control, or rRNA) for the 5S rRNA (ANOVA, $F_{2,7} = 20,545$, $p = 3.11 \times 10^{-12}$), 16S (ANOVA, $F_{2,9} = 3070$, $p = 5.4 \times 10^{-11}$), and 23S (ANOVA, $F_{2,6} = 1.5 \times 10^{33}$, $p < 2 \times 10^{-16}$). The data derived by looking at amino acid similarities rather than identities was virtually identical from a statistical perspective; adding the similar amino acids within each sequence to the identities merely increased the quality of the matches without further differentiating the rRNAs from the controls, and so these data have not been displayed.

The average length of the rRNA matches is seven times that of the random RNA matches, and the number of identical amino acids within those matches is 14 times as many as in the random RNA matches. The mean length of identities in matches differed by sample type (protein control, random control, or rRNA) for the 5S rRNA (ANOVA, $F_{2,7} = 23.54$, $p = 0.001$), for 16S, (ANOVA, $F_{2,9} = 1470$, $p = 5.43 \times 10^{-11}$), and for 23S (ANOVA, $F_{2,6} = 24,496$, $p = 1.05 \times 10^{-10}$). Altogether, the rRNA matches account for 11 (16S rRNA) to 19 (23S rRNA) times the amount of the *E. coli* genome that is recognized as similar, as do the random RNA matches or the protein-derived mRNA matches. rRNA sequences are clearly highly over-represented compared with either control group.

In sum, rRNA sequences are significantly more likely to match the genome than are sets of mRNA derived from *E. coli* proteins of similar lengths. The average length of an rRNA match is six times that of an mRNA match (for 16S- and 23S-length sequences) and, as with the random sequences, the rRNA matches are significantly more likely than the mRNA matches to encode more amino acid identities so that overall the rRNAs encode four to six times as many identical amino acids within their matches (and this, in spite of the fact that BLAST yielded three times as many mRNA matches than 23S rRNA matches). Most significantly, while a typical mRNA sequence had 95% or more identity over its entire length (an “identity”) with only one genome sequence (as would be expected for matching its gene), the rRNA sequences yielded dozens of exact homologies. Thus, while protein sequences generally occur in only one copy in one reading frame, rRNA sequences occur in many copies in multiple reading frames. No random sequence in our study yielded any exact homologies. Protein mRNA matches are more similar to random sequence matches than to rRNA matches.

Overall, this statics-based similarity investigation into the frequency and qualities of rRNA-like, mRNA-like and randomized RNA-like sequences in the *E. coli* genome confirms that rRNA-like sequences occur very significantly more often than one would expect by chance, and far more often than mRNA duplications do. Indeed, it is possible to calculate the total number of base pairs involved in the similarities revealed in the Tables above, which yields the surprising result that, although the 5S, 16S, and 23S rRNAs, which total 4566 bp and account for only 0.098% of the 4,646,332 bp of the *E. coli* K12 genome, similar sequences (at $E = 10$ in the BLAST parameters) are present in sufficient

copies (in all six reading frames) to account for 1/20 or 5% of the genome (240,171 bp) with quite high fidelity. This compares with 1/250 or 0.4% of the genome matching randomized RNA sequences (18,702 bp) and about 1/70 or 1.3% matching random mRNA (protein-encoding) sequences (66,768 bp), but with far less fidelity than for the rRNA sequences in terms of percent identities within each match. Thus, our hypothesis that rRNA may have been the basis for the first cellular genomes is supported while alternative theories such as a protein-world origin for the genome, or genome-before-ribosome, are not.

3. Discussion

To summarize, we have reviewed existing studies that demonstrate evidence that ribosome-related RNAs are present in unexpectedly large quantities in all genomes, and that these rRNAs, tRNAs and ribosome-related proteins play a wide range of extra-ribosomal, pleiofunctional roles in modern cellular systems. Large numbers of rRNA-like sequences in both prokaryotes and eukaryotes are transcribed as polyadenylated mRNAs and, in many cases, translated into functional proteins. rRNA- and tRNA-like transposons are known to make up a significant proportion of all genomes, which may provide a mechanism by which ribosome-related genes have proliferated within cellular genomes. rRNAs and tRNAs and their fragments play multiple extra-ribosomal roles as regulators of transcription and translation, as metal chelators, by carrying out various forms of non-ribosomal protein synthesis, as mRNAs, and in cell signaling functions. Nearly all ribosomal proteins are known to perform at least one, and often multiple, extra-ribosomal functions ranging from autologous regulation of their own mRNAs by means of binding to them, to control of replication, transcription and translation, participation in polyphosphate synthesis and metabolism, and carrying out various enzyme functions. Aminoacyl tRNA synthetases similarly play a wide range of extra-ribosomal functions that include replication, transcription, and translation regulation, polyphosphate metabolism, dipeptide synthesis, and cell signaling. These synthetases have given rise to a large number of paralogs that carry out additional enzymatic functions including amino acid synthesis, RNA editing, and energy metabolism (Figure 6). Thus, ribosome-related molecules are capable of carrying out most key cellular functions as predicted on the basis of the ribosome-first theory of cellular evolution. In addition, we have demonstrated here for the first time that although rRNA-encoding genes make only a few hundredths of a percent of the *E. coli* genome, rRNA-like sequences constitute no less than 1/20th or 5% of the *E. coli* genome. This prevalence is many times greater than would be predicted if the genome evolved from random sequences of RNA or from divergence from protein-encoding genes. Note further that our analysis did not investigate the occurrence of tRNA-like sequences, SINES or LINES, which the literature cited above suggests may account as much as 10–20% of many genomes. In short, our analysis suggests why ribosome-related genes are the core of the approximately 100 orthologous genes conserved across all species that make up the Universal Gene Set of Life (UGSL) [1–3].

Consider these data in light of the three possible scenarios outlined in our Introduction. A genome-first approach does not fit the evidence for multiple ribosomal molecule function at all. In a genome-first approach, translation is added to cellular functions after an RNA-world has already evolved autonomously, without feedback from protein–world interactions. rRNA and tRNA genes would have been selected without regard for protein function and so have no predictive value for other cellular functions. Additionally, if ribosomes evolved separately from an RNA- or DNA-based genetic world, then one would expect to find, as current dogma maintains, that ribosomal genes have no functions other than to encode rRNA and that rRNA contains no genetic information. These predictions are not borne out by the evidence gathered here.

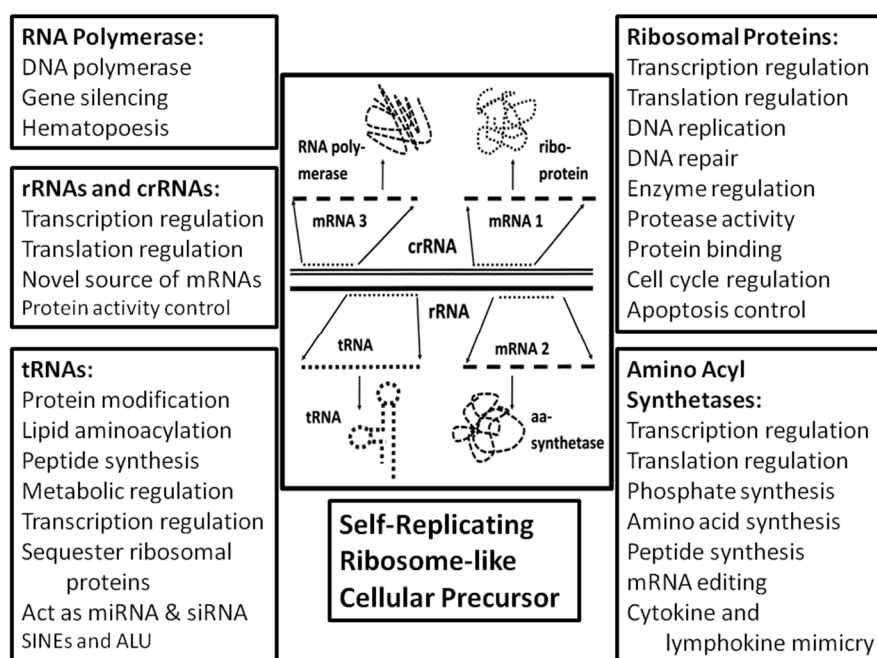


Figure 6. A summary of the evidence for pleiofunctionality in rRNA-related molecules supporting the hypothesis of a self-replicating ribosome. rRNA = ribosomal ribonucleic acid (only one rRNA is shown though there are three in the ribosome, the 5S, 16S, and 23S). crRNA = complementary strand of rRNA posited to exist to make possible rRNA replication in self-replicating ribosome. The crRNA would also act as a messenger RNA (mRNA), or act itself as a source of transcription to produce mRNAs (see Root-Bernstein and Root-Bernstein, 2016; 2017). tRNA = transfer ribonucleic acid. aa-synthetase = amino acyl synthetase, which loads amino acids onto tRNA. Ribo-protein = ribosomal proteins (see Tables 3–6). Boxes summarize the data from Tables 1–7 and the related text.

A metabolism-first approach to cellular evolution runs into a different conundrum: the possibility that rRNAs, tRNAs, and ribosomal proteins all evolved separately for non-translational functions, and then reached a critical point, at which they suddenly integrated to perform translation functions seems not only miraculous in its implausibility, but leads to the inherent contradiction that cellular life evolved without the means to reliably produce the proteins required to make it possible. Thus, while there is every reason to predict that ribosomal proteins would have non-ribosomal functions in a metabolism-first scenario, the difficulties in evolving a ribosome in the first place produce significant barriers to a metabolism-before-genes model, and our evidence suggests that a protein-first model of genome evolution is only marginally better than a totally random origin of genome sequences.

The ribosome-first theory is the only current theory that is consistent with all data gathered here and the only one that explains the co-evolution of genetic information and translation function from the outset. The ribosome first theory is consistent with the prebiotic evolution of an RNA–protein world [12,146–150], thereby integrating observations that both RNA species (“RNA world” scenarios [11,150–156]) and amino acids (“protein world” scenarios [157,158]) were present simultaneously from the outset. More importantly, the ribosome-first theory is consistent with multiple taxonomic studies that all place the origin of functional ribosomes earlier than the last universal common ancestor (LUCA) and therefore, with high probability, prior to the inception of cellular life itself [159–163]. If the cell evolved, at least in part, to optimize preexisting ribosomal functions by providing a protected, homeostatic environment with a ready and steady supply of precursors and energy, then the incorporation of rRNA sequences into the very fabric of the cellular genome in multiple copies and multiple reading frames becomes logically comprehensible. Ribosomal RNA would already function like a small genome encoding the ribosome and its attendant tRNAs and proteins. Having already optimized the information content of their “genomes” in order to exist as “free-living” selfish

agents, ribosome-like entities would have been among the primary donors of genetic material to emergent cellular life. Such cellular life would have vastly increased the metabolic efficiency, stability, and probability of successful replication of such ribosome-like entities by accumulating necessary resources for the ribosome and providing an optimized environment for its function. As part of this process, the cellular chromosome would have “unpacked” the highly integrated information within the rRNAs (in which every reading frame encodes functional protein segments as well as serving as mRNAs and tRNAs) and transferred these separate functions to the cellular genome. As separate genes, this information would be more readily accessible for transcription and translation than when “packed” as overlapping sequences; it would be more amenable to control through the incorporation of promoter and repressor sequences; and it could be stored in the much more stable form of DNA genes. The evolution of DNA-encoded genes may have provided a high degree of stability to a previously highly varying ecology of self-replicating ribosome-like entities, and selection for stabilization of the translation function of these ribosome-like entities at the expense of their evolvability. Evolutionary processes would have replicated and produced divergences in the DNA-encoded genetic sequences and recombined these sequences with genetic material from other sources to yield molecular novelties so that the various reading frames of the rRNAs would have found novel uses throughout the new cellular genome (Figure 7). By building up the cellular genome from already tried-and-true interactive modules, the integrity of the cellular interactome would have been assured as cell functions expanded and diversified. The present-day ribosome would then have been able to divest some of its functions by off-loading them to the cellular genome in order to optimize translation efficiency.

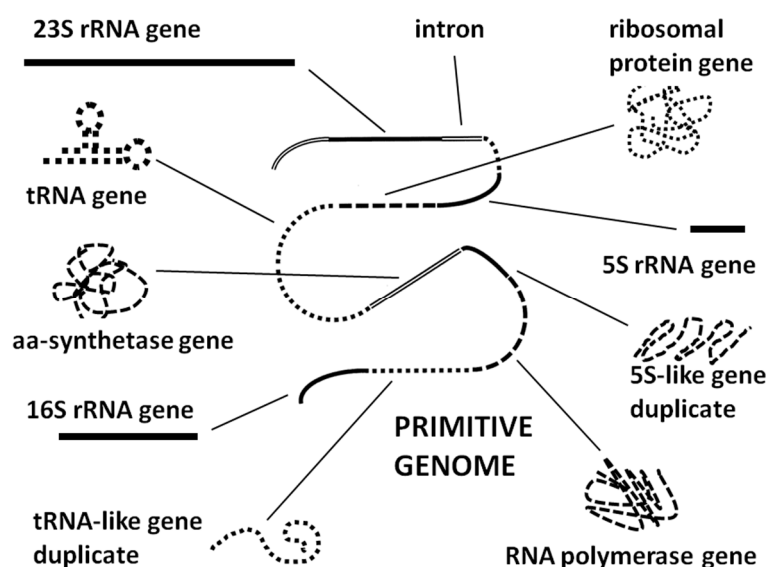


Figure 7. Schematic model of how RNAs perform various ribosomal functions and/or encoding various ribosome-related proteins may have become ligated to form the first genome. The visual formalisms are the same as in Figure 6 and illustrate the idea that “genes” that may have existed as overlapping sequences in a variety of reading frames within a self-replicating set or rRNAs may have been separated out into a single reading frame for better functional control. This schema does not identify whether the original primitive genome was constituted of RNA or DNA (though RNA seems likely if a ribosomal origin is correct) but given a common origin for RNA and DNA polymerases (see text) the existence of either one could have provided the basis for the other. It is also important to remember that SINEs and other insertion-facilitating RNAs most likely evolved from tRNAs or that they have common ancestors (see text), so that molecular mechanisms of genome growth would have been present from the outset in a ribosome-first scenario.

This is not to say that the ribosome-first theory of cellular origins explains all aspects of cellular function or genome structure. It is evident from the list of functions that are associated with ribosomal

molecules that these do not include critical aspects such as glucose metabolism (e.g., the origins of the Krebs and Calvin cycles), nor does it explain the origins of other key cellular organelles such as acidocalcisomes and cell membranes that are found in all cells. The origins of these structures should also attract serious investigation and mechanisms by which cells evolved to integrate them with ribosomes and genomes need to be explored.

We suggest in conclusion that our modular approach to evaluating the evolution of such bacterial genomes reveals a fundamental mechanism by which complexity and diversity resulted from mixing and matching simple, already-proven, functionally interactive, genetically encoded protein fragments. By using such pre-selected modules as building blocks for more complex genes and proteins, the number of different permutations required to find new functional elements was minimized, the probability of retaining functionality greatly increased, and the interactivity of these elements maximized. The specific results of our analysis can therefore be seen as a concrete example of an evolutionary self-assembly principle proposed by Simon [164], who demonstrated theoretically that complex systems are much more likely to evolve from interactive, semi-stable modules than from the random assortment of components that evolve independently of each other (see also [165,166]). Our approach also suggests new avenues for research on the origins of cellular life. Looking backwards to the evolution of self-replicating ribosome-like entities, the transposon-like functions of so many tRNA- and rRNA-like genes suggests that ribosomes themselves may have evolved from a distributed ecology of virus-like particles capable of rapid evolution through gene swapping and rearrangement. The ribosome would have been a critical step in integrating this distributed, co-evolving set of molecules into a stable, efficient replication-transcription-translation system.

4. Materials and Methods

Escherichia coli K12 rRNA (5S, 16S, and 23S) sequences were obtained from http://ecoliwiki.net/colipedia/index.php/16S_rRNA:Gene_Product%28s%29. Each RNA was translated into its six protein translation reading frames using a translate tool (<https://web.expasy.org/translate/>). The protein sequences resulting from the six reading frames of the 5S, 16S, and 23S rRNA were compared with the *E. coli* K12 genome in the Nucleotide database using TBLASTN (version 2.2.31+), which searches protein sequences against the nucleotide sequence of the genome thereby permitting all possible reading frames to be examined. Parameters were set as follows: E = 10, the low complexity filter on, no gapped alignments, and 100 of the best scoring sequences and alignments to show. The resulting homologies were recorded for each frame in each rRNA sequence, in terms of the overall length of the protein match, how many identical amino acids were present in the homology, and how many similar amino acids were present. The same procedure was performed with a series of mRNA from control proteins of the *E. coli* K12 genome accessed from the U. C. Santa Cruz microbial genome list: http://microbes.ucsc.edu/lists/eschColi_K12/refSeq-list.html. The rRNA sequences and the proteins resulting from their six reading frames are provided in Supplementary Table V. Three sets of mRNA encoding *E. coli* proteins were explored for comparison, one set of five averaging about 120 bp in length to mimic the 5S rRNA; as second set of five averaging about 1500 bp in length to mimic the 16S rRNA, and a third set of two averaging about 2900 bp in length to mimic the 23S rRNA. The mRNA sequences encoding the proteins are provided in Supplementary Table W. The results for each set of five proteins were averaged for comparison with the *E. coli* K12 rRNA results obtained above. The proteins and their identifiers are listed in Table 1. The same procedure was performed one more time with three sets of further control RNA sequences generated using an RNA random sequence generator (<http://www.molbiotools.com/randomsequencegenerator.html>). Lengths were matched to the lengths of the 5S rRNA (120 bp), 16S rRNA (1500 bp) and 23S rRNA (2900 bp). The random sequences are provided in Supplementary Table X.

Ribosomal RNA-encoded protein location and copy number in *E. coli* K12 genome was determined using a function that has, unfortunately, since been removed from the program during a recent upgrade. Each reading frame for each rRNA (5S, 16S, and 23S rRNA) was compared using the nblast

function at the Swiss Institute of Bioinformatics website (<http://www.expasy.org>) function with two *E. coli* K12 genomes (“ECOLI”, 2 versions: U00096 *Escherichia coli* MG1655 Chromosome and AP009048 *Escherichia coli* ATCC 27325 Chromosome). The resulting data did not identify specific protein homologies but provided information about where the sequences appeared within the genome and how many copies of that particular rRNA-encoded protein were present in the genome. (The same kind of data can still be obtained, but must now be graphed onto the genome “by hand” using the bp location data provided by the search).

Statistics: ANOVA using the “aov” command in R version 3.3.3 (<https://cran.r-project.org/bin/macosx/>).

Author Contributions: R.R.-B. co-co-conceived the project, carried out the literature review, performed the proteomic searches and wrote the first draft of the manuscript; M.R.-B. co-conceived the project, carried out the statistical analyses, and revised the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Charlebois, R.L.; Doolittle, W.F. Computing prokaryotic gene ubiquity: Rescuing the core from extinction. *Genome Res.* **2004**, *14*, 2469–2477. [[CrossRef](#)] [[PubMed](#)]
2. Koonin, E.V. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* **2003**, *1*, 127–136. [[CrossRef](#)] [[PubMed](#)]
3. Harris, J.K.; Kelley, S.T.; Spiegelman, G.B.; Pace, N.R. The genetic core of the universal ancestor. *Genome Res.* **2003**, *13*, 407–412. [[CrossRef](#)] [[PubMed](#)]
4. Ortiz, J.O.; Förster, F.; Kürner, J.; Linaroudis, A.A.; Baumeister, W. Mapping 70S ribosomes in intact cells by cryoelectron tomography and pattern recognition. *J. Struct. Biol.* **2006**, *156*, 334–341. [[CrossRef](#)] [[PubMed](#)]
5. Scott, M.; Gunderson, C.W.; Mateescu, E.M.; Zhang, Z.; Hwa, T. Interdependence of cell growth and gene expression: Origins and consequences. *Science* **2010**, *330*, 1099–1102. [[CrossRef](#)]
6. Root-Bernstein, M.; Root-Bernstein, R. The ribosome as a missing link in the evolution of life. *J. Theor. Biol.* **2015**, *367*, 130–158. [[CrossRef](#)] [[PubMed](#)]
7. Root-Bernstein, R.S.; Root-Bernstein, M.M. The ribosome as a missing link in prebiotic evolution II: Ribosomes encode ribosomal proteins that bind to common regions of their own mRNAs and rRNAs. *J. Theor. Biol.* **2016**, *397*, 115–127. [[CrossRef](#)]
8. Root-Bernstein, R.S.; Root-Bernstein, M.M. From compositional chemical ecologies to self-replicating ribosomes and on to functional trait ecological networks. In *Evolutionary Biology—19th Meeting 2015*; Panterotti, P., Ed.; Springer: New York, NY, USA; London, UK, 2016; pp. 327–346.
9. Bloch, D.; McArthur, B.; Widdowson, R.; Spector, D.; Guimaraes, R.C.; Smith, J. tRNA-rRNA sequence homologies: A model for the origin of a common ancestral molecule, and prospects for its reconstruction. *Orig. Life* **1984**, *14*, 571–578. [[CrossRef](#)]
10. Bloch, D.P.; McArthur, B.; Guimarães, R.C.; Smith, J.; Staves, M.P. tRNA-rRNA sequence matches from inter- and intraspecies comparisons suggest common origins for the two RNAs. *Braz. J. Med. Biol. Res.* **1989**, *22*, 931–944.
11. De Farias, S.T.; do Rêgo, T.G.; José, M.V. Evolution of transfer RNA and the origin of the translation system. *Front. Genet.* **2014**, *5*, 303. [[CrossRef](#)]
12. Bowman, J.C.; Hud, N.V.; Williams, L.D. The ribosome challenge to the RNA world. *J. Mol. Evol.* **2015**, *80*, 143–161. [[CrossRef](#)] [[PubMed](#)]
13. Mauro, V.P.; Edelman, G.M. rRNA-like sequences occur in diverse primary transcripts: Implications for the control of gene expression. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 422–427. [[CrossRef](#)] [[PubMed](#)]
14. Tranque, P.; Hu, M.C.; Edelman, G.M.; Mauro, V.P. rRNA complementarity within mRNAs: A possible basis for mRNA-ribosome interactions and translational control. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12238–12243. [[CrossRef](#)] [[PubMed](#)]

15. Hu, M.C.; Tranque, P.; Edelman, G.M.; Mauro, V.P. rRNA-complementarity in the 5' untranslated region of mRNA specifying the Gtx homeodomain protein: Evidence that base-pairing to 18S rRNA affects translational efficiency. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 1339–1344. [CrossRef] [PubMed]
16. Dresios, J.; Chappell, S.A.; Zhou, W.; Mauro, V.P. An mRNA-rRNA base-pairing mechanism for translation initiation in eukaryotes. *Nat. Struct. Mol. Biol.* **2006**, *13*, 30–34. [CrossRef] [PubMed]
17. Kong, Q.; Stockinger, M.P.; Chang, Y.; Tashiro, H.; Lin, C.L. The presence of rRNA sequences in polyadenylated RNA and its potential functions. *Biotechnol J.* **2008**, *3*, 1041–1046. [CrossRef] [PubMed]
18. Lee, C.; Yen, K.; Cohen, P. Humanin: A harbinger of mitochondrial-derived peptides? *Trends Endocrinol. Metab.* **2013**, *24*, 222–228. [CrossRef] [PubMed]
19. Bonawitz, N.D.; Chatenay-Lapointe, M.; Wearn, C.M.; Shadel, G.S. Expression of the rDNA-encoded mitochondrial protein Tar1p is stringently controlled and responds differentially to mitochondrial respiratory demand and dysfunction. *Curr. Genet.* **2008**, *54*, 83–94. [CrossRef]
20. Coelho, P.S.; Bryan, A.C.; Kumar, A.; Shadel, G.S.; Snyder, M. A novel mitochondrial protein, Tar1p, is encoded on the antisense strand of the nuclear 25S rDNA. *Genes Dev.* **2002**, *16*, 2755–2760. [CrossRef]
21. Mercer, T.R.; Neph, S.; Dinger, M.E.; Crawford, J.; Smith, M.A.; Shearwood, A.M.; Haugen, E.; Bracken, C.P.; Rackham, O.; Stamatoyannopoulos, J.A.; et al. The human mitochondrial transcriptome. *Cell* **2011**, *146*, 645–658. [CrossRef]
22. Baserga, S.J.; Linnenbach, A.J.; Malcolm, S.; Ghosh, P.; Malcolm, A.D.; Takeshita, K.; Forget, B.G.; Benz, E.J., Jr. Polyadenylation of a human mitochondrial ribosomal RNA transcript detected by molecular cloning. *Gene* **1985**, *35*, 305–312. [CrossRef]
23. Amikura, R.; Kashikawa, M.; Nakamura, A.; Kobayashi, S. Presence of mitochondria-type ribosomes outside mitochondria in germ plasm of *Drosophila* embryos. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 9133–9138. [CrossRef] [PubMed]
24. Chooi, W.Y.; Leiby, K.R. The in vivo expression of pseudo ribosomal RNA genes in *Drosophila melanogaster*. *Mol. Gen. Genet.* **1981**, *182*, 245–251. [CrossRef] [PubMed]
25. Villegas, J.; Araya, P.; Bustos-Obregon, E.; Burzio, L.O. Localization of the 16S mitochondrial rRNA in the nucleus of mammalian spermatogenic cells. *Mol. Hum. Reprod.* **2002**, *8*, 977–983. [CrossRef] [PubMed]
26. Mignone, F.; Pesole, G. rRNA-like sequences in human mRNAs. *Appl. Bioinform.* **2002**, *1*, 145–154.
27. Jain, S.K.; Crampton, J.; Gonzalez, I.L.; Schmickel, R.D. Complementarity between ferritin H mRNA and 28 S ribosomal RNA. *Biochem. Biophys. Res. Commun.* **1985**, *131*, 863–867. [CrossRef]
28. Dooley, S.; Farber, U.; Welter, C.; Theisinger, B. Crosshybridization between the avian myeloblastosis oncogene and eukaryotic 28S ribosomal RNA. *Gene* **1992**, *110*, 263–264. [CrossRef]
29. Matveeva, O.V.; Shabalina, S.A. Intermolecular mRNA-rRNA hybridization and the distribution of potential interaction regions in murine 18S rRNA. *Nucleic Acids Res.* **1993**, *21*, 1007–1011. [CrossRef]
30. Stewart, F.J.; Ottesen, E.A.; DeLong, E.F. Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *ISME J.* **2010**, *4*, 896–907. [CrossRef]
31. Hei, S.; Wurtzel, O.; Singh, K.; Froula, J.L.; Yilmaz, S.; Tringe, S.G.; Wang, Z.; Chen, F.; Lindquist, E.A.; Sorek, R.; et al. Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. *Nat. Methods* **2010**, *7*, 807–812. [CrossRef]
32. Huang, Y.; Gilna, P.; Li, W. Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics* **2009**, *25*, 1338–1340. [CrossRef] [PubMed]
33. Schneider, R.; Lim, Y.W.; Edwards, R. Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics* **2011**, *28*, 433–435. [CrossRef] [PubMed]
34. Höps, W.; Jeffries, M.; Bateman, A. Gene Unprediction with Spurio: A tool to identify spurious protein sequences. *F1000Research* **2018**, *7*, 261. [CrossRef] [PubMed]
35. Ribopicker. 2017. Available online: <http://ribopicker.sourceforge.net/manual.html> (accessed on 29 December 2018).
36. Tripp, H.J.; Hewson, I.; Boyarsky, S.; Stuart, J.M.; Zehr, J.P. Misannotations of rRNA can now generate 90% false positive protein matches in metatranscriptomic studies. *Nucleic Acids Res.* **2011**, *39*, 8792–8802. [CrossRef] [PubMed]
37. Aziz, R.K.; Bartels, D.; Best, A.A.; DeJongh, M.; Disz, T.; Edwards, R.A.; Formsma, K.; Gerdes, S.; Glass, E.M.; Kubal, M. The RAST Server: Rapid annotations using subsystems technology. *BMC Genom.* **2008**, *9*, 75. [CrossRef] [PubMed]

38. Kermekchiev, M.; Ivanova, L. Ribin, a protein encoded by a message complementary to rRNA, modulates ribosomal transcription and cell proliferation. *Genes Dev.* **2002**, *16*, 2755–2760. [[CrossRef](#)]
39. Scharf, M.E.; Wu-Scharf, D.; Zhou, X.; Pittendrigh, B.R.; Bennett, G.W. Gene expression profiles among immature and adult reproductive castes of the termite *Reticulitermes flavipes*. *Insect Mol. Biol.* **2005**, *14*, 31–44. [[CrossRef](#)]
40. Mumtaz, M.A.; Couso, J.P. Ribosomal profiling adds new coding sequences to the proteome. *Biochem. Soc. Trans.* **2015**, *43*, 1271–1276. [[CrossRef](#)]
41. Treangen, T.J.; Salzberg, S.L. Repetitive DNA and next-generation sequencing: Computational challenges and solutions. *Nat. Rev. Genet.* **2012**, *13*, 36–46. [[CrossRef](#)]
42. Schmitz, J. SINEs as driving forces in genome evolution. *Genome Dyn.* **2012**, *7*, 92–107. [[CrossRef](#)]
43. Smit, A.F.; Riggs, A.D. MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. *Nucleic Acids Res.* **1995**, *23*, 98–102. [[CrossRef](#)] [[PubMed](#)]
44. Longo, M.S.; Brown, J.D.; Zhang, C.; O'Neill, M.J.; O'Neill, R.J. Identification of a recently active mammalian SINE derived from ribosomal RNA. *Genome Biol. Evol.* **2015**, *7*, 775–788. [[CrossRef](#)] [[PubMed](#)]
45. Munro, J.; Burdon, R.H.; Leader, D.P. Characterization of a human orphion 28 S ribosomal DNA. *Gene* **1986**, *48*, 65–70. [[CrossRef](#)]
46. Wang, S.; Pirtle, I.L.; Pirtle, R.M. A human 28S ribosomal RNA retropseudogene. *Gene* **1997**, *196*, 105–111. [[CrossRef](#)]
47. Nishihara, H.; Smit, A.F.A.; Okada, N. Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res.* **2006**, *16*, 864–874. [[CrossRef](#)] [[PubMed](#)]
48. Gogolevsky, K.P.; Vassetzky, N.S.; Kramerov, D.A. 5S rRNA-derived and tRNA-derived SINEs in fruit bats. *Genomics* **2009**, *93*, 494–500. [[CrossRef](#)] [[PubMed](#)]
49. Kanhayuwa, L.; Coutts, R.H. Short Interspersed Nuclear Element (SINE) Sequences in the Genome of the Human Pathogenic Fungus *Aspergillus fumigatus* Af293. *PLoS ONE* **2016**, *11*, e0163215. [[CrossRef](#)] [[PubMed](#)]
50. Kapitonov, V.V.; Jurka, J. A novel class of SINE elements derived from 5S rRNA. *Mol. Biol. Evol.* **2003**, *20*, 694–702. [[CrossRef](#)]
51. Jurka, J.; Kapitonov, V.V.; Pavlicek, A.; Klonowski, P.; Kohany, O.; Walichiewicz, J. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **2005**, *110*, 462–467. [[CrossRef](#)]
52. Frenkel, F.E.; Chaley, M.B.; Korotkov, E.V.; Skryabin, K.G. Evolution of tRNA-like sequences and genome variability. *Gene* **2004**, *335*, 57–71. [[CrossRef](#)]
53. Wegrzyn, G.; Wegrzyn, A. Is tRNA only a translation factor or also a regulator of other processes? *J. Appl. Genet.* **2008**, *49*, 115–122. [[CrossRef](#)]
54. Katz, A.; Elgamal, S.; Rajkovic, A.; Ibba, M. Non-canonical roles of tRNAs and tRNA mimics in bacterial cell biology. *Mol. Microbiol.* **2016**, *101*, 545–558. [[CrossRef](#)] [[PubMed](#)]
55. Lalaouna, D.; Carrier, M.C.; Massé, E. Every little piece counts: The many faces of tRNA transcripts. *Transcription* **2015**, *6*, 74–77. [[CrossRef](#)]
56. Bąkowska-Żywicka, K.; Mleczko, A.M.; Kasprzyk, M.; Machtel, P.; Żywicki, M.; Twardowski, T. The widespread occurrence of tRNA-derived fragments in *Saccharomyces cerevisiae*. *FEBS Open Bio* **2016**, *6*, 1186–1200. [[CrossRef](#)] [[PubMed](#)]
57. Thompson, D.M.; Lu, C.; Green, P.J.; Parker, R. tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA* **2008**, *14*, 2095–2103. [[CrossRef](#)]
58. Raab, J.R.; Chiu, J.; Zhu, J.; Katzman, S.; Kurukuti, S.; Wade, P.A.; Haussler, D.; Kamakaka, R.T. Human tRNA genes function as chromatin insulators. *EMBO J.* **2012**, *31*, 330–350. [[CrossRef](#)] [[PubMed](#)]
59. Ariza-Mateos, A.; Gómez, J. Viral tRNA Mimicry from a Biocommunicative Perspective. *Front. Microbiol.* **2017**, *8*, 2395. [[CrossRef](#)]
60. Sauguet, L.; Moutiez, M.; Li, Y.; Belin, P.; Seguin, J.; Le Du, M.H.; Thai, R.; Masson, C.; Fonvielle, M.; Pernodet, J.L.; et al. Cyclodipeptide synthases, a family of class-I aminoacyl-tRNA synthetase-like enzymes involved in non-ribosomal peptide synthesis. *Nucleic Acids Res.* **2011**, *39*, 4475–4489. [[CrossRef](#)] [[PubMed](#)]
61. Bärbel, J.F.; Thuss, W.S.; Wildum, S.; Binder, S. Mapping of mitochondrial mRNA termini in *Arabidopsis thaliana*: T-elements contribute to 5' and 3' end formation. *Nucleic Acids Res.* **2007**, *35*, 3676–3692. [[CrossRef](#)]

62. Hanic-Joyce, P.J.; Spencer, D.F.; Gray, M.W. In vitro processing of transcripts containing novel tRNA-like sequences ('t-elements') encoded by wheat mitochondrial DNA. *Plant Mol. Biol.* **1990**, *15*, 551–559. [[CrossRef](#)] [[PubMed](#)]
63. Babski, J.; Maier, L.K.; Heyer, R.; Jaschinski, K.; Prasse, D.; Jäger, D.; Randau, L.; Schmitz, R.A.; Marchfelder, A.; Soppa, J. Small regulatory RNAs in Archaea. *RNA Biol.* **2014**, *11*, 484–493. [[CrossRef](#)] [[PubMed](#)]
64. Keam, S.P.; Hutvagner, G. tRNA-Derived Fragments (tRFs): Emerging New Roles for an Ancient RNA in the Regulation of Gene Expression. *Life* **2015**, *5*, 1638–1651. [[CrossRef](#)] [[PubMed](#)]
65. Zhu, L.; Ow, D.W.; Dong, Z. Transfer RNA-derived small RNAs in plants. *Sci. China Life Sci.* **2018**, *61*, 155–161. [[CrossRef](#)] [[PubMed](#)]
66. Li, S.; Xu, Z.; Sheng, J. tRNA-Derived Small RNA: A Novel Regulatory Small Non-Coding RNA. *Genes* **2018**, *9*, 246. [[CrossRef](#)] [[PubMed](#)]
67. Mleczo, A.M.; Celichowski, P.; Bąkowska-Żywicka, K. Transfer RNA-derived fragments target and regulate ribosome-associated aminoacyl-transfer RNA synthetases. *Biochim. Biophys. Acta* **2018**, *1861*, 647–656. [[CrossRef](#)]
68. Diebel, K.W.; Zhou, K.; Clarke, A.B.; Bemis, L.T. Beyond the Ribosome: Extra-translational Functions of tRNA Fragments. *Biomark. Insights* **2016**, *11*, 1–8. [[CrossRef](#)] [[PubMed](#)]
69. Martinez, G. tRNA-derived small RNAs: New players in genome protection against retrotransposons. *RNA Biol.* **2018**, *15*, 170–175. [[CrossRef](#)] [[PubMed](#)]
70. Gebetsberger, J.; Wyss, L.; Mleczo, A.M.; Reuther, J.; Polacek, N. A tRNA-derived fragment competes with mRNA for ribosome binding and regulates translation during stress. *RNA Biol.* **2017**, *14*, 1364–1373. [[CrossRef](#)]
71. Giessen, T.W.; Marahiel, M.A. The tRNA-dependent biosynthesis of modified cyclic dipeptides. *Int. J. Mol. Sci.* **2014**, *15*, 14610–14631. [[CrossRef](#)]
72. Stewart, T.S.; Roberts, R.J.; Strominger, J.L. Novel species of tRNA. *Nature* **1971**, *230*, 36–38. [[CrossRef](#)]
73. Roy, H.; Ibbá, M. RNA-dependent lipid remodeling by bacterial multiple peptide resistance factors. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4667–4672. [[CrossRef](#)] [[PubMed](#)]
74. Wang, D.; Zhan, Y.; Cai, D.; Li, X.; Wang, Q.; Chen, S. Regulation of the Synthesis and Secretion of the Iron Chelator Cyclodipeptide Pulcherriminic Acid in *Bacillus licheniformis*. *Appl. Environ. Microbiol.* **2018**, *27*, 1–14. [[CrossRef](#)] [[PubMed](#)]
75. Mishra, A.K.; Choi, J.; Choi, S.J.; Baek, K.H. Cyclodipeptides: An Overview of Their Biosynthesis and Biological Activity. *Molecules* **2017**, *22*, 1796. [[CrossRef](#)] [[PubMed](#)]
76. Staves, M.P.; Lacey, J.C., Jr. On the probability of a common origin for tRNA and 5S rRNA. *Z. Naturforsch. C* **1989**, *44*, 296–306. [[CrossRef](#)] [[PubMed](#)]
77. Fu, Y.; Bannach, O.; Chen, H.; Teune, J.H.; Schmitz, A.; Steger, G.; Xiong, L.; Barbazuk, W.B. Alternative splicing of anciently exonized 5S rRNA regulates plant transcription factor TFIIIA. *Genome Res.* **2009**, *19*, 913–921. [[CrossRef](#)] [[PubMed](#)]
78. Barbazuk, W.B. A conserved alternative splicing event in plants reveals an ancient exonization of 5S rRNA that regulates TFIIIA. *RNA Biol.* **2010**, *7*, 397–402. [[CrossRef](#)] [[PubMed](#)]
79. Hammond, M.C.; Wachter, A.; Breaker, R.R. A plant 5S ribosomal RNA mimic regulates alternative splicing of transcription factor IIIA pre-mRNAs. *Nat. Struct. Mol. Biol.* **2009**, *16*, 541–549. [[CrossRef](#)]
80. Donati, G.; Peddigari, S.; Mercer, C.A.; Thomas, G. 5S Ribosomal RNA Is an Essential Component of a Nascent Ribosomal Precursor Complex that Regulates the Hdm2-p53 Checkpoint. *Cell Rep.* **2013**, *4*, 87–98. [[CrossRef](#)]
81. Fox, P.L. Discovery and investigation of the GAIT translational control system. *RNA* **2015**, *21*, 615–618. [[CrossRef](#)]
82. Arkov, A.L.; Mankin, A.; Murgola, E.J. An rRNA Fragment and Its Antisense Can Alter Decoding of Genetic Information. *J. Bacteriol.* **1998**, *180*, 2744–2748.
83. Chernyaeva, N.S.; Murgola, E.J.; Mankin, A.S. Suppression of Nonsense Mutations Induced by Expression of an RNA Complementary to a Conserved Segment of 23S rRNA. *J. Bacteriol.* **1999**, *181*, 5257–5262. [[PubMed](#)]
84. Feagin, J.E.; Mericle, B.L.; Werner, E.; Morris, M. Identification of additional rRNA fragments encoded by the *Plasmodium falciparum* 6 kb element. *Nucleic Acids Res.* **1997**, *25*, 438–446. [[CrossRef](#)] [[PubMed](#)]
85. Chen, Z.; Sun, Y.; Yang, X.; Wu, Z.; Guo, K.; Niu, X.; Wang, Q.; Ruan, J.; Bu, W.; Gao, S. Two featured series of rRNA-derived RNA fragments (rRFs) constitute a novel class of small RNAs. *PLoS ONE* **2017**, *12*, e0176458. [[CrossRef](#)]

86. Shi, J.; Ko, E.A.; Sanders, K.M.; Chen, Q.; Zhou, T. SPORTS1.0: A Tool for Annotating and Profiling Non-coding RNAs Optimized for rRNA- and tRNA-derived Small RNAs. *Genom. Proteom. Bioinform.* **2018**, *16*, 144–151. [[CrossRef](#)] [[PubMed](#)]
87. Hadjiargyrou, M.; Delihias, N. The intertwining of transposable elements and non-coding RNAs. *Int. J. Mol. Sci.* **2013**, *14*, 13307–13328. [[CrossRef](#)] [[PubMed](#)]
88. Cerutti, H.; Casas-Mollano, J.A. On the origin and functions of RNA-mediated silencing: From protists to man. *Curr. Genet.* **2006**, *50*, 81–99. [[CrossRef](#)] [[PubMed](#)]
89. Bhavsarc, R.B.; Makley, L.N.; Tsonis, P.A. The other lives of ribosomal proteins. *Hum. Genom.* **2010**, *4*, 327–344. [[CrossRef](#)]
90. Warner, J.R.; McIntosh, K.B. How common are extraribosomal functions of ribosomal proteins? *Mol. Cell* **2009**, *34*, 3–11. [[CrossRef](#)]
91. Wool, I.G. Extraribosomal functions of ribosomal proteins. *Trends Biochem. Sci.* **1996**, *21*, 164–165. [[CrossRef](#)]
92. Graifer, D.; Malygin, A.; Zharkov, D.O.; Karpova, G. Eukaryotic ribosomal protein S3: A constituent of translational machinery and an extraribosomal player in various cellular processes. *Biochimie* **2014**, *99*, 8–18. [[CrossRef](#)]
93. Naora, H.; Naora, H. Involvement of ribosomal proteins in regulating cell growth and apoptosis: Translational modulation or recruitment for extraribosomal activity? *Immunol. Cell Biol.* **1999**, *77*, 197–205. [[CrossRef](#)]
94. Chen, J.; Zhao, R.; Semba, U.; Oda, M.; Suzuki, T.; Toba, K.; Hattori, S.; Okada, S.; Yamamoto, T. Involvement of cross-linked ribosomal protein S19 oligomers and C5a receptor in definitive erythropoiesis. *Exp. Mol. Pathol.* **2013**, *95*, 364–375. [[CrossRef](#)] [[PubMed](#)]
95. Lu, H.; Zhu, Y.F.; Xiong, J.; Wang, R.; Jia, Z. Potential extra-ribosomal functions of ribosomal proteins in *Saccharomyces cerevisiae*. *Microbiol. Res.* **2015**, *177*, 28–33. [[CrossRef](#)]
96. Scheiman, J.; Jamieson, K.V.; Ziello, J.; Tseng, J.C.; Meruelo, D. Extraribosomal functions associated with the C terminus of the 37/67 kDa laminin receptor are required for maintaining cell viability. *Cell Death Dis* **2010**, *1*, e42. [[CrossRef](#)] [[PubMed](#)]
97. Zhou, X.; Liao, W.J.; Liao, J.M.; Liao, P.; Lu, H. Ribosomal proteins: Functions beyond the ribosome. *J. Mol. Cell Biol.* **2015**, *7*, 92–104. [[CrossRef](#)] [[PubMed](#)]
98. Han, S.H.; Chung, J.H.; Kim, J.; Kim, K.S.; Han, Y.S. New role of human ribosomal protein S3: Regulation of cell cycle via phosphorylation by cyclin-dependent kinase 2. *Oncol. Lett.* **2017**, *13*, 3681–3687. [[CrossRef](#)] [[PubMed](#)]
99. Patil, A.V.; Hsieh, T.S. Ribosomal Protein S3 Negatively Regulates Unwinding Activity of RecQ-like Helicase 4 through Their Physical Interaction. *J. Biol. Chem.* **2017**, *292*, 4313–4325. [[CrossRef](#)] [[PubMed](#)]
100. Lee, S.B.; Kwon, I.S.; Park, J.; Lee, K.H.; Ahn, Y.; Lee, C.; Kim, J.; Choi, S.Y.; Cho, S.W.; Ahn, J.Y. Ribosomal protein S3, a new substrate of Akt, serves as a signal mediator between neuronal apoptosis and DNA repair. *J. Biol. Chem.* **2010**, *285*, 29457–29468. [[CrossRef](#)]
101. Lindstrom, M.S. Emerging functions of ribosomal proteins in gene-specific transcription and translation. *Biochem. Biophys. Res. Commun.* **2009**, *379*, 167–170. [[CrossRef](#)]
102. Sudhamalla, B.; Kumar, M.; Kumar, R.S.; Sashi, P.; Yasin, U.M.; Ramakrishna, D.; Rao, P.N.; Bhuyan, A.K. Enzyme dimension of the ribosomal protein S4 across plant and animal kingdoms. *Biochim. Biophys. Acta* **2013**, *830*, 5335–5341. [[CrossRef](#)]
103. Ba, Q.; Li, X.; Huang, C.; Li, J.; Fu, Y.; Chen, P.; Duan, J.; Hao, M.; Zhang, Y.; Li, J.; et al. BCCIP β modulates the ribosomal and extraribosomal function of S7 through a direct interaction. *J. Mol. Cell Biol.* **2017**, *9*, 209–219. [[CrossRef](#)] [[PubMed](#)]
104. Dosil, M. Ribosome synthesis-unrelated functions of the preribosomal factor Rrp12 in cell cycle progression and the DNA damage response. *Mol. Cell. Biol.* **2011**, *31*, 2422–2438. [[CrossRef](#)] [[PubMed](#)]
105. Russo, A.; Esposito, D.; Catillo, M.; Pietropaolo, C.; Crescenzi, E.; Russo, G. Human rpL3 induces G₁/S arrest or apoptosis by modulating p21^{waf1/cip1} levels in a p53-independent manner. *Cell Cycle* **2013**, *12*, 76–87. [[CrossRef](#)]
106. Mazumder, B.; Poddar, D.; Basu, A.; Kour, R.; Verbovetskaya, V.; Barik, S. Extraribosomal I13a is a specific innate immune factor for antiviral defense. *J. Virol.* **2014**, *88*, 9100–9110. [[CrossRef](#)] [[PubMed](#)]
107. Jia, J.; Arif, A.; Willard, B.; Smith, J.D.; Stuehr, D.J.; Hazen, S.L.; Fox, P.L. Protection of extraribosomal RPL13a by GAPDH and dysregulation by S-nitrosylation. *Mol. Cell* **2012**, *47*, 656–663. [[CrossRef](#)]

108. Zhang, Y.; O'Leary, M.N.; Peri, S.; Wang, M.; Zha, J.; Melov, S.; Kappes, D.J.; Feng, Q.; Rhodes, J.; Amieux, P.S.; et al. Ribosomal Proteins Rpl22 and Rpl22l1 Control Morphogenesis by Regulating Pre-mRNA Splicing. *Cell Rep.* **2017**, *18*, 545–556. [[CrossRef](#)] [[PubMed](#)]
109. Zhang, Y.; Duc, A.C.; Rao, S.; Sun, X.L.; Bilbee, A.N.; Rhodes, M.; Li, Q.; Kappes, D.J.; Rhodes, J.; Wiest, D.L. Control of hematopoietic stem cell emergence by antagonistic functions of ribosomal protein paralogs. *Dev. Cell* **2013**, *24*, 411–425. [[CrossRef](#)] [[PubMed](#)]
110. Yang, M.; Sun, H.; He, J.; Wang, H.; Yu, X.; Ma, L.; Zhu, C. Interaction of ribosomal protein L22 with casein kinase 2 α : A novel mechanism for understanding the biology of non-small cell lung cancer. *Oncol. Rep.* **2014**, *32*, 139–144. [[CrossRef](#)]
111. Zsögön, A.; Szakonyi, D.; Shi, X.; Byrne, M.E. Ribosomal Protein RPL27a Promotes Female Gametophyte Development in a Dose-Dependent Manner. *Plant Physiol.* **2014**, *165*, 1133–1143. [[CrossRef](#)]
112. Provost, E.; Weier, C.A.; Leach, S.D. Multiple ribosomal proteins are expressed at high levels in developing zebrafish endoderm and are required for normal exocrine pancreas development. *Zebrafish* **2013**, *10*, 161–169. [[CrossRef](#)]
113. Caprara, M.G.; Lehnert, V.; Lambowitz, A.M.; Westhof, E. A tyrosyl-tRNA synthetase recognizes a conserved tRNA-like structural motif in the group I intron catalytic core. *Cell* **1996**, *87*, 1135–1145. [[CrossRef](#)]
114. Mohr, G.; Rennard, R.; Cherniack, A.D.; Stryker, J.; Lambowitz, A.M. Function of the *Neurospora crassa* mitochondrial tyrosyl-tRNA synthetase in RNA splicing. Role of the idiosyncratic N-terminal extension and different modes of interaction with different group I introns. *J. Mol. Biol.* **2001**, *307*, 75–92. [[CrossRef](#)] [[PubMed](#)]
115. Boniecki, M.T.; Rho, S.B.; Tukalo, M.; Hsu, J.L.; Romero, E.P.; Martinis, S.A. Leucyl-tRNA synthetase-dependent and -independent activation of a group I intron. *J. Biol. Chem.* **2009**, *284*, 26243–26250. [[CrossRef](#)]
116. Rho, S.B.; Lincecum, T.L., Jr.; Martinis, S.A. An inserted region of leucyl-tRNA synthetase plays a critical role in group I intron splicing. *EMBO J.* **2002**, *21*, 6874–6881. [[CrossRef](#)] [[PubMed](#)]
117. Zamecnik, P.C.; Stephenson, M.L.; Janaway, C.M.; Randerath, K. Enzymatic synthesis of diadenosine tetraphosphate and diadenosine triphosphate with a purified lysyl-sRNA synthetase. *Biochem. Biophys. Res. Commun.* **1966**, *24*, 91–105. [[CrossRef](#)]
118. Randerath, K.; Janaway, C.M.; Stephenson, M.L.; Zamecnik, P.C. Isolation and characterization of dinucleoside tetra- and tri-phosphates formed in the presence of lysyl-sRNA synthetase. *Biochem. Biophys. Res. Commun.* **1966**, *24*, 98–105. [[CrossRef](#)]
119. Romby, P.; Caillet, J.; Ebel, C.; Sacerdot, C.; Graffe, M.; Eyermann, F.; Brunel, C.; Moine, H.; Ehresmann, C.; Ehresmann, B.; et al. The expression of *E. coli* threonyl-tRNA synthetase is regulated at the translational level by symmetrical operator-repressor interactions. *EMBO J.* **1966**, *15*, 5976–5987. [[CrossRef](#)]
120. Wakasugi, K.; Schimmel, P. Two distinct cytokines released from a human aminoacyl-tRNA synthetase. *Science* **1999**, *284*, 147–151. [[CrossRef](#)] [[PubMed](#)]
121. Ivakhno, S.S.; Kornelyuk, A.I. Cytokine-like activities of some aminoacyl-tRNA synthetases and auxiliary p43 cofactor of aminoacylation reaction and their role in oncogenesis. *Exp. Oncol.* **2004**, *26*, 250–255.
122. Sajish, M.; Zhou, Q.; Kishi, S.; Valdez, D.M., Jr.; Kapoor, M.; Guo, M.; Lee, S.; Kim, S.; Yang, X.L.; Schimmel, P. Trp-tRNA synthetase bridges DNA-PKcs to PARP-1 to link IFN- γ and p53 signaling. *Nat. Chem. Biol.* **2012**, *8*, 547–554. [[CrossRef](#)]
123. Mirando, A.C.; Francklyn, C.S.; Lounsbury, K.M. Regulation of angiogenesis by aminoacyl-tRNA synthetases. *Int. J. Mol. Sci.* **2014**, *15*, 23725–23748. [[CrossRef](#)] [[PubMed](#)]
124. Brown, M.V.; Reader, J.S.; Tzima, E. Mammalian aminoacyl-tRNA synthetases: Cell signaling functions of the protein translation machinery. *Vasc. Pharmacol.* **2010**, *52*, 21–26. [[CrossRef](#)] [[PubMed](#)]
125. Lee, S.W.; Cho, B.H.; Park, S.G.; Kim, S. Aminoacyl-tRNA synthetase complexes: Beyond translation. *J. Cell Sci.* **2004**, *117*, 3725–3734. [[CrossRef](#)] [[PubMed](#)]
126. Ivanov, K.A.; Moor, N.A.; Lavrik, O.I. Non-canonical functions of aminoacyl-tRNA synthetases. *Biochemistry* **2000**, *65*, 888–897. [[PubMed](#)]
127. Pang, Y.L.; Poruri, K.; Martinis, S.A. tRNA synthetase: TRNA aminoacylation and beyond. *Wiley Interdiscip. Rev. RNA* **2014**, *5*, 461–480. [[CrossRef](#)] [[PubMed](#)]
128. Schimmel, P.; De Pouplana, L.R. Footprints of aminoacyl-tRNA synthetases are everywhere. *Trends Biochem. Sci.* **2000**, *25*, 207–209. [[CrossRef](#)]

129. Salazar, J.C.; Ambrogelly, A.; Crain, P.F.; McCloskey, J.A.; Soll, D. A truncated aminoacyl-tRNA synthetase modifies RNA. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7536–7541. [[CrossRef](#)]
130. Campanacci, V.; Dubois, D.Y.; Becker, H.D.; Kern, D.; Spinelli, S.; Valencia, C.; Pagot, F.; Salomoni, A.; Grisel, S.; Vincentelli, R. The Escherichia coli YadB gene product reveals a novel aminoacyl-tRNA synthetase like activity. *J. Mol. Biol.* **2004**, *337*, 273–283. [[CrossRef](#)]
131. Sissler, M.; Delorme, C.; Bond, J.; Ehrlich, S.D.; Renault, P.; Francklyn, C. An aminoacyl-tRNA synthetase paralog with a catalytic role in histidine biosynthesis. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 8985–8990. [[CrossRef](#)]
132. Xu, W.; Li, L.; Du, L.; Tan, N. Various mechanisms in cyclopeptide production from precursors synthesized independently of non-ribosomal peptide synthetases. *Acta Biochim. Biophys. Sin.* **2011**, *43*, 757–762.
133. Bonnefond, L.; Arai, T.; Sakaguchi, Y.; Suzuki, T.; Ishitani, R.; Nureki, O. Structural basis for nonribosomal peptide synthesis by an aminoacyl-tRNA synthetase paralog. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3912–3917. [[CrossRef](#)]
134. Chaliotis, A.; Vlastaridis, P.; Mossialos, D.; Ibba, M.; Becker, H.D.; Stathopoulos, C.; Amoutzias, G.D. The complex evolutionary history of aminoacyl-tRNA synthetases. *Nucleic Acids Res.* **2017**, *45*, 1059–1068. [[CrossRef](#)] [[PubMed](#)]
135. Francklyn, C. tRNA synthetase paralogs: Evolutionary links in the transition from tRNA-dependent amino acid biosynthesis to de novo biosynthesis. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 9650–9652. [[CrossRef](#)] [[PubMed](#)]
136. Guo, M.; Schimmel, P. Essential nontranslational functions of tRNA synthetases. *Nat. Chem. Biol.* **2013**, *9*, 145–153. [[CrossRef](#)] [[PubMed](#)]
137. Park, S.G.; Ewalt, K.L.; Kim, S. Functional expansion of aminoacyl-tRNA synthetases and their interacting factors: New perspectives on housekeepers. *Trends Biochem. Sci.* **2005**, *30*, 569–574. [[CrossRef](#)] [[PubMed](#)]
138. Yakobov, N.; Debard, S.; Fischer, F.; Senger, B.; Becker, H.D. Cytosolic aminoacyl-tRNA synthetases: Unanticipated relocations for unexpected functions. *Biochim. Biophys. Acta* **2018**, *1861*, 387–400. [[CrossRef](#)] [[PubMed](#)]
139. Singh, I.; Yadav, A.R.; Mohanty, K.K.; Katoch, K.; Sharma, P.; Mishra, B.; Bisht, D.; Gupta, U.D.; Sengupta, U. Molecular mimicry between Mycobacterium leprae proteins (50S ribosomal protein L2 and Lysyl-tRNA synthetase) and myelin basic protein: A possible mechanism of nerve damage in leprosy. *Microbes Infect.* **2015**, *17*, 247–257. [[CrossRef](#)]
140. Steitz, T.A. A mechanism for all polymerases. *Nature* **1998**, *391*, 231–232. [[CrossRef](#)]
141. Cramer, P. Multisubunit RNA polymerases. *Curr. Opin. Struct. Biol.* **2002**, *12*, 89–97. [[CrossRef](#)]
142. Iyer, L.M.; Koonin, E.V.; Aravind, L. Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Struct. Biol.* **2003**, *3*, 1. [[CrossRef](#)]
143. Maida, Y.; Masutomi, K. Telomerase reverse transcriptase moonlights: Therapeutic targets beyond telomerase. *Cancer Sci.* **2015**, *106*, 1486–1492. [[CrossRef](#)]
144. Romaniuk, A.; Paszal-Jaworska, A.; Totoń, E.; Lisiak, N.; Hołysz, H.; Królak, A.; Grodecka-Gazdecka, S.; Rubiś, B. The non-canonical functions of telomerase: To turn off or not to turn off. *Mol. Biol. Rep.* **2018**. [[CrossRef](#)] [[PubMed](#)]
145. Alcaraz-Pérez, F.; García-Castillo, J.; García-Moreno, D.; López-Muñoz, A.; Anhelín, M.; Angosto, D.; Zon, L.I.; Mulero, V.; Cayuela, M.L. A non-canonical function of telomerase RNA in the regulation of developmental myelopoiesis in zebrafish. *Nat. Commun.* **2014**, *5*, 3228. [[CrossRef](#)] [[PubMed](#)]
146. Brandman, R.; Brandman, Y.; Pande, V.S. Sequence coevolution between RNA and protein characterized by mutual information between residue triplets. *PLoS ONE* **2012**, *7*, e30022. [[CrossRef](#)] [[PubMed](#)]
147. Altman, S. The RNA-protein world. *RNA* **2013**, *19*, 589–590. [[CrossRef](#)] [[PubMed](#)]
148. Liberman, J.A.; Wedekind, J.E. Base ionization and ligand binding: How small ribozymes and riboswitches gain a foothold in a protein world. *Curr. Opin. Struct. Biol.* **2011**, *21*, 327–334. [[CrossRef](#)] [[PubMed](#)]
149. Wong, T.-F.; Ng, S.-K.; Mat, W.-K.; Hu, T.; Xue, H. Coevolution theory of the genetic code at age forty: Pathway to translation and synthetic life. *Life* **2016**, *6*, 12. [[CrossRef](#)] [[PubMed](#)]
150. Carter, C.W., Jr. What RNA world? Why a peptide/RNA partnership merits renewed experimental attention. *Life* **2015**, *5*, 294–320. [[CrossRef](#)]
151. Cech, T.R. The RNA worlds in context. *Cold Spring Harb. Perspect. Biol.* **2012**, *4*, a006742. [[CrossRef](#)]
152. Cech, T.R. Structural biology. The ribosome is a ribozyme. *Science* **2000**, *289*, 878–879. [[CrossRef](#)]

153. Lilley, D.M.J.; Sutherland, J. The chemical origins of life and its early evolution: An introduction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2011**, *366*, 2853–2856. [[CrossRef](#)]
154. Van der Gulik, P.T.S.; Speijer, D. How amino acids and peptides shaped the RNA World. *Life* **2015**, *5*, 230–246. [[CrossRef](#)] [[PubMed](#)]
155. Wächtershäuser, G. The place of RNA in the origin and early evolution of the genetic machinery. *Life* **2014**, *4*, 1050–1091. [[CrossRef](#)] [[PubMed](#)]
156. Steitz, T.A.; Moore, P.B. RNA, the first macromolecular catalyst: The ribosome is a ribozyme. *Trends Biochem. Sci.* **2003**, *28*, 411–418. [[CrossRef](#)]
157. Ikehara, K. [GADV]-Protein world hypothesis on the origin of life. *Orig. Life Evol. Biosph.* **2014**, *44*, 299–302. [[CrossRef](#)] [[PubMed](#)]
158. Sharov, A.A. Coenzyme world model of the origin of life. *Biosystems* **2016**, *144*, 8–17. [[CrossRef](#)] [[PubMed](#)]
159. Fox, G.E. Origin and evolution of the ribosome. *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a003483. [[CrossRef](#)] [[PubMed](#)]
160. Hury, J.; Nagaswamy, U.; Larios-Sanz, M.; Fox, G.E. Ribosome origins: The relative age of 23S rRNA Domains. *Orig. Life Evol. Biosph.* **2006**, *36*, 421–429. [[CrossRef](#)]
161. Lecompte, O.; Ripp, R.; Thierry, J.C.; Moras, D.; Poch, O. Comparative analysis of ribosomal proteins in complete genomes: An example of reductive evolution at the domain scale. *Nucleic Acids Res.* **2002**, *30*, 5382–5390. [[CrossRef](#)]
162. Söding, J.; Lupa, A.N. More than the sum of their parts: On the evolution of proteins from peptides. *Bioessays* **2003**, *25*, 837–846. [[CrossRef](#)]
163. Kovacs, N.A.; Petrov, A.S.; Lanier, K.A.; Williams, L.D. Frozen in time: The history of proteins. *Mol. Biol. Evol.* **2017**, *34*, 1252–1260. [[CrossRef](#)]
164. Simon, H.A. *The Sciences of the Artificial*; MIT Press: Cambridge, MA, USA, 1969.
165. Hunding, A.; Kepes, F.; Lancet, D.; Minsky, A.; Norris, V.; Raine, D.; Sriram, K.; Root-Bernstein, R. Compositional complementarity and prebiotic ecology in the origin of life. *Bioessays* **2006**, *28*, 399–412. [[CrossRef](#)] [[PubMed](#)]
166. Root-Bernstein, R.S.; Dillon, P.F. Molecular complementarity I: The complementarity theory of the origin and evolution of life. *J. Theor. Biol.* **1997**, *188*, 447–479. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).