Expansion segments in bacterial and archaeal 5S ribosomal RNAs

VICTOR G. STEPANOV and GEORGE E. FOX

Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204-5001, USA

ABSTRACT

The large ribosomal RNAs of eukaryotes frequently contain expansion sequences that add to the size of the rRNAs but do not affect their overall structural layout and are compatible with major ribosomal function as an mRNA translation machine. The expansion of prokaryotic ribosomal RNAs is much less explored. In order to obtain more insight into the structural variability of these conserved molecules, we herein report the results of a comprehensive search for the expansion sequences in prokaryotic 5S rRNAs. Overall, 89 expanded 5S rRNAs of 15 structural types were identified in 15 archaeal and 36 bacterial genomes. Expansion segments ranging in length from 13 to 109 residues were found to be distributed among 17 insertion sites. The strains harboring the expanded 5S rRNAs belong to the bacterial orders *Clostridiales*, *Halanaerobiales*, *Thermoanaerobacterales*, and *Alteromonadales* as well as the archaeal order *Halobacterales*. When several copies of a 5S rRNA gene are present in a genome, the expanded versions may coexist with normal 5S rRNA genes. The insertion sequences are typically capable of forming extended helices, which do not seemingly interfere with folding of the conserved core. The expanded 5S rRNAs have largely been overlooked in 5S rRNA databases.

Keywords: ribosome; 5S rRNA; expansion segment; archaea; bacteria

INTRODUCTION

5S ribosomal RNA (5S rRNA) is a small RNA that is an integral constituent of the central protuberance of the large ribosomal subunit. Although not decisively proven, 5S rRNA is believed to be a mediator of allosteric interactions between the ribosomal functional centers (Dontsova and Dinman 2005; Kiparisov et al. 2005; Kouvela et al. 2007; Gongadze 2011) and an important contributor to efficient and correct ribosome assembly (Huang et al. 2020). On the ribosome, it forms multiple contacts with several structural elements that are linked to the peptidyl transferase center, elongation factor binding region, and mRNA decoding site. Such an intense involvement in intraribosomal interactions imposes severe structural and functional constraints on evolutionary changes, thus making the 5S rRNA one of the most conserved ribosomal components with respect to sequence, secondary structure, and three-dimensional layout (Szymanski et al. 1999; Smimov et al. 2008; Sun and Caetano-Anolles 2009). The 5S rRNA topology remains essentially the same across all three domains of life. Substantial deviations from the consensus structure are extremely rare and therefore are targets of particular interest.

Expansion segments in 5S rRNAs can certainly be regarded as a very exotic structural abnormality that deserves special attention. The term “expansion segment” is usually applied to taxon-specific nucleotide stretches intervening between the universally conserved elements of large ribosomal RNAs (Melnikov et al. 2012; Parker et al. 2014; Weisser and Ban 2019). Despite being long thought of as a unique distinctive feature of eukaryotic ribosomes, the expansion segments have recently been identified in several bacterial and archaeal 16S and 23S rRNAs (Roller et al. 1992; Everett et al. 1999; Armache et al. 2012; Greber et al. 2012; Yang et al. 2017; Kushwaha and Bhushan 2020; Penev et al. 2020). Meantime, they were never observed in 5S rRNAs apart from two remarkable exceptions. In 1975 Sutton and Woese reported that the 5S rRNA pool from the ribosomes of thermophilic hemicellulolytic bacterium *Thermoanaerobacterium thermosaccharolyticum* (formerly known as *Clostridium thermosaccharolyticum*) consists of two 5S rRNA types, one of the usual length of approximately 120 residues and the other significantly longer containing about 160 nt (Sutton and Woese 1975). After partially assessing RNA sequences using an oligonucleotide fingerprinting assay, they suggested that the oversized 5S rRNA corresponds to an underprocessed

Corresponding author: fox@uh.edu

Article is online at <http://www.rnajournal.org/cgi/doi/10.1261/rna.077123.120>. Freely available online through the RNA Open Access option.

© 2021 Stepanov and Fox This article, published in *RNA*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

precursor with 40 unremoved 3'-terminal residues that is able to functionally substitute for the properly matured 5S rRNA in the ribosomes. They ruled out the possibility that the oversized 5S rRNA represents a novel 5S rRNA species on the grounds that such a drastic change would be very unlikely within the established ribosomal framework made of multiple coevolved components. However, these conclusions were disproven after the whole-genome sequences of several *Th. thermosaccharolyticum* strains became available in the 2010s (Hemme et al. 2010; Jiang et al. 2018; Li et al. 2018). It is now known that this oversized 5S rRNA is encoded separately from the normal 5S rRNA, has a total length of 155 nt, and contains a single 46-nt insertion embedded deep within an otherwise typical 5S rRNA core sequence. In the context of the canonical secondary structure of 5S rRNA, the insertion extends from the universally conserved bulge in helix III. It is predicted to fold as an independent structural module while leaving undisturbed the conventional base-pairing scheme of the 5S rRNA core.

In 1981, a second 5S rRNA with an expansion segment was discovered (Luehrsen et al. 1981). It was found that the entire 5S rRNA pool from the halophilic archaeon *Halococcus morrhuae* consists of molecules far exceeding the size of an average archaeal 5S rRNA. Direct enzymatic sequencing of this extra-long 5S rRNA species revealed a 108-nt insertion located immediately adjacent to the so-called "loop E" in the conserved core. The length of the whole molecule was determined to be 231 nt, which is nearly double the usual length. Since no other 5S rRNA was found in the *H. morrhuae* cells, this oversized 5S rRNA was assumed to be functional by default. However, the involvement of the 5S rRNA expansion segment in ribosomal mechanisms has remained unexplored. Intriguingly, it has been noticed that a closely related archaeon *Halobacterium salinarum* (referred to as *Halobacterium cutirubrum* at the time of the study) has a conventional 5S rRNA without an expansion segment but otherwise highly similar to the *H. morrhuae* 5S rRNA with 89% identity over aligned residues.

Subsequent studies confirmed the presence of expanded 5S rRNAs in all then-known *H. morrhuae* strains and in several uncategorized *Halococcus* isolates (Nicholson 1982). In addition to *H. morrhuae*, the expansion segments were later identified in 5S rRNAs from two other *Halococcus* species, *H. salifodinae* and *H. saccharolyticus* (Stan-Lotter et al. 1999). At the same time, no insertions was found in 5S rRNAs from other haloarchaeal genera, *Halobacterium* and *Haloarcula*, in spite of their close phylogenetic relationships with the genus *Halococcus* (Nicholson 1982; Nicholson and Fox 1983; Daniels et al. 1985).

It is worth noting that the only known 5S rRNAs with expansion segments have been observed in the prokaryotic realm where they would be the least anticipated. This

means that both bacterial and archaeal ribosomes are apparently capable of accommodating relatively large additional modules in the immediate vicinity of the 5S rRNA core without losing activity. It does not seem improbable because 5S rRNA is located at the periphery of the ribosome and only partially buried in it. The expansion segments may therefore be rooted in the exposed parts of the 5S rRNA core, and either occupy the nearby cavities or protrude outwards into the surrounding environment (Tirumalai et al. 2020). At the same time, it remains puzzling why so few examples of the expanded 5S rRNAs have been reported so far if the steric constraints on 5S rRNA enlargement are not particularly prohibitive.

In order to assess the abundance of the expanded 5S rRNAs in prokaryotes, we have performed a broad screen of 5S rRNA sequence databases and publicly available genome sequences of bacteria and archaea. A number of novel oversized 5S rRNAs have been identified, some of them harboring not just one but two or even three expansion segments. In the present paper, we describe our search strategy and findings, discuss putative origin and function of the expansion segments in 5S rRNAs, and explain why the expanded 5S rRNAs have been largely overlooked up until now. Throughout the paper, the term "5S rRNA" designates the particular type of ribosomal RNAs rather than the actual size of the RNA.

RESULTS

Description and classification of the expansion segments requires a universal nucleotide numbering system compatible with significant size variations in both bacterial and archaeal 5S rRNAs. For the purpose of the present study, the 5S rRNA bases were numbered according to a previously proposed scheme, which treats the conserved core and each of the insertions as separate nucleotide series (Fox 1985; Erdmann and Wolters 1986; Wolters and Erdmann 1988). Position numbering in the conserved core follows the numbering scheme of the *E. coli* 5S rRNA, which is a long-standing structural archetype of the 5S rRNA family. Positions, which do not belong to the core, are presented in the form of a decimal fraction preceded with a number of the upstream core position as a prefix. Such an approach allows one to accommodate the expansion segments at various insertion points in any 5S rRNA without the necessity of revising nucleotide numbering throughout the entire RNA chain after each new finding (Fig. 1).

Some 5S rRNAs have a shortened Helix I as compared to the *E. coli* 5S rRNA, which makes ambiguous the position numbering in this region. To provide consistent nucleotide nomenclature, we used conserved base pair G7-C113 in Helix I as a landmark. Once the G7-C113 base pair is backtracked from the interior of the molecule, the adjacent residues are numbered relative to this position. Another problematic area is Helix IV, which in some cases is shorter

FIGURE 1. Generalized secondary structure model of prokaryotic 5S rRNA with indicated locations of expansion segments. Universally conserved nucleotides are shown as white characters in black filled circles. Dashed open circles represent nucleotides unique to archaeal 5S rRNAs. Dashed open diamonds represent a unique base pair observed only in 5S rRNAs from the bacterial family *Lachnospiraceae*. Canonical Watson–Crick bonding is depicted with solid lines. Non-Watson–Crick bonding is presented as small open circles. Dotted lines indicate that the bases may be bound by either Watson–Crick or non-Watson–Crick interactions. Expansion segments are schematically presented as thick red (Bacteria) or magenta (Archaea) lines connecting the flanking residues of the conserved 5S rRNA core.

than standard. The solution here was to locate the universal U77–A99 and A78–G98 interactions, and preserve four base Loop D as positions 87, 88, 89, and 89.1 while the stem residues between these two markers were numbered to maximize compatibility with the nucleotide conservation pattern observed in Helix IV of regular length. In archaeal 5S rRNAs, base pairs within the “loop E” motif were arranged in order to comply with its layout as seen in the X-ray crystal structure of the *Haloarcula marismortui* 50S ribosomal subunit (Ban et al. 2000; Szymanski et al. 2002; Gabdulkhakov et al. 2013). A single-nucleotide bulge loop was placed between positions 76 and 77 instead of between positions 74 and 75 as in earlier secondary structure models (Erdmann and Wolters 1986; Erdmann et al. 1987; Wolters and Erdmann 1988, 1989).

The database search for 5S rRNAs with expansion segments initially used the two previously reported examples from the archaeon *H. morrhuae* (Fig. 2A) and the bacterium

Th. thermosaccharolyticum (Fig. 2B). Both of these 5S rRNAs have one large insertion in the conserved 5S rRNA core. Using these two sequences as queries for BLAST searches of the NCBI nt database ([ftp://ftp.ncbi.nlm.nih.gov/blast/db/](http://ftp.ncbi.nlm.nih.gov/blast/db/)) made it possible to identify similar 5S rRNAs in 15 archaeal strains from the genera *Halococcus*, *Halalkalicoccus*, *Halomarina* and *Halodesulfurarchaeum*, and in 20 bacterial strains from the genera *Thermoanaerobacterium*, *Thermoanaerobacter*, *Caldanaerobacter*, and *Caldanaerobius*, respectively. Next, a more rigorous analysis of higher taxa that contain these genera was conducted. A local FASTA database was constructed of 107 complete and draft RefSeq genomes of all strains belonging to the bacterial order *Thermoanaerobacterales*. This database was searched for 5S rRNA-like sequences using RNAMotif software (Macke et al. 2001) with a search pattern that included three 5S rRNA fragments of fixed size corresponding to residues 3–11, 72–80, and 109–117, interspersed with two arbitrary sequences of variable length (50 to 180 nt and 20 to 130 nt, respectively). This search discovered a 172 nt-long 5S rRNA with an expansion segment of novel structural type in *Thermacetogenium phaeum*. A similar search of a database constructed from 257 genomic

sequences of all members of the archaeal class *Halobacteria* did not yield any new 5S rRNA sequences with insertions.

In light of these initial findings, we sought to conduct a more comprehensive search for expanded 5S rRNAs in prokaryotes. Initially, we searched two major collections of 5S rRNA sequences, the 5SrRNAdb database (Szymanski et al. 2016) and the RF00001 family in the Rfam database (Kalvari et al. 2018), for sequences with the length of 130 nt and above. In 5SrRNAdb, the search yielded only six sequences with the length of 130–132 nt, all of them from prokaryotes. None of these 5S rRNAs had an insertion larger than 2 nt. On the other hand, the RF00001 family of the Rfam database was found to contain 1524 RNA sequences within a size range from 130 to a maximum of 346 nt. After removal of sequences with unidentified nucleotides, the number of oversized 5S rRNAs decreased to 1474, with lengths ranging from 130 to 291 nt. Of these, only 34 examples were from

FIGURE 2. Predicted secondary structures of several representative 5S rRNAs with expansion segments. Nucleotides of the conserved core are highlighted yellow. (A) 5S rRNA from *Halococcus morrhuae* DSM 1307 (NZ_AOMC01000054.1, 66311–66541). (B) 5S rRNA from *Thermoanaerobacterium thermosaccharolyticum* DSM 571 (NC_014410.1, 84899–85053). (C) 5S rRNA from *Desulfotobacterium dichloroeliminans* LMG P-21439 (NC_019903.1, complement [2241573–2241816]). (D) 5S rRNA from *Thermacetogenium phaeum* DSM 12270 (NC_018870.1, complement [1414325–1414496]). (E) 5S rRNA from *Halobacteroides halobius* DSM 5150 (NC_019978.1, 20186–20360). (F) 5S rRNA from *Halanaerobium* sp. DL-01 (NZ_QPJN01000058.1, complement [9–185]). (G) 5S rRNA from *Lachnospiraceae* bacterium, strain COE1 (NZ_KE159617.1, 786025–786157). (H) 5S rRNA from *Pseudoalteromonas piscicida* DE2-B (NZ_CP021646.1, complement [2103409–2103538]). (I) 5S rRNA from *Desulfallus geothermicus* DSM 3669 (NZ_FOYM01000008.1, complement [126967–127097]).

prokaryotes (seven from archaea and 27 from bacteria) while the rest mostly represented the 5S rDNA tandems of eukaryotic origin. Each of the selected prokaryotic sequences was evaluated individually by querying the

NCBI nt database using BLAST, and by aligning it to the closest BLAST hits. This allowed us to identify and exclude from further consideration chimeric sequences and sequences with internal duplication. Of the 16 remaining

RNAs, seven were already known from previous searches, and nine were newly uncovered 5S rRNAs from the bacterial genera *Desulfallas* and *Desulfofarcimen*, both belonging to the family *Peptococcaceae*, and from an unclassified bacterial genus of the family *Lachnospiraceae*.

In addition to the 5S rRNA database searches, we retrieved all annotated 5S rRNA sequences longer than 129 nt from 1046 archaeal and 184,876 bacterial RefSeq genomes. In total, 159 oversized 5S rRNA sequences were collected. Of those, 110 sequences had no stretches of unidentified nucleotides, and therefore were retained for deeper analysis. The selected sequences were evaluated for validity as described above, and 52 sequences were discarded because of incorrect boundary assignment or internal sequence duplication. The remaining 58 sequences corresponded to valid 5S rRNAs with insertions. However, the majority of these were already known from the previous searches except for six oversized 5S rRNAs from the bacterial genera *Halanaerobium* and *Pseudoalteromonas*.

Finally, we inspected genomes of all organisms carrying oversized 5S rRNAs and their closest relatives for the presence of nonannotated 5S rRNA genes. A local BLAST database was constructed using 478 genomic DNA sequences from the bacterial orders *Clostridiales*, *Halanaerobiales*, *Thermoanaerobacterales*, and *Alteromonadales*, and from the archaeal order *Halobacteriales*. The database was queried with all normal and expanded 5S rRNA sequences available from annotations of the selected genomes. We specifically looked for cases where 5'- and 3'-terminal parts of a query made separate high-scoring hits to a genomic DNA at a distance compatible with the presence of an insertion in the 5S rRNA sequence. This approach allowed us to discover eight novel expanded 5S rRNAs in *Desulfotomaculum nigrificans*, *D. reducens*, *Desulfitobacterium dichloroeliminans*, and *Halobacteroides halobius* genomes. In the end, the list of the expanded 5S rRNAs contained 89 specimens of 15 structural types found in 15 archaeal and 36 bacterial strains. The number of expansion segments per 5S rRNA molecule varied from 1 to 3, and each of them ranged in length from 13 to 109 nt (Table 1; Supplemental Table S1; Supplemental File S1). In all cases, secondary structure prediction strongly suggested that the expansion segments have extensive internal base-pairing, which may allow them to exist as independent modules extending away from an unperturbed 5S rRNA core (Fig. 2A–I; Supplemental Figs. S1–S4).

In the course of these searches, several *Peptococcaceae* and *Pseudoalteromonadaceae* genomes were found to contain sequences that exhibit high similarity to 5S rRNA genes except they lack one or both of the 5S rRNA terminal parts and have large insertions in the 5S rRNA-like core (Supplemental File S2). They are usually located downstream from normal 5S rRNA genes at the distal ends of rRNA operons, sometimes in multiple copies, each of

which is preceded by a putative transcription terminator. These sequences are likely not transcribed. However, even if they are transcribed, the product RNA likely won't be processed and folded as a normal 5S rRNA because of the missing parts. Still, they retain enough of the 5S rRNA distinctive features to be identified as former 5S rRNA genes, and therefore may represent pseudogenes. Species that contain such remnants include *Desulfitobacterium metallireducens*, *Desulfosporosinus orientis*, *D. youngiae*, *Desulfotomaculum nigrificans*, *Desulfallas arcticus*, *Desulfitobacterium dichloroeliminans*, *Pseudoalteromonas piscicida*, *P. distincta*, *P. aliena*, *P. luteoviolacea*, *P. marina*, *P. phenolica*, *P. tetraodonis*, *P. flavipulchra*, and several unclassified strains from the same genera.

Interestingly, some of the insertions in these pseudogenes are nearly identical to the expansion segments of the oversized 5S rRNA genes in the cases when both types are present simultaneously in the same genome. Such a situation was observed in two *Peptococcaceae* strains (*Desulfitobacterium dichloroeliminans* LMG P-21439 and *Desulfotomaculum nigrificans* CO-1-SRB), and three *Pseudoalteromonadaceae* strains (*Pseudoalteromonas piscicida* strains DE2-A and DE2-B, and *P. flavipulchra* ATCC BAA-314). At the same time, several members of these families do not have the expanded 5S rRNA genes but still harbor their pseudogenized counterparts similar to those seen in *D. nigrificans* CO-1-SRB and *P. piscicida* DE2-A/DE2-B genomes, respectively. In two instances, both coming from the genus *Pseudoalteromonas*, the expansion segments embedded in a fragment of 5S rRNA sequence were found on a broad-host-range plasmid at the end of a plasmid-borne ribosomal RNA operon (Supplemental File S2). Association of these potential homologous recombination donors with mobile genetic elements might facilitate horizontal transfer of the expansion segments to 5S rRNA genes of closely related strains.

The observed resemblance between the obvious pseudogenes and expanded 5S rRNAs raises a question of whether the 5S rRNAs with expansion segments remain functional or have become inactive. There is little doubt about functionality of those expanded 5S rRNAs, which constitute the one and only 5S rRNA type present in a cell, since there is no alternative to them. Such a situation is observed in all species from the archaeal genera *Halococcus*, *Halalkalicoccus*, *Halomarina*, and *Halodesulfurarchaeum*, and in the bacterial strains *Thermacetogenium phaeum* DSM 12270, *Thermoanaerobacterium thermosaccharolyticum* TG57, *Thermoanaerobacter* sp. X513 and X514, *Halanaerobium* sp. DL-01 and MA284_MarDTE_T2, and unclassified *Lachnospiraceae* strains COE1 and A4. In all these cases, the expansion segments are expected to not interfere with the 5S rRNA function in the ribosome.

However, when a genome harbors both normal and expanded 5S rRNA genes, it cannot be excluded that the

TABLE 1. Prokaryotic strains harboring expanded 5S rRNAs

Species	Strain	Expanded 5S rRNA genes/ Total 5S rRNA genes	Expanded 5S rRNA length (nt)	Number of insertions	Insertion length (nt)	Insertion position
<i>Archaea; Euryarchaeota; Halobacteria; Halobacteriales</i>						
<i>Halococcus morrhuae</i>	DSM 1307	1/1	231	1	108	104.1–105
<i>Halococcus thailandensis</i>	JCM 13552	1/1	231	1	108	104.1–105
<i>Halococcus sediminicola</i>	CBA1101	1/1	231	1	108	104.1–105
<i>Halococcus hamelinensis</i>	100A6	1/1	231	1	108	104.1–105
<i>Halococcus salifodinae</i>	DSM 8989	1/1	232	1	108	104.1–105
<i>Halococcus saccharolyticus</i>	DSM 5350	1/1	232	1	108	104.1–105
<i>Halococcus agarilyticus</i>	197A	1/1	232	1	108	104.1–105
<i>Halococcus</i> sp.	IIIV-5B	1/1	231	1	108	104.1–105
<i>Halalkalicoccus jeotgali</i>	B3	1/1	231	1	108	104.1–105
<i>Halalkalicoccus paucihalophilus</i>	DSM 24557	1/1	231	1	108	104.1–105
<i>Halalkalicoccus</i> sp.	GSM28	1/1	231	1	108	104.1–105
<i>Halomarina oriensis</i>	SPP-AMP-1	1/1	228	1	105	104.1–105
<i>Halomarina oriensis</i>	JCM 16495	1/1	228	1	105	104.1–105
<i>Halodesulfurarchaeum formicum</i>	HSR6	1/1	232	1	109	104.1–105
<i>Halodesulfurarchaeum formicum</i>	HTSR1	1/1	232	1	109	104.1–105
<i>Bacteria; Firmicutes; Clostridia; Thermoanaerobacterales</i>						
<i>Thermacetogenium phaeum</i>	DSM 12270	3/3	172	1	61	36–37
<i>Thermoanaerobacterium thermosaccharolyticum</i>	TG57	5/5	155	1	46	52–53
<i>Thermoanaerobacterium thermosaccharolyticum</i>	DSM 571	3/5	155	1	46	52–53
<i>Thermoanaerobacterium thermosaccharolyticum</i>	M0795	4/5	155	1	46	52–53
<i>Thermoanaerobacterium thermosaccharolyticum</i>	GSU5	2/?	155	1	46	52–53
<i>Thermoanaerobacterium thermosaccharolyticum</i>	M5	1/?	?	1	?	52–53
<i>Thermoanaerobacterium thermosaccharolyticum</i>	SP-H2	1/?	155	1	46	52–53
<i>Thermoanaerobacterium saccharolyticum</i>	JW/SL-YS485	2/5	155	1	46	52–53
<i>Thermoanaerobacterium saccharolyticum</i>	NTOU1	3/5	155	1	46	52–53
<i>Thermoanaerobacterium aotearoense</i>	SCUT27	2/5	155	1	46	52–53
<i>Thermoanaerobacter pseudethanolicus</i>	ATCC 33223	2/4	152	1	43	52–53
<i>Thermoanaerobacter ethanolicus</i>	CCSD1	1/?	152	1	43	52–53
<i>Thermoanaerobacter italicus</i>	Ab9	1/4	152	1	43	52–53
<i>Thermoanaerobacter brockii</i> subsp. <i>finii</i>	Ako-1	2/4	152	1	43	52–53
<i>Thermoanaerobacter kivui</i>	DSM 2030	3/5	152	1	43	52–53
<i>Thermoanaerobacter</i> sp.	X513	4/4	152	1	43	52–53
<i>Thermoanaerobacter</i> sp.	X514	4/4	152	1	43	52–53
<i>Thermoanaerobacter</i> sp.	X561	3/4	152	1	43	52–53

(Continued)

TABLE 1. (Continued)

Species	Strain	Expanded 5S rRNA genes/ Total 5S rRNA genes	Expanded 5S rRNA length (nt)	Number of insertions	Insertion length (nt)	Insertion position
<i>Thermoanaerobacter</i> sp.	UBA8867	1/?	152	1	43	52–53
<i>Caldanaerobacter subterraneus</i>	DSM 13054	1/?	152	1	43	52–53
<i>Caldanaerobius polysaccharolyticus</i>	DSM 13641	1/?	150	1	41	52–53
<i>Bacteria; Firmicutes; Clostridia; Halanaerobiales</i>						
<i>Halobacteroides halobius</i>	DSM 5150	4/5	175	1	58	36–37
<i>Halanaerobium</i> sp.	MA284_MarDTE_T2	1/?	177	1	60	36–37
<i>Halanaerobium</i> sp.	DL-01	1/?	177	1	60	36–37
<i>Bacteria; Firmicutes; Clostridia; Clostridiales</i>						
<i>Desulfotomaculum acetoxidans</i>	DSM 771	2/9	147; 166	1; 2	32; 19 + 32	29–30, 42–43
<i>Desulfallus geothermicus</i>	DSM 3669	1/5	131	1	19	69–70
<i>Desulfallus gibsoniae</i>	DSM 7213	3/9	153; 212; 262	2; 2; 3	19 + 20; 39 + 57; 23 + 39 + 85	42–43, 67–68, 96–98
<i>Desulfotomaculum nigrificans</i>	CO-1-SRB	1/9	206	1	89	57–58
<i>Desulfotomaculum reducens</i>	MI-1	2/10	205	2	13 + 78	59–63, 66–67
<i>Desulfotomaculum dichloroeliminans</i>	LMG P-21439	1/6	244	2	64 + 69	89.1–93, 105–109
<i>Lachnospiraceae bacterium</i>	COE1	4/?	133	1	15	34–35
<i>Lachnospiraceae bacterium</i>	A4	1/?	133	1	15	34–35
<i>Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales</i>						
<i>Pseudoalteromonas piscicida</i>	DE2-A	1/13	130	1	16	68–69
<i>Pseudoalteromonas piscicida</i>	DE2-B	1/13	130	1	16	68–69
<i>Pseudoalteromonas flavipulchra</i>	NCIMB 2033	1/10	131	1	16	68–69
<i>Pseudoalteromonas distincta</i>	ATCC 700518	1/10	131	1	16	68–69

Question mark (?) indicates a lack of precise information about 5S rRNA genes due to incompleteness of genome assemblies.

expanded 5S rRNAs are either not expressed or inactive, and only the normal 5S rRNAs are associated with translating ribosomes. In such cases, we sought to assess the functional status of the expanded 5S rRNAs by comparing the immediate genomic context of their coding sequences with that of either normal 5S rRNA genes from the same genome or oversized 5S rRNA genes from closely related organisms harboring only 5S rRNAs of the expanded type. The comparison was performed under the assumption that a loss of function would correlate with accumulation of mutations in the conserved intergenic sites involved in 5S rRNA synthesis and maturation since they would no longer remain under selective pressure. It was found that the majority of the oversized 5S rRNA genes in question can be related to at least one reference gene deemed to produce functional 5S rRNA and embedded within the

same genetic context. The strongest indications of functional competence were noted for the expanded 5S rRNAs from the bacterial genera *Thermoanaerobacterium*, *Thermoanaerobacter*, and *Halobacteroides*. However, in several other cases, significant deviations from the conserved layout of prokaryotic 5S rRNA expression units were observed. The most remarkable example is an oversized 5S rRNA gene from *Desulfallus gibsoniae* DSM 7213 (genome accession number NC_021184.1) located on the forward strand at positions 1,812,738 to 1,812,949. It is not a part of an rRNA operon. Instead, it is a stand-alone gene situated between a phosphate transporter pseudogene (DESGI_RS08420) and a sodium-dependent transporter gene with indications of horizontal transfer from a distant species (DESGI_RS08425). While the 5S rRNA coding sequence is preceded by a 107-bp-long region typical for

regular 5S rRNA genes in *D. gibsoniae*, it is followed by a transcription terminator, which is not related to any rRNA operon but rather inherited from a protein expression unit similar to those at remote loci DESGI_RS19955 and DESGI_RS20640. Putative promoters are located within the adjacent pseudogene and show little similarity to the promoter groups controlling the expression of rRNA operons. Overall, this 5S rRNA gene looks like a displaced piece of DNA with accidentally acquired transcriptional control elements. As such, it may or may not produce functional 5S rRNA depending on the promoter activity and compatibility of the primary transcript with *D. gibsoniae* 5S rRNA maturation machinery.

Other examples of the oversized 5S rRNA genes with abnormal surroundings include those from *Desulfotomaculum reducens* MI-1 (NC_009253.1, positions 361,111 to 361,315 on the forward strand) embedded in a severely damaged rRNA operon with part of its promoter group lost, *Desulfallas gibsoniae* DSM 7213 (NC_021184.1, positions 1,557,019 to 1,557,280 on the forward strand) with unusual transcription terminator, and *Pseudoalteromonas piscicida* strains DE2-A and DE2-B with multiple substitutions in the promoter group of the corresponding rRNA operon. Also, the oversized 5S rRNA genes from *Pseudoalteromonas flavipulchra* NCIMB 2033, *P. distincta* ATCC 700518, *Caldanaerobacter subterraneus* DSM 13054, and *Caldanaerobius polysaccharolyticus* DSM 13641 were not evaluated due to insufficient information for the comparative sequence analysis. For all these problematic cases, the functional state of 5S rRNAs would likely need to be assessed through experimental studies.

It is worth mentioning that the insertional heterogeneity observed in some genomes among 5S rRNA gene copies can also be seen in 16S and 23S rRNA gene families when the expanded rRNA genes coexist with the normal ones (Supplemental File S3). At the same time, there is no strict correlation between the distribution of large (>10 bp) insertions in 5S, 16S, and 23S rRNA coding sequences in a single genome with multiple rRNA operons. Of the prokaryotic strains harboring the oversized 5S rRNAs, insertional heterogeneity was detected in *Desulfitobacterium dichloroeliminans* LMG P-21439 5S, 16S, and 23S rRNAs, *Desulfofarcimen acetoxidans* DSM 771 5S and 23S rRNAs (but not in 16S rRNAs), *Desulfallas gibsoniae* DSM 7213 5S, 16S, and 23S rRNAs, *Thermoanaerobacterium thermosaccharolyticum* DSM 571 5S, 16S, and 23S rRNAs, *Th. thermosaccharolyticum* TG57 23S rRNAs (but not in 5S and 16S rRNAs), *Th. thermosaccharolyticum* M0795 5S and 23S rRNAs (but not in 16S rRNAs), *Th. aotearoense* SCUT27 5S and 16S rRNAs (but not in 23S rRNAs), *Thermoanaerobacter brockii* subsp. *finnii* Ako-1 5S, 16S, and 23S rRNAs, *Th. pseudethanolicus* ATCC 33223 5S, 16S, and 23S rRNAs, and *Th. italicus* Ab9 5S and 23S rRNAs (but not in 16S rRNAs). It seems likely that the mechanisms underlying the insertion of novel

modules into the 5S rRNA chain also act on large rRNAs of the mentioned organisms. At the same time, sequences and distribution of the expansion segments in the rRNA gene families point to rather independent acquisition of such modules by 5S, 16S, and 23S rRNA genes.

DISCUSSION

Our search for the expansion segments in prokaryotic 5S rRNAs revealed 87 novel examples of the expanded 5S rRNA molecules in addition to the previously known two. The Rfam RF00001 and 5SrRNAdb databases and the entire RefSeq collection of annotated bacterial and archaeal genomes have been exhaustively screened for the oversized 5S rRNA molecules. It should be noted that the outcome directly depended on the quality of the genome annotations with respect to 5S rRNA gene mapping, and on the content curation practices in the case of 5S rRNA databases. Once identified, the sequences of the expanded 5S rRNAs were used as queries for a BLAST similarity search in the NCBI *nt* nucleotide sequence database and whole-genome collections. This approach allowed us to discover expanded 5S rRNAs in non- or poorly annotated genomes of species closely related to those from which the queries were derived. Another annotation-independent search method utilized 5S rRNA sequence models, which were constructed of short, conserved motifs interspersed with random sequences of variable size, and partly constrained with characteristic secondary structure patterns. It was used to screen only a small fraction of prokaryotic genomes because of the high rate of false-positive detections inevitably associated with high redundancy of the search probes. Overall, the most comprehensive search has been performed in the bacterial order *Thermoanaerobacteriales* and archaeal class *Halobacteria*, where the initial screening uncovered the majority of novel 5S rRNAs with expansion segments. At the same time, we cannot exclude that non- or poorly annotated genomes outside of these taxonomic groups may still harbor some expanded 5S rRNAs that escaped detection.

One may wonder why the expanded 5S rRNAs were largely overlooked up until now, given that the first two examples had been discovered about four decades ago. One explanation is that a substantial number of them have been incorrectly presented as much shorter sequences in genomic annotations and databases in order to make them better fit the consensus-derived standard model of 5S rRNA. For example, all genes encoding the expanded 5S rRNAs in sequenced archaeal genomes have been found misannotated with respect to their boundaries. In all these cases, a fragment of the expansion segment with partial complementarity to the 5'-terminus of 5S rRNA has been misinterpreted as a 3'-terminal sequence in order to keep 5S rRNA length within standard range. However, even when correctly mapped in genomic

sequences, the expanded 5S rRNAs may remain not deposited in databases and thus evade public attention. The most prominent collection of 5S rRNA sequences, 5SrRNAdb, contains only molecules not longer than 132 residues, possibly because of strict curation practices, which strongly disfavor deposition of anomalous specimens. As a result, the database hosts only those expanded 5S rRNAs, which were erroneously annotated as short enough to match the standard 5S rRNA model. Another specialized database, Rfam RF00001, does not apparently set a rigid size limit for the deposited sequences, and therefore contains several expanded 5S rRNAs with correctly identified boundaries. However, these entries are obscured by numerous artefactual oversized sequences, which originate from sequencing, assembly, and annotation errors.

Interestingly, it has been a chance for some of the expanded 5S rRNAs to be spotted during a study on intragenomic variation of multicopy 5S rRNA genes in prokaryotes performed in 2012 by Pei et al. (Pei et al. 2012). Since in a number of cases the expanded 5S rRNAs coexist with the normals in the same organism, direct comparison of their sequences would immediately expose the presence of expansion segments in the longer molecules. Among the examined genomes, the study mentions those of *Thermoanaerobacter pseudethanolicus*, *Thermoanaerobacter italicus*, *Desulfotomaculum reducens*, and *Desulfotomaculum acetoxidans* (now reclassified as *Desulfofarcimen acetoxidans*), which, as we now know, encode both normal and extended 5S rRNA types. For unclear reasons, the variation between *Th. italicus* 5S rRNA gene copies was not noticed at all, while in the genomes of *Th. pseudethanolicus*, *D. reducens*, and *D. acetoxidans*, the observed differences have been interpreted exclusively in terms of point mutations and single-nucleotide indels. At some point, the authors specifically claim that all the studied 5S rRNA genes have been found to have no “intervening sequences,” that is, long insertions, as opposed to 16S and 23S rRNA genes. We think that such a conclusion could only be reached if incomplete sequences of the expanded 5S rRNA genes have been used in the study.

In both prokaryotic superkingdoms, the expansion segments are present in only a tiny fraction of 5S rRNAs. At the same time, patterns of their taxonomic distribution within each superkingdom are markedly different. In *Archaea*, the expanded 5S rRNAs are predominantly found within two closely related genera, *Halococcus* and *Halalkalicoccus*, which together form a compact cluster well separated from other members of the family *Halobacteriaceae* (Fig. 3). In every species in this cluster, 5S rRNA has a 108-nt insertion between positions 104.1 and 105 of the archaeal core. The only other clades that harbor the expanded 5S rRNAs are two small genera, *Halodesulfurarchaeum* and *Halomarina*. The recently proposed single-species genus *Halodesulfurarchaeum* is rather remote from the *Halococ-*

cus-Halalkalicoccus branch, and grouped with genera *Halobacterium* and *Halarchaeum*, which both are devoid of the expanded 5S rRNAs. Similarly, the genus *Halomarina* belongs to a subfamily whose other members carry only normal 5S rRNA genes. Since all the expansion segments in the oversized 5S rRNAs of the *Halobacteriaceae* family members belong to the same structural type, it is natural to suggest a horizontal transfer of the segment-coding sequence between the *Halococcus-Halalkalicoccus* cluster, *Halomarina*, and *Halodesulfurarchaeum* genera.

In bacteria, the expansion segments are mostly found in 5S rRNAs of clostridia belonging to the orders *Thermoanaerobacterales*, *Clostridiales*, and *Halanaerobiales*. The only nonclostridial taxon harboring the expanded 5S rRNAs is the gammaproteobacterial genus *Pseudoalteromonas*. The distribution of expanded 5S rRNAs across bacterial taxa is intermittent. This is further complicated by high structural diversity of the expansion modules, which implies their independent origin. It is common for the bacterial species carrying expanded 5S rRNAs to be found together with those with only normal 5S rRNAs within the same genus. Sometimes, even strains belonging to the same species have a different representation of normal and oversized 5S rRNA types as in the case of the *Pseudoalteromonas piscicida* branch. However, the variations in distribution of the expanded 5S rRNA genes seen at subspecies level are not always genuine. They may be caused by strain misassignment as apparently happened with the *Pseudoalteromonas flavipulchra* and *Thermoanaerobacter ethanolicus* lineages.

The most prominent bacterial group associated with the expanded type of 5S rRNA is a cluster of four interconnected thermoanaerobacterial genera, *Thermoanaerobacterium*, *Thermoanaerobacter*, *Caldanaerobacter*, and *Caldanaerobius* (Fig. 4). The group does not look entirely coherent since several branches within it have apparently lost the oversized 5S rRNA genes. All expansion segments within the group belong to the same structural type, which is represented by a 41–46-nt insertion rooted in the universally conserved bulge at the positions 52 and 53. Alignment of the inserted sequences reveals three distinctive subtypes strictly correlated with phylogenetic branching (Supplemental Fig. 5). One of the subtypes, which originates from the genus *Caldanaerobius*, is apparently ancestral to two others associated with the genus *Thermoanaerobacterium*, and the group *Thermoanaerobacter-Caldanaerobacter*, respectively.

In other cases, the expanded 5S rRNAs are scattered through bacterial taxa as isolated single strains or small clusters consisting of several closely related strains (Figs. 5–8). At the above-species levels, the expansion sequences may propagate via horizontal transfer apparently mediated by mobile genetic elements. One such a long-range horizontal link connects the *Halobacteroides halobius* DSM 5150 strain from the family *Halobacteroidaceae*

FIGURE 3. Phylogenetic tree of the archaeal order *Halobacteriales*. Branches associated with the strains carrying the expanded 5S rRNA genes are highlighted yellow. Branch support values represent branch recovery percentage in 100 jackknife resampling replicates. Scale bar represents the number of nucleotide substitutions per position. The tree was generated using the maximum likelihood algorithm on a 106,995-bp-long concatenated codon-wise alignment of 100 conserved single-copy protein-coding genes shared across all of the 50 strains selected for tree construction. The tree was rooted using *Haloferax elongans* ATCC BAA-1513 as an outgroup.

with a group of two unclassified *Halanaerobium* strains, DL-01 and MA284_MarDTE_T2, from the family *Halanaerobiaceae* (Fig. 6). Another horizontal transfer of the expansion sequence seemingly happened between a group of two *Pseudoalteromonas piscicida* strains, DE2-A and DE2-B, and a group consisting of *P. distincta* ATCC 700518 and *P. flavipulchra* ATCC BAA-314 strains (Fig. 7). It should be noted, however, that multiple occurrences of 5S rRNA remnants with the same insertion as in expanded 5S rRNAs hinder precise identification of a donor of the transferred sequence among members of the family *Pseudoalteromonadaceae*. In addition to the horizontal route, acquisition of the expansion segments might also proceed through vertical inheritance of the pseudogenes followed by their homologous recombination with normal 5S rRNA genes. Similar considerations can be applied to the distribution pattern of the expanded 5S rRNAs in the family *Peptococcaceae*, whose members also occasionally harbor the 5S rRNA gene remnants with insertions.

Unlike their counterparts in eukaryotic 18S and 25S/28S rRNAs, the expansion segments in prokaryotic 5S rRNAs do not follow any recognizable trend that correlates with

phylogenetic branching order across the bacterial and archaeal superkingdoms. Instead, they are mostly divided in taxonomically remote and structurally unrelated groups, each of which represents a rather uniform set of sequences. The noncontiguous taxonomic distribution alongside with rare occurrence point to a nonuniversal character of their adaptive value, if any exists. Therefore, even if an expansion of 5S rRNA contributes to the fitness of the host organism, this contribution is likely restricted to the conditions uncommon for the majority of prokaryotes. Otherwise it would be difficult to explain why the expanded 5S rRNAs have not spread through the entire prokaryotic realm but instead remain confined within boundaries of several particular taxa of subfamily rank.

Analysis of environmental preferences of the prokaryotic hosts shows a correlation between the presence of expanded 5S rRNA genes and microaerobic/anaerobic lifestyle accompanied with strong halophilic or thermophilic traits. Thus, the insertions in 5S rRNAs might, for example, be associated with peculiar metabolic pathways involved in niche-specific adaptation to hypoxic saline or thermal habitats. At the same time, we cannot completely rule out a possibility that some of the expansion segments do

FIGURE 4. Partial phylogenetic tree of the bacterial order *Thermoanaerobacterales* represented by the related members of families *Thermoanaerobacteraceae*, *Thermoanaerobacterales* Family III and Family IV *incertae sedis*. Branches associated with the strains carrying the expanded 5S rRNA genes are highlighted yellow. The inset shows a scaled-up fragment of the tree. Branch support values represent branch recovery percentage in 100 jackknife resampling replicates. Scale bar represents the number of nucleotide substitutions per position. The tree was generated using the maximum likelihood algorithm on an 80,655-bp-long concatenated codon-wise alignment of 95 conserved single-copy protein-coding genes shared across all of the 37 strains selected for tree construction. The tree was rooted using *Anoxybacter fermentans* DY22613 as an outgroup.

not confer any adaptive benefit for the hosts, and instead represent accidentally acquired features undergoing neutral evolution. While this may not seem likely in the case of relatively large and genetically cohesive taxonomic clusters, such as *Halococcus-Halalkalicoccus* and *Thermoanaerobacterium-Thermoanaerobacter-Caldanaerobacter-Caldanaerobius* groups, it is certainly possible for the small and incoherent ones, like the set of strains from the family *Peptococcaceae* (Fig. 8).

The observation that bacterial genomes can host both normal and expanded 5S rRNA genes leads to interesting consequences in the form of heterogeneity of the translation apparatus. The two types of 5S rRNA must confer different operational specializations on the ribosome in order to simultaneously persist within the same organism over an evolutionary significant period of time. Thus, the ribosomal population would consist of two subpopulations optimized for different tasks. In the cases when 16S and 23S rRNAs are also represented by several different structural

types, the number of possible rRNA combinations would increase accordingly, giving rise to an even more diversified ribosomal pool. Such a structural heterogeneity of the translation apparatus is anticipated to increase its functional flexibility. It would allow cells to finely adjust the translation efficiency and accuracy to current metabolic requirements through differential expression of rRNA genes of various types. Up until now, the concept of ribosomal heterogeneity has mostly been associated with variations in protein composition of eukaryotic ribosomes (Byrgazov et al. 2013; Sauert et al. 2015; Shi et al. 2017; Genuth and Barna 2018a,b; Guo 2018; Ferretti and Karbstein 2019; López García de Lomana et al. 2020). Only a few reports addressed the issue of rRNA variability as a source of diversity in an intracellular ribosome pool. The latter was primarily attributed to single-nucleotide polymorphisms (López-López et al. 2007; Sato et al. 2017; Kurylo et al. 2018; Parks et al. 2018; Song et al. 2019) and differential patterns of posttranscriptional

FIGURE 5. Phylogenetic tree of the bacterial family *Lachnospiraceae*. Branches associated with the strains carrying the expanded 5S rRNA genes are highlighted yellow. Branch support values represent branch recovery percentage in 100 jackknife resampling replicates. Scale bar represents the number of nucleotide substitutions per position. The tree was generated using the maximum likelihood algorithm on a 98,322-bp-long concatenated codon-wise alignment of 100 conserved single-copy protein-coding genes shared across all of the 37 strains selected for tree construction. The tree was rooted using *Thermacetogenium phaeum* DSM 12270 as an outgroup.

FIGURE 6. Phylogenetic tree of the bacterial order *Halanaerobiales*. Branches associated with the strains carrying the expanded 5S rRNA genes are highlighted yellow. Branch support values represent branch recovery percentage in 100 jackknife resampling replicates. Scale bar represents the number of nucleotide substitutions per position. The tree was generated using the maximum likelihood algorithm on a 29,271-bp-long concatenated codon-wise alignment of 27 conserved single-copy protein-coding genes shared across all of the 30 strains selected for tree construction. The tree was rooted using *Thermacetogenium phaeum* DSM 12270 as an outgroup.

FIGURE 7. Phylogenetic tree of the bacterial genus *Pseudoalteromonas*. Branches associated with the strains carrying the expanded 5S rRNA genes are highlighted yellow. Branch support values represent branch recovery percentage in 100 jackknife resampling replicates. Scale bar represents the number of nucleotide substitutions per position. The tree was generated using the maximum likelihood algorithm on a 126,927-bp-long concatenated codon-wise alignment of 100 conserved single-copy protein-coding genes shared across all of the 38 strains selected for tree construction. The tree was rooted using *Alvicola sagamiensis* DSM 14643 as an outgroup.

rRNA modification (Marcel et al. 2015; Babaian et al. 2020). The present study implies that in certain prokaryotes the ribosomal heterogeneity can also be caused by the combinatorial assembly of the ribosomes from a plethora of normal and expanded rRNA types.

The distribution of expansion segments across the elements of the prokaryotic 5S rRNA secondary structure model is expectedly uneven (Fig. 1). The insertion sites are concentrated in the loop B—helix III—loop C fragment, in the vicinity of the triple helical junction, and in the 3'-terminal half of the helix IV. The inserted modules are anticipated to not interfere with folding of the 5S rRNA core because of their extensive internal base-pairing and lack of noticeable complementarity to the core sequence. None of them interrupts conserved sequence motifs except the one in oversized *Desulfotomaculum nigrificans* CO-1-SRB 5S rRNA. Also, the insertions rarely displace bases of the core or disrupt tertiary interactions, which contribute to maintenance of the core structure. At the same time, not all of the insertions seem compatible with the immediate surroundings of 5S rRNA on the ribosome. In archaeal 5S rRNAs, the sole identified insertion site (pos. 104.1–105) allows unhindered sequence expansion away from other ribosomal constituents (Fig. 9A). In bacteria, the in-

sertion site at positions 29–30 overlaps with the binding site of ribosomal protein uL18 while positions 42–43 and 57–58 are blocked by protein uL5. In addition, substantial steric constraints disfavor a large sequence expansion at positions 68–70 and 89.1–93 whereas new modules rooted at positions 66–68 and 96–98 would encounter significantly milder spatial restrictions. The least obstructed insertion sites in bacterial 5S rRNAs are located at positions 34–37 in loop C, and positions 52–53 corresponding to the universally conserved bulge in helix III (Fig. 9B).

Taking these considerations into account, one might conclude that the expanded 5S rRNAs from *Desulfofarcimen acetoxidans*, *Desulfallas gibsoniae*, and *Desulfotomaculum nigrificans* cannot properly bind to the large ribosomal subunit because their interactions with associated ribosomal proteins uL5 and uL18 are compromised. Unless the expansion segments can functionally substitute for uL5 and uL18, these 5S rRNAs are unlikely to correctly assemble with other ribosomal components to produce a functional ribosome. Likewise, a 69-nt insertion in the helix IV of the expanded 5S rRNA from *Desulfobacterium dichloroeliminans* may prevent binding of the whole molecule to the ribosome due to steric constraints at the helix IV docking cavity. In all these cases, the expansion of the

FIGURE 8. Phylogenetic tree of the bacterial family *Peptococcaceae*. Branches associated with the strains carrying the expanded 5S rRNA genes are highlighted yellow. Branch support values represent branch recovery percentage in 100 jackknife resampling replicates. Scale bar represents the number of nucleotide substitutions per position. The tree was generated using the maximum likelihood algorithm on a 29,898-bp-long concatenated codon-wise alignment of 49 conserved single-copy protein-coding genes shared across all of the 45 strains selected for tree construction. The tree was rooted using *Thermacetogenium phaeum* DSM 12270 as an outgroup.

5S rRNA chain possibly represents the initial stage of pseudogenization or a shift to a novel extraribosomal function.

The most conspicuous types of expansion segments are predictably associated with the three most exposed insertion sites. In halobacterial 5S rRNAs, the expansion segment is located between the “loop E” and triple helical junction and expected to extend outwards of the ribosome. It is shaped as an asymmetric “V” with the shorter arm folded as a hairpin, and longer arm as an unbranched protrusion consisting of several helical stems interspersed with internal loops (Fig. 2A; Supplemental Fig. S6). The distal stem-loop element is the most conserved part of the insertion, and therefore a prime candidate for a putative functional domain.

Expansion of thermoanaerobacterial 5S rRNAs predominantly occurs at the conserved bulge in helix III in a form of two hairpins of unequal size (Fig. 2B; Supplemental Fig. S7). The entire fragment is situated at the very top of the central protuberance. The longer hairpin is homoiteron-rich, which is a feature often seen in the expansion segments of eukaryotic rRNAs (Parker et al. 2015). Interestingly, similar motifs are present in several prokary-

otic genomes other than thermoanaerobacterial. They reside mostly in intergenic regions, including the leaders and spacers in rRNA operons. Their possible functions inferred from their locations include transcription control and RNA processing. It is important to note that in the thermoanaerobacterial genomes, such motifs are found only in the expanded 5S rRNA genes.

The insertion sites at positions 34–37 in the loop C harbor expansion segments in 5S rRNAs from several distant clostridial strains belonging to the families *Thermoanaerobacteraceae*, *Halobacteroidaceae*, *Halanaerobiaceae*, and *Lachnospiraceae*. Apart from the latter case, the inserted modules represent long unbranched helices with bulges and internal loops. In this respect, they bear a superficial resemblance to the long arm of the expansion segment from halobacterial 5S rRNAs. Despite being in close proximity to the binding sites of proteins uL5 and uL18, the inserted helices have enough space for accommodation without any interference with the nearby ribosomal constituents.

Summarizing our findings, we would like to note that intrinsic limitations of bioinformatic methods do not allow us

FIGURE 9. Apical part of large ribosomal subunit viewed from the solvent side. Ribosomal proteins are presented in space-filling mode; rRNAs are shown as pipes. 5S rRNA is colored blue except for the rooting sites of the most exposed expansion segments, which are shown in red. 23S rRNA is colored cyan. (A) Archaeal 50S subunit (*Haloarcula marismortui*, PDB 4V9F). (B) Bacterial 50S subunit (*Escherichia coli*, PDB 4YBB).

to draw definitive conclusions about the true nature of the insertions in 5S rRNA coding sequences without supplementing computational genome analysis with appropriate experimental data. Even if the extended 5S rRNA genes are properly expressed, the insertions might not be retained in the primary transcripts in the form of expansion segments but excised as it happens with rRNA introns and intervening sequences (Evguenieva-Hackenberg 2005; Rogers 2019). The most solidly confirmed expansion segment is by far the one found in *H. morrhuae* 5S rRNA. This insertion was reproducibly observed as an integral part of the only 5S rRNA type present in total RNA pools of several *H. morrhuae* strains (Luehrsen et al. 1981; Nicholson 1982). It is also visible in a 6.4 Å cryo-EM reconstruction of the *H. morrhuae* ribosome as a density bulge protruding out from the 5S rRNA core (Tirumalai et al. 2020). Another case of valid 5S rRNA expansion is represented by the oversized 5S rRNA species isolated from the large subunit of *Th. thermosaccharolyticum* ribosome (Sutton and Woese 1975). Although missing in the original publication, the complete 5S rRNA sequence can be easily reconstructed from the experimental data using its gene sequence as a reference. In both these cases, the insertion seen in the 5S rRNA gene does not prevent

the latter from being expressed into mature 5S rRNA, which then harbor an additional ribonucleotide stretch inserted into conserved core. The inserted fragments are apparently retained by the oversized 5S rRNAs throughout their life cycle, and do not interfere with 5S rRNA integration into the ribosome. Therefore, they can confidently be regarded as expansion segments rather than introns or intervening sequences.

By analogy with the expanded 5S rRNA genes from *H. morrhuae* and *Th. thermosaccharolyticum*, their direct homologs in halobacterial and thermoanaerobacterial genomes, respectively, are naturally anticipated to produce oversized 5S rRNAs with nonexcisable insertions, which are compatible with ribosomal architecture. In all other cases, however, the term “expansion segment” should be used with caution when describing the expression products from 5S rRNA genes with insertions. For members of the bacterial families *Halanaerobiaceae*, *Halobacteroidaceae*, and *Lachnospiraceae*, there is no reason to doubt that such genes can be expressed into mature 5S rRNAs but it remains to be proven that the inserted sequences are not removed from these 5S rRNAs at some point of their life cycle. The situation with the expanded 5S rRNA genes in strains belonging to the families *Peptococcaceae* and *Pseudoalteromonadaceae* is somewhat more obscure because frequent abnormalities in the corresponding rRNA operons make us question if they can produce functional rRNA in the first place. In the case of several *Peptococcaceae* family members, these concerns are aggravated by anticipated steric incompatibility of the expanded 5S rRNAs with the established ribosomal framework, which among other things implies a possibility of excision of the inserted fragments prior ribosome assembly. Hopefully, the forthcoming experimental studies will resolve these uncertainties.

Seemingly unhindered accommodation of the long helical fragments inserted at positions 34–37, 52–53, and 104.1–105 of the conserved 5S rRNA core highlights a potential value of these sites for biotechnological applications. A regular prokaryotic 5S rRNA could be modified to harbor artificial RNA modules at these locations in order to confer novel properties and functions on the ribosome. The modules may carry aptamers, ribozymes, riboswitches, antisense sequences, purification tags and other RNAs of interest. Genetically modified 5S rRNA might be supplemented with altered 16S and 23S rRNAs if several artificial modules must be introduced into a ribosomal particle. Mobile expression constructs with engineered 5S rRNA genes might also be used to create and control multiple ribosomal subpopulations in a cell with its own set of natural 5S rRNA genes.

While being commonly regarded as a molecule with a highly conserved structure, 5S rRNA can apparently accommodate relatively large expansion segments comparable in size with its core. This finding gives us a better

understanding of structural constraints imposed on 5S rRNA by the surrounding ribosomal framework. It also provides us with new insight into admissible variations in the prokaryotic translation apparatus. The acquired knowledge could be useful in studies on function and evolution of 5S rRNA, and in ribosome redesign for biotechnological applications.

MATERIALS AND METHODS

Databases

The nucleotide *nt* database was downloaded from the National Center for Biotechnology Information (NCBI) FTP site <ftp://ftp.ncbi.nlm.nih.gov/blast/db/> in April 2020 and used for local BLAST searches. Complete and draft RefSeq assemblies of all available archaeal and bacterial genomes were downloaded from the NCBI FTP site <ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/> on April 25, 2018, using the corresponding assembly_summary.txt files as a source of ftp paths to the assembly directories. A homemade Perl script was used to retrieve 5S rRNA sequences with minimum size limit of 130 nt from the [accession number]_[assembly name]_rna_from_genomic.fna.gz files. When necessary, local BLAST or FASTA databases were created from genomic DNA sequences of the selected assemblies. 5SrRNAdb (Szymanski et al. 2016) and Rfam RF00001 (Kalvari et al. 2018) collections of 5S rRNA sequences were last accessed on May 4, 2020, at <http://combio.pl/rna/> and <http://rfam.xfam.org/family/RF00001/>, respectively.

Sequence analysis

Searches of the genomic nucleotide databases for 5S rRNA-coding sequences were performed using the Basic Local Alignment Search Tool (BLAST) 2.7.1+ (Camacho et al. 2009). Handling and analysis of short DNA/RNA sequences was managed using Seaview 4.6.1 software package (Gouy et al. 2010). Annotated genomes were explored using the Artemis genome viewer 17.0.1 (Rutherford et al. 2000; Carver et al. 2008).

A 5S rRNA pattern matching search of the *Thermoanaerobacterales* and *Halobacteria* genomes was performed using RNAMotif 3.1.1 (Macke et al. 2001). Descriptors for the 5S rRNA-like secondary structure pattern were constructed to allow large insertions anywhere in the structure except helix I and the upstream half of "loop E." In the case of *Thermoanaerobacterales*, the search pattern was $YYYGGYGRN_{50-180}GAUGRUASUN_{20-150}.YYGCCRRR$, with the 5'-terminal fragment $YYYGGYGR$ required to be complementary to the 3'-terminal fragment $YYGCCRRR$. In the case of *Halobacteria*, the search pattern was $MGGCGGCCN_{50-180}.BNNMSUACUN_{20-150}TCGCCGCC$, where the 5'-terminal fragment $GGCGGC$ was base-paired with the 3'-terminal fragment $GCGGCC$.

The identified 5S rRNA expansion segments were folded as self-sufficient structural modules using an energy-based algorithm from RNAstructure 6.0 (Xu and Mathews 2016). RNA secondary structure prediction driven by both energy constraints and sequence covariation data was performed using R-scape 0.3.3 (Rivas et al. 2017) and the RNAalifold program from the

ViennaRNA 2.2.5 software package (Hofacker and Lorenz 2014; Lorenz et al. 2016). 5S rRNA secondary structures were visualized using RNAviz 2.0.3 (De Rijk et al. 2003).

Phylogenetic tree construction

Phylogenetic trees were generated using tools provided at the Pathosystems Resource Integration Center (PATRIC) website, <https://patricbrc.org> (Wattam et al. 2014, 2017). The trees were inferred from concatenated multiple alignments of the conserved single-copy genes of homologous proteins using the Codon Tree method. Typically, up to 100 of shared proteins were selected specifically for each group of prokaryotic genomes under analysis. The combined protein/coding DNA sequence alignments were processed using the program RAXML for maximum likelihood tree building (Stamatakis 2014). Robustness of the built trees was assessed using the gene-wise jackknifing procedure (Simmons et al. 2019) with 100 resampling replicates per tree. The trees were visualized using FigTree software (<http://tree.bio.ed.ac.uk/software/figtree/>).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

This work was supported in part by NASA Exobiology grant NNX14AK36G and NASA contract 80NSSC18K1139 under the Center for Origin of Life, Georgia Institute of Technology to G.E.F.

Received July 28, 2020; accepted November 9, 2020.

REFERENCES

- Armache J-P, Anger AM, Marquez V, Franckenberg S, Fröhlich T, Villa E, Berninghausen O, Thomm M, Arnold GJ, Beckmann R. 2012. Promiscuous behaviour of archaeal ribosomal proteins: implications for eukaryotic ribosome evolution. *Nucleic Acids Res* **41**: 1284–1293. doi:10.1093/nar/gks1259
- Babaian A, Rothe K, Girodat D, Minia I, Djondovic S, Milek M, Spencer Miko SE, Wieden HJ, Landthaler M, Morin GB, et al. 2020. Loss of m¹acp^{3Ψ} ribosomal RNA modification is a major feature of cancer. *Cell Rep* **31**: 107611. doi:10.1016/j.celrep.2020.107611
- Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**: 905–920. doi:10.1126/science.289.5481.905
- Byrgazov K, Vesper O, Moll I. 2013. Ribosome heterogeneity: another level of complexity in bacterial translation regulation. *Curr Opin Microbiol* **16**: 133–139. doi:10.1016/j.mib.2013.01.009
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421. doi:10.1186/1471-2105-10-421
- Carver T, Berriman M, Tivey A, Patel C, Bohme U, Barrell BG, Parkhill J, Rajandream MA. 2008. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* **24**: 2672–2676. doi:10.1093/bioinformatics/btn529

- Daniels CJ, Hofman JD, MacWilliam JG, Doolittle FW, Woese CR, Luehrsén KR, Fox GE. 1985. Sequence of 5S ribosomal RNA gene regions and their products in the archaeobacterium *Halobacterium volcanii*. *Mol Gen Genet* **198**: 270–274. doi:10.1007/BF00383005
- De Rijk P, Wuyts J, De Wachter R. 2003. RnaViz 2: an improved representation of RNA secondary structure. *Bioinformatics* **19**: 299–300. doi:10.1093/bioinformatics/19.2.299
- Dontsova OA, Dinman JD. 2005. 5S rRNA: structure and function from head to toe. *Int J Biomed Sci* **1**: 1–7.
- Erdmann VA, Wolters J. 1986. Collection of published 5S, 5.8S and 4.5S ribosomal RNA sequences. *Nucleic Acids Res* **14**: r1–r59. doi:10.1093/nar/14.suppl.r1
- Erdmann VA, Wolters J, Pieler T, Digweed M, Specht T, Ulbrich N. 1987. Evolution of organisms and organelles as studied by comparative computer and biochemical analyses of ribosomal 5S RNA structure. *Ann N Y Acad Sci* **503**: 103–124. doi:10.1111/j.1749-6632.1987.tb40602.x
- Everett KDE, Kahane S, Bush RM, Friedman MG. 1999. An unspliced group I intron in 23S rRNA links *Chlamydiales*, chloroplasts, and mitochondria. *J Bacteriol* **181**: 4734–4740. doi:10.1128/JB.181.16.4734-4740.1999
- Evguenieva-Hackenberg E. 2005. Bacterial ribosomal RNA in pieces. *Mol Microbiol* **57**: 318–325. doi:10.1111/j.1365-2958.2005.04662.x
- Ferretti MB, Karbstein K. 2019. Does functional specialization of ribosomes really exist? *RNA* **25**: 521–538. doi:10.1261/rna.069823.118
- Fox GE. 1985. The structure and evolution of archaeobacterial ribosomal RNA. In *The bacteria: a treatise on structure and function*, Vol. VIII: archaeobacteria (ed. Woese CR, Wolfe RS), pp. 257–310. Academic Press, New York.
- Gabdulkhakov A, Nikonov S, Garber M. 2013. Revisiting the *Haloarcula marismortui* 50S ribosomal subunit model. *Acta Crystallogr D Biol Crystallogr* **69**: 997–1004. doi:10.1107/S0907444913004745
- Genuth NR, Barna M. 2018a. The discovery of ribosome heterogeneity and its implications for gene regulation and organismal life. *Mol Cell* **71**: 364–374. doi:10.1016/j.molcel.2018.07.018
- Genuth NR, Barna M. 2018b. Heterogeneity and specialized functions of translation machinery: from genes to organisms. *Nat Rev Genet* **19**: 431–452. doi:10.1038/s41576-018-0008-z
- Gongadze GM. 2011. 5S rRNA and ribosome. *Biochemistry (Mosc)* **76**: 1450–1464. doi:10.1134/S0006297911130062
- Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**: 221–224. doi:10.1093/molbev/msp259
- Greber BJ, Boehringer D, Godinic-Mikulcic V, Crnkovic A, Ibba M, Weygand-Durasevic I, Ban N. 2012. Cryo-EM structure of the archaeal 50S ribosomal subunit in complex with initiation factor 6 and implications for ribosome evolution. *J Mol Biol* **418**: 145–160. doi:10.1016/j.jmb.2012.01.018
- Guo H. 2018. Specialized ribosomes and the control of translation. *Biochem Soc Trans* **46**: 855–869. doi:10.1042/BST20160426
- Hemme CL, Mouttaki H, Lee YJ, Zhang G, Goodwin L, Lucas S, Copeland A, Lapidus A, Glavina del Rio T, Tice H, et al. 2010. Sequencing of multiple clostridial genomes related to biomass conversion and biofuel production. *J Bacteriol* **192**: 6494–6496. doi:10.1128/JB.01064-10
- Hofacker IL, Lorenz R. 2014. Predicting RNA structure: advances and limitations. *Methods Mol Biol* **1086**: 1–19. doi:10.1007/978-1-62703-667-2_1
- Huang S, Aleksashin NA, Loveland AB, Klepacki D, Reier K, Kefi A, Szal T, Remme J, Jaeger L, Vazquez-Laslop N, et al. 2020. Ribosome engineering reveals the importance of 5S rRNA autonomy for ribosome assembly. *Nat Commun* **11**: 2900. doi:10.1038/s41467-020-16694-8
- Jiang Y, Liu J, Dong W, Zhang W, Fang Y, Ma J, Jiang M, Xin F. 2018. The draft genome sequence of thermophilic *Thermoanaerobacterium thermosaccharolyticum* M5 capable of directly producing butanol from hemicellulose. *Curr Microbiol* **75**: 620–623. doi:10.1007/s00284-017-1425-5
- Kalvari I, Nawrocki EP, Argasinska J, Quinones-Olvera N, Finn RD, Bateman A, Petrov AI. 2018. Non-coding RNA analysis using the Rfam database. *Curr Protoc Bioinformatics* **62**: e51. doi:10.1002/cpbi.51
- Kiparisov S, Petrov A, Meskauskas A, Sergiev P, Dontsova OA, Dinman JD. 2005. Structural and functional analysis of 5S rRNA in *Saccharomyces cerevisiae*. *Mol Genet Genomics* **274**: 235–247. doi:10.1007/s00438-005-0020-9
- Kouvela EC, Gerbanas GV, Xaplanteri MA, Petropoulos AD, Dinos GP, Kalpaxis DL. 2007. Changes in the conformation of 5S rRNA cause alterations in principal functions of the ribosomal nanomachine. *Nucleic Acids Res* **35**: 5108–5119. doi:10.1093/nar/gkm546
- Kurylo CM, Parks MM, Juetz MF, Zinshteyn B, Altman RB, Thibado JK, Vincent CT, Blanchard SC. 2018. Endogenous rRNA sequence variation can regulate stress response gene expression and phenotype. *Cell Rep* **25**: 236–248. doi:10.1016/j.celrep.2018.08.093
- Kushwaha AK, Bhushan S. 2020. Unique structural features of the *Mycobacterium* ribosome. *Prog Biophys Mol Biol* **152**: 15–24. doi:10.1016/j.pbiomolbio.2019.12.001
- Li T, Zhang C, Yang KL, He J. 2018. Unique genetic cassettes in a *Thermoanaerobacterium* contribute to simultaneous conversion of cellulose and monosugars into butanol. *Sci Adv* **4**: e1701475. doi:10.1126/sciadv.1701475
- López García de Lomana A, Kusebauch U, Raman AV, Pan M, Turkarslan S, Lorenzetti APR, Moritz RL, Baliga NS. 2020. Selective translation of low abundance and upregulated transcripts in *Halobacterium salinarum*. *mSystems* **5**: e00329-20. doi:10.1128/mSystems.00329-20
- López-López A, Benlloch S, Bonfá M, Rodríguez-Valera F, Mira A. 2007. Intragenomic 16S rDNA divergence in *Haloarcula marismortui* is an adaptation to different temperatures. *J Mol Evol* **65**: 687–696. doi:10.1007/s00239-007-9047-3
- Lorenz R, Hofacker IL, Stadler PF. 2016. RNA folding with hard and soft constraints. *Algorithms Mol Biol* **11**: 8. doi:10.1186/s13015-016-0070-z
- Luehrsén KR, Nicholson DE, Eubanks DC, Fox GE. 1981. An archaeobacterial 5S rRNA contains a long insertion sequence. *Nature* **293**: 755–756. doi:10.1038/293755a0
- Macke TJ, Ecker DJ, Gutell RR, Gautheret D, Case DA, Sampath R. 2001. RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res* **29**: 4724–4735. doi:10.1093/nar/29.22.4724
- Marcel V, Catez F, Diaz JJ. 2015. Ribosome heterogeneity in tumorigenesis: the rRNA point of view. *Mol Cell Oncol* **2**: e983755. doi:10.4161/23723556.2014.983755
- Melnikov S, Ben-Shem A, Garreau de Loubresse N, Jenner L, Yusupova G, Yusupov M. 2012. One core, two shells: bacterial and eukaryotic ribosomes. *Nat Struct Mol Biol* **19**: 560–567. doi:10.1038/nsmb.2313
- Nicholson DE. 1982. “Structure of 5S ribosomal RNA from *Halobacteriaceae*.” PhD thesis, University of Houston, Houston, Texas.
- Nicholson DE, Fox GE. 1983. Molecular evidence for a close phylogenetic relationship among box-shaped halophilic bacteria, *Halobacterium vallismortis*, and *Halobacterium marismortui*. *Can J Microbiol* **29**: 52–59. doi:10.1139/m83-009

- Parker MS, Sah R, Balasubramaniam A, Sallee FR, Park EA, Parker SL. 2014. On the expansion of ribosomal proteins and RNAs in eukaryotes. *Amino Acids* **46**: 1589–1604. doi:10.1007/s00726-014-1704-4
- Parker MS, Sallee FR, Park EA, Parker SL. 2015. Homoiterons and expansion in ribosomal RNAs. *FEBS Open Bio* **5**: 864–876. doi:10.1016/j.fob.2015.10.005
- Parks MM, Kurylo CM, Dass RA, Bojmar L, Lyden D, Vincent CT, Blanchard SC. 2018. Variant ribosomal RNA alleles are conserved and exhibit tissue-specific expression. *Sci Adv* **4**: eaao0665. doi:10.1126/sciadv.aao0665
- Pei A, Li H, Oberdorf WE, Alekseyenko AV, Parsons T, Yang L, Gerz EA, Lee P, Xiang C, Nossa CW, et al. 2012. Diversity of 5S rRNA genes within individual prokaryotic genomes. *FEMS Microbiol Lett* **335**: 11–18. doi:10.1111/j.1574-6968.2012.02632.x
- Penev PI, Fakhretaha-Aval S, Patel VJ, Cannone JJ, Gutell RR, Petrov AS, Williams LD, Glass JB. 2020. Supersized ribosomal RNA expansion segments in Asgard archaea. *Genome Biol Evol* **12**: 1694–1710. doi: org/10.1093/gbe/evaa170
- Rivas E, Clements J, Eddy SR. 2017. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat Methods* **14**: 45–48. doi:10.1038/nmeth.4066
- Rogers SO. 2019. Integrated evolution of ribosomal RNAs, introns, and intron nurseries. *Genetica* **147**: 103–119. doi:10.1007/s10709-018-0050-y
- Roller C, Ludwig W, Schleifer KH. 1992. Gram-positive bacteria with a high DNA G + C content are characterized by a common insertion within their 23S rRNA genes. *J Gen Microbiol* **138**: 1167–1175. doi:10.1099/00221287-138-6-1167
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944–945. doi:10.1093/bioinformatics/16.10.944
- Sato Y, Fujiwara T, Kimura H. 2017. Expression and function of different guanine-plus-cytosine content 16S rRNA genes in *Haloarcula hispanica* at different temperatures. *Front Microbiol* **8**: 482. doi:10.3389/fmicb.2017.00482
- Sauert M, Temmel H, Moll I. 2015. Heterogeneity of the translational machinery: variations on a common theme. *Biochimie* **114**: 39–47. doi:10.1016/j.biochi.2014.12.011
- Shi Z, Fujii K, Kovary KM, Genuth NR, Röst HL, Teruel MN, Barna M. 2017. Heterogeneous ribosomes preferentially translate distinct subpools of mRNAs genome-wide. *Mol Cell* **67**: 71–83. doi:10.1016/j.molcel.2017.05.021
- Simmons MP, Sloan DB, Springer MS, Gatesy J. 2019. Gene-wise resampling outperforms site-wise resampling in phylogenetic coalescence analyses. *Mol Phylogenet Evol* **131**: 80–92. doi:10.1016/j.ympev.2018.10.001
- Smirnov AV, Entelis NS, Krashennnikov IA, Martin R, Tarassov IA. 2008. Specific features of 5S rRNA structure: its interactions with macromolecules and possible functions. *Biochemistry (Mosc)* **73**: 1418–1437. doi:10.1134/S000629790813004X
- Song W, Joo M, Yeom JH, Shin E, Lee M, Choi HK, Hwang J, Kim YI, Seo R, Lee JE, et al. 2019. Divergent rRNAs as regulators of gene expression at the ribosome level. *Nat Microbiol* **4**: 515–526. doi:10.1038/s41564-018-0341-1
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313. doi:10.1093/bioinformatics/btu033
- Stan-Lotter H, McGenity TJ, Legat A, Denner EBM, Glaser K, Stetter KO, Wanner G. 1999. Very similar strains of *Halococcus salifodinae* are found in geographically separated Permo-Triassic salt deposits. *Microbiology* **145**: 3565–3574. doi:10.1099/00221287-145-12-3565
- Sun FJ, Caetano-Anolles G. 2009. The evolutionary history of the structure of 5S ribosomal RNA. *J Mol Evol* **69**: 430–443. doi:10.1007/s00239-009-9264-z
- Sutton LA, Woese CR. 1975. Stable large variant of 5S rRNA in *Clostridium thermosaccharolyticum*. *Nature* **256**: 64–66. doi:10.1038/256064a0
- Szymanski M, Barciszewska MZ, Barciszewski J, Erdmann VA. 1999. 5S Ribosomal RNA data bank. *Nucleic Acids Res* **27**: 158–160. doi:10.1093/nar/27.1.158
- Szymanski M, Barciszewska MZ, Erdmann VA, Barciszewski J. 2002. 5S Ribosomal RNA database. *Nucleic Acids Res* **30**: 176–178. doi:10.1093/nar/30.1.176
- Szymanski M, Zielezinski A, Barciszewski J, Erdmann VA, Karlowski WM. 2016. 5SRNAdb: an information resource for 5S ribosomal RNAs. *Nucleic Acids Res* **44**: D180–D183. doi:10.1093/nar/gkv1081
- Tirumalai MR, Kaelber JT, Park DR, Tran Q, Fox GE. 2020. Cryo-electron microscopy visualization of a large insertion in the 5S ribosomal RNA of the extremely halophilic archaeon *Halococcus morrhuae*. *FEBS Open Bio* **10**: 1938–1946. doi:10.1002/2211-5463.12962
- Wattam AR, Gabbard JL, Shukla M, Sobral BW. 2014. Comparative genomic analysis at the PATRIC, a bioinformatic resource center. *Methods Mol Biol* **1197**: 287–308. doi:10.1007/978-1-4939-1261-2_17
- Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, et al. 2017. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res* **45**: D535–D542. doi:10.1093/nar/gkw1017
- Weisser M, Ban N. 2019. Extensions, extra factors, and extreme complexity: ribosomal structures provide insights into eukaryotic translation. *Cold Spring Harb Prospect Biol* **11**: a032367. doi:10.1101/cshperspect.a032367
- Wolters J, Erdmann VA. 1988. Compilation of 5S rRNA and 5S rRNA gene sequences. *Nucleic Acids Res* **16**: r1–r70. doi:10.1093/nar/16.suppl.r1
- Wolters J, Erdmann VA. 1989. The structure and evolution of archaeobacterial ribosomal RNAs. *Can J Microbiol* **35**: 43–51. doi:10.1139/m89-007
- Xu ZZ, Mathews DH. 2016. Secondary structure prediction of single sequences using RNAstructure. *Methods Mol Biol* **1490**: 15–34. doi:10.1007/978-1-4939-6433-8_2
- Yang K, Chang JY, Cui Z, Li X, Meng R, Duan L, Thongchol J, Jakana J, Huwe CM, Sacchettini JC, et al. 2017. Structural insights into species-specific features of the ribosome from the human pathogen *Mycobacterium tuberculosis*. *Nucleic Acids Res* **45**: 10884–10894. doi:10.1093/nar/gkx785

RNA

A PUBLICATION OF THE RNA SOCIETY

Expansion segments in bacterial and archaeal 5S ribosomal RNAs

Victor G. Stepanov and George E. Fox

RNA 2021 27: 133-150 originally published online November 12, 2020

Access the most recent version at doi:[10.1261/rna.077123.120](https://doi.org/10.1261/rna.077123.120)

Supplemental Material

<http://rnajournal.cshlp.org/content/suppl/2020/11/12/rna.077123.120.DC1>

References

This article cites 69 articles, 9 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/27/2/133.full.html#ref-list-1>

Open Access

Freely available online through the RNA Open Access option.

Creative Commons License

This article, published in RNA, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to RNA go to:

<http://rnajournal.cshlp.org/subscriptions>