Diversity of 5S rRNA genes within individual prokaryotic genomes

Anna Pei[1], **Hongru Li**[2], **William E Oberdorf**[3], **Alexander V. Alekseyenko**[3], **Tamasha Parsons**[3], **Liying Yang**[3], **Erika A. Gerz**[3], **Peng Lee**[4,5], **Charlie Xiang**[6], **Carlos W. Nossa**[7], and **Zhiheng Pei**[3,4,5]

[1]New York University College of Arts and Science, New York, NY 10012

[2]Mount Sinai School of Medicine, Graduate School of Biological Sciences, New York, NY, USA 10029

[3]Department of Medicin, New York University School of Medicine, New York, NY 10016

[4]Department of Pathology, New York University School of Medicine, New York, NY 10016

[5]Department of Veterans Affairs New York Harbor Healthcare System, New York, NY 10010

[6]Genome and Bioinformatics Center, Zhejiang University College of Medicine, China 310003

[7]Ecology and Evolutionary Biology, Rice University, Houston, TX 77005

Abstract

We examined intragenomic variation of paralogous 5S rRNA genes to evaluate the concept of ribosomal constraints. In a dataset containing 1168 genomes from 779 unique species, 96 species exhibited >3% diversity. Twenty seven species with >10% diversity contained a total of 421 mismatches between all pairs of the most dissimilar copies of 5S rRNA genes. The large majority (401 of 421) the diversified positions were conserved at the secondary structure level. The high diversity was associated with partial rRNA operon, split operon, or spacer length-related divergence. In total, these findings indicated that there were tight ribosomal constraints on paralogous 5S rRNA genes in a genome despite of the high degree of diversity at the primary structure level.

There is supplementary material.

Keywords

rRNA diversity; Ribosomal constraints; 5S rRNA

INTRODUCTION

Ribosomal RNA genes (rRNA genes) are widely used for documentation of evolutionary history and taxonomic assignment of individual organisms (8, 12, 20, 21, 22). The choice of rRNA genes as optimal tools for such purposes is based on both observations and assumptions of rRNA gene conservation (10, 21). The rRNA genes are essential components of the ribosome consisting of more than 50 proteins and three classes of RNA molecules; precise spatial relationships may be essential for assembly of functional ribosomes,

Corresponding author: Zhiheng Pei, MD, PhD, Associate Professor of Pathology and Medicine, New York University School of Medicine, Phone: (212) 951-5492, Fax: (212) 263-4108, peiz01@med.nyu.edu.

constraining rRNA genes from drastic change (4, 7). The concept of ribosomal constraints has been examined by analysis of intragenomic variation among paralogous 23S rRNA (17) as well as 16S rRNA genes (2, 6, 16). Evidence supporting the concept includes similarity at the primary structure level and conservation of the secondary structure in cases with significant diversity in the primary structure. 5S rRNA is the smallest gene in a ribosomal operon, with an average length of only 120 nt. Whether paralogous 5S rRNA genes comply with ribosomal constraints has not been evaluated.

With the increasing database of whole microbial genomes available from the National Center for Biotechnology Information (NCBI), we systemically evaluated the extent of 5S rRNA gene diversity within single organisms and addressed the theory of ribosomal constraints.

MATERIALS AND METHODS

Annotation of rRNA genes

5S gene sequences were obtained from the Complete Microbial Genomes database at the NCBI website (http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi). For some species with more than one genome available in the database, only the most completely annotated genome was included for analysis to avoid overrepresentation of any species. In genomes lacking adequate annotations, 5S rRNA genes were identified by using experimentally defined 5S rRNA sequences from the closest relatives available and verified by secondary structure analysis based on minimizing free energy, using RNAstructure (15) and Rnaviz 2.0 (6), with experimentally defined 5S rRNA used for reference. The number of 5S rRNA genes present in a genome was determined by whole genome BLAST search based on the known 5S rRNA sequence.

Analysis of intragenomic diversity in rRNA operons

Genomes that contained only a single 5S rRNA gene operon were not further analyzed. Copies of 5S rRNA genes from each remaining genome were aligned with Clustalw (18). To calculate diversity, we normalized the number of revealed mismatches and indels by the total number of positions, including gaps in the alignment.

Comparison of secondary structures

To compare two related secondary structures, a mismatch was defined as conserved if it did not cause a stem-loop transition (16, 17). For example, a mismatch located in a loop was considered conserved because it maintained the loop structure and a mismatch located in a stem but causing GC:GU conversions or covariation was also considered conserved because it did not cause a change in base-pairing or disruption of the stem. In contrast, a non-conserved mismatch was one that altered base-pairing and converted a loop to a stem or a stem to a loop.

RESULTS

5S rRNA genes dataset

In total, 1161 complete prokaryotic genomes were available for analysis, 86 from Archaea and 1075 from Bacteria, representing 779 unique species (75 Archaea and 704 Bacteria) (Table S1). Of the 779 species, 174 genomes contained only a single 5S rRNA gene. Remaining were 605 unique species (40 Archaea, 565 Bacteria) whose genomes contained multiple 5S rRNA genes, representing 27 phyla. *Proteobacteria* was the most abundant phylum (344 species) in the dataset followed by *Firmicutes* (123 species), *Actinobacteria*

(82 species), *Euryarchaeota* (53 species) and *Bacteroidetes/Chlorobi* (36 species). The remaining 22 phyla were represented by only 141 species.

Diversity of 5S rRNA genes

The 605 genomes examined contained 2–19 copies of 5S rRNA genes (median 4 copies/ genome, interquartile range (IQR) 2–6); 388 genomes had 5S rRNA genes that were identical, and 217 had 5S rRNA genes that were diversified. For each of the 217 diversified species, the most dissimilar 5s rRNA gene pair has been identified by pair-wise analysis of all possible pairs. Maximal diversity ranged from 0.60% to 26.15% (median 2.50%, interquartile range (IQR) 0.88–5.91%) (24). Sixteen genomes with >13.44% diversity between the most dissimilar pair of 5S rRNA genes—*Staphylococcus saprophyticus subsp. saprophyticus, Actinobacillus pleuropneumoniae, Thermoanaerobacter pseudethanolicus, Desulfotomaculum acetoxidans, Bifidobacterium adentium, Lactococcus lactis subsp. cremoris, Francisella novicida, Syntrophomonas wofei subsp. Wolfei, Methanosphaerula palustris, Francisella tularnesis subsp. holarctica, Psychromonas ingrahamii, Bacillus megaterium, Actinobacillus succinogenes, Symbiobacterium thermophilum, Aggregatibacter aphrophilus,* and *Haemophilus influenzae*, were classified as outliers, using Tukey's boxplot (24). In 158 genomes, the maximal diversity exceeded 1% and 96 genomes had more than 3% intragenomic diversity. Comparison of the intragenomic diversity of 5S rRNA, 16S rRNA and 23S rRNA was made, and 5S rRNA has the most widespread intragenomic variation (Figure 1). The diversity was due to point mutations or single-nucleotide indels; intervening sequences (IVS), commonly present in 16S and 23S rRNA genes, were not found in 5S rRNA genes.

Impact of diversity on secondary structures

Twenty seven genomes with >10% intragenomic diversity between their 5S rRNA genes were further examined for the impact of the diversity on secondary structure. The two most diversified 5S rRNA genes were selected for the analysis. Secondary structures of the 5S rRNA genes were constructed based on the principle of minimization of free energy (15), using experimentally defined rRNA as references. In the 27 genomes, there were a total of 421 diversified positions between all pairs of the most dissimilar 5S rRNA genes. Conservative mutations comprised 401 (95.25%) positions, including 125 in loops, 202 co-variations, and 74 GU:GC conversions (Table 1). Only 20 (4.75%) of the 421 diversified positions caused changes in the secondary structures of 5S rRNA genes in 14 genomes (*S. amazonensis, A. prevotii, C. beijerinckii, T. auensis, H. somnus, H. influenzae, A. aphrophilus, S. thermophilum, B. megaterium, P. ingrahamii, L. lactis subsp. cremoris, T. pseudethanolicus, A. pleuropneumoniae, S. saprophyticus subsp. saprophyticus*). Only five genomes (*C. beijerinckii, T. auensis, H. influenzae, L. lactis subsp. cremoris,* and *A. pleuropneumoniae*) had the secondary structures altered at more than one position in the 5S rRNA genes (Figure 2). Insertions/deletions (indels) occurred at 46 of the 421 positions.

Characteristics of the most diversified 5S rRNA genes

The 96 genomes with >3% diversity between 5S rRNA genes (Table S1) can be categorized into five groups based on the potential mechanisms that may explain the observed high diversity (Figure 3). (i). Partial operon in which an orphan 5S rRNA gene, unassociated with 16S and 23S rRNA gene, was near an intact rRNA operon (Figure 3a). In 52 of the 96 genomes with >3% diversity, the maximal diversity occurred between the orphan 5S rRNA genes and 5S rRNA genes in a complete operon (Table 2), reaching 15.45% in *F. tularensis subsp. holarctica* and 13.04% in *H. ducreyi*. (ii). Split operon. In 8 of the 96 genomes, the 5S rRNA gene most dissimilar to the majority of other 5S rRNA gene copies was physically separated from the rRNA operon it belongs to (Table 3). For example, in *C. perfringens*, the 5S rRNA gene rrnH5S (12.61% diversity) was located ~240,000-nt from rrnH16S and

rrnH23S. Similarly, in *G. kaustophilus*, the minor 5S rRNA gene (4.92% diversity) was located ~2,800,000-nt from the remaining rRNA operon that contained 16S and 23S rRNA genes. (iii). 5S-23S spacer-length lineage divergence. In *Bacillus*, 5S rRNA genes can be grouped based on the 23S-5S spacer-length variation. In 24 of the 96 diversified genomes, maximal diversity was related to divergence of the spacer variants (Table 4). For example, in *B. subtilis*, maximal diversity (6.90%) existed between the nine 5S rRNA genes that had 56-nt 23S-5S spacers and the one 5S rRNA gene with a 112-nt spacer. (iv). Divergent operon. In *T. tengcongensis*, the rrnC operon differed from the other three operons by 3.70% at 5S, 6.70% at 16S, and 4.04% at the 23S rRNA gene loci. (v). Unusual alteration of secondary structures. In *A. pleuropneumoniae, C. beijerinckii, H. influenza, L. lactis subsp. cremoris,* and *T. auensis*, the secondary structures were altered between the two most dissimilar 5S rRNA genes at 3, 2, 2, 2 and 2 positions (Figure 2), respectively. In comparison, none of the other genomes analyzed had altered secondary structures of 5S rRNA genes at more than one position.

DISCUSSION

We analyzed 5S rRNA genes from genomes representing 779 prokaryotic species to look for evidence of ribosomal constraint of rRNA structures at the intragenomic level. Our findings indicated that individual 5S rRNA genes within a genome were conserved due to such structural constraints, with rare exceptions. The large majority of genomes (683 of 779) in which diversity is <3% between primary sequences of paralogous rRNA genes provided one type of evidence for constraints. Another type of constraint was at the level of secondary structures; 27 genomes with >10% rRNA gene diversity showed striking conservation of more than 95.25% of diversified positions at the secondary structure level. Significant differences between rRNA genes in single organisms, albeit few, have been discovered in all three domains of life, and in all three classes of rRNA genes. The amphibian *Xenopus laevis* and the loach *Misgurnus fossilis* have two types of 5S rRNA genes that are specific to either somatic or oocyte ribosomes (14, 19). The parasite *Plasmodium berghei* contains two types of 18S rRNA genes that differ at 3.5% of the nucleotide positions and that are life-cycle stage-specific (9). In *A. pleuropneumoniae, C. beijerinckii, H. influenzae, L. lactis subsp. cremoris,* and *T. auensis*, the abnormally high diversity among their 5S rRNA genes with significant alterations of secondary structures suggested diminished ribosomal constraints in some individual rRNA genes, or constraints in higher order structures (3, 10, 23). Alternatively, the apparent violation of ribosomal constraints in 5S rRNA genes in these species may be explained by physiological requirement of different types of 5S rRNA genes by these prokaryotic organisms, similar to the above-mentioned eukaryotes.

Multiple mechanisms might be responsible for generating the observed diversity in 5S rRNA genes in a genome. In organisms containing multiple rRNA genes, the homogeneity of primary structures is believed to be maintained through gene conversion by homologous recombination (11), as a form of concerted evolution (1). Although the observed homogeneity of 5S rRNA genes in the majority of species analyzed could be attributed to the effect of homologous recombination, the recombination appeared to be compartmentalized or ineffective in some genomes. The observed high degree of diversity in the primary structures of the 5S rRNA genes in partial or split rRNA gene operons and the rrnC operon in *T. tengcongensis* suggested that these rRNA genes have been excluded from participation in concerted evolution. Such compartmentalization was also present in *Bacillus subtilis* that has two similarity groups of rRNA genes appeared to have evolved independently, as evidenced by their relation to different 5S rRNA genes-rrn23S spacers. Despite the lack of sequence homogeneity, secondary structures of these genes were well conserved, most likely due to the life and death driving force of ribosomal constraints.

Compared with whole 16S and 23S rRNA genes, 5S rRNA genes are a less ideal taxonomical marker for use in analyses of complex microbiomes. The main reason is the widespread intragenomic 5S rRNA gene diversity. Approximately 12.3% (96 of 779) of the unique species analyzed had >3% intragenomic variation of 5S rRNA genes, compared to only about 1% of species with similar degree of variation in 16S and 23S rRNA genes (2, 5, 16). This high degree of diversity most often occurs between a standalone 5S rRNA gene (orphan or split) and a 5S rRNA gene in a complete rRNA operon. The lack of standalone 16S or 23 S rRNA genes is the main reason for the lower intragenomic diversity among 16S or 23S rRNA genes. Orphan 5S rRNA genes are sometimes overlooked by a whole genome annotation program because of their small size. Compared with rrnDB (13), a publically accessible database that collects existing data on structure RNA genes from whole genome seuqnecing projects, 11 genomes listed in Table 1 had additional 5S rRNA genes in our study not listed in rrnDB. The additional 5S rRNA genes would have been invisible if blast search of 5S rRNA genes against the whole genomes were not performed. Nevertheless, in 26 genomes of the 52 genomes listed in Table 1, correct records of the orphan 5S rRNA genes can be found in rrnDB. The remaining 15 of the 52 genomes have no entries in rrnDB. Divergent evolution between paralogous 5S rRNA genes in a genome may corrupt the record of evolutionary history and obscure the true identity of an organism. With substantial variation, use of 5S rRNA gene as a taxonomic marker may lead to the artificial classification of an organism into more than one species. For a cultivable organism, the highly diversified 5S rRNA genes can be correctively traced to a single species when pure culture is available for verification. However, cultivation-independent techniques have become a standard in studies of complex microbiomes that contain mixed species, such as the Human Microbiome Project. In this type of study, highly diversified 5S rRNA genes from the same genome would be misinterpreted as being from different species, leading to overestimation of species richness.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Abdulkarim F, Hughes D. Homologous recombination between the tuf genes of Salmonella typhimurium. J Mol Biol. 1996; 260:506–522. [PubMed: 8759317]

2. Acinas SG, Marcelino LA, Klepac-Ceraj V, Polz MF. Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. J Bacteriol. 2004; 186:2629–35. [PubMed: 15090503]

3. Babin P, Dolan M, Wollenzien P, Gutell RR. Identity and geometry of a base triple in 16S rRNA determined by comparative sequence analysis and molecular modeling. RNA. 1999; 5:1430–9. [PubMed: 10580471]

4. Clayton RA, Sutton G, Hinkle PS, Bult C, Fields C. Intraspecific variation in small subunit rRNA sequences in GenBank: why single sequences may not adequately represent prokaryotic taxa. Int J Syst Bacteriol. 1995; 45:595–599. [PubMed: 8590690]

5. Coenye T, Vandamme P. Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. FEMS Microbiol Lett. 2003; 228:45–9. [PubMed: 14612235]

6. De Rijk P, De Wachter R. RnaViz, a program for the visualization of RNA secondary structure. Nucleic Acids Res. 1997; 25:4679–4684. [PubMed: 9358182]

7. Doolittle WF. Phylogenetic classification and the universal tree. Science. 1999; 284:2124–2129. [PubMed: 10381871]

8. Eigen M, LB, Winkler-Oswatitsch R, Clarke CH. Pattern analysis of 5S rRNA. Proc Natl Acad Sci U S A. 1985; 82:2437–2441. [PubMed: 3921961]

9. Gunderson JH, Sogin ML, Wollet G, Hollingdale M, de la Cruz VF, Waters AP, McCutchan TF. Structurally distinct, stage-specific ribosomes occur in Plasmodium. Science. 1987; 238:933–937. [PubMed: 3672135]

10. Gutell RR, Noller HF, Woese CR. Higher order structure in ribosomal RNA. EMBO J. 1986; 5:1111–1113. [PubMed: 3720727]

11. Hashimoto JG, Stevenson BS, Schmidt TM. Rates and consequences of recombination between rRNA operons. J Bacteriol. 2003; 185:966–972. [PubMed: 12533472]

12. Küntzel H, Heidrich M, Piechulla B. Phylogenetic tree derived from bacterial, cytosol and organelle 5S rRNA sequences. Nucleic Acids Res. 1981; 9:1451–1461. [PubMed: 6785727]

13. Lee ZM, Bussema C 3rd, Schmidt TM. rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. Nucleic Acids Res. 2009; 37(Database issue):D489–93. [PubMed: 18948294]

14. Mashkova TD, Serenkova TL, Mazo AM, Avdonina TA, Timofeyeva Y, Kisselev LL. The primary structure of oocyte and somatic 5S rRNAs from the loach Misgurnus fossils. Nucleic Acids Res. 1981; 9:2141–2151. [PubMed: 7197777]

15. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci USA. 2004; 101:7287–7292. [PubMed: 15123812]

16. Pei A, Oberdorf WE, Nossa CW, Agarwal A, Chokshi P, Gerz EA, Jin Z, Lee P, Yang L, Pei Z. Diversity of 16S rRNA genes within individual prokaryotic genomes. Applied and environmental microbiology. 2010:3886–3897. [PubMed: 20418441]

17. Pei A, Nossa CW, Chokshi P, Blaser MJ, Yang L, Rosmarin DM, Pei Z. Diversity of 23S rRNA genes within individual prokaryotic genomes. PLoS One. 2009; 4:e5437. [PubMed: 19415112]

18. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22:4673–4680. [PubMed: 7984417]

19. Wegnez MR, Monier R, Denis H. Sequence heterogeneity of 5S rRNA in *Xenopus laevis*. FEBS Lett. 1972; 25:13–21. [PubMed: 11946710]

20. Woese CR. The universal ancestor. Proc Natl Acad Sci USA. 1998; 95:6854–6859. [PubMed: 9618502]

21. Woese CR. Bacterial evolution. Microbial Rev. 1987; 51:221–271.

22. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eucarya. Proc Natl Acad Sci USA. 1990; 87:4576–4579. [PubMed: 2112744]

23. Woese CR, Gutell RR. Evidence for several higher order structural elements in ribosomal RNA. Proc Natl Acad Sci USA. 1989; 86:3119–22. [PubMed: 2654936]

24. Wonacott, TH.; Wonacott, RJ. Introductory Statistics. John Willey & Son; 1990.
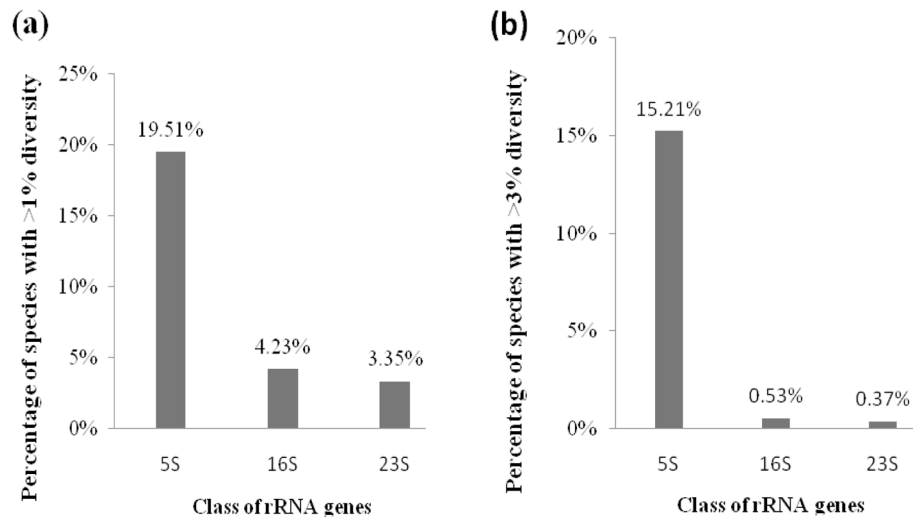
Fig. 1.
Difference of rRNA classes in intragenomic diversity, shown by percentage of species with >1% (a) or >3% (b) intragenomic diversity (16, 17).
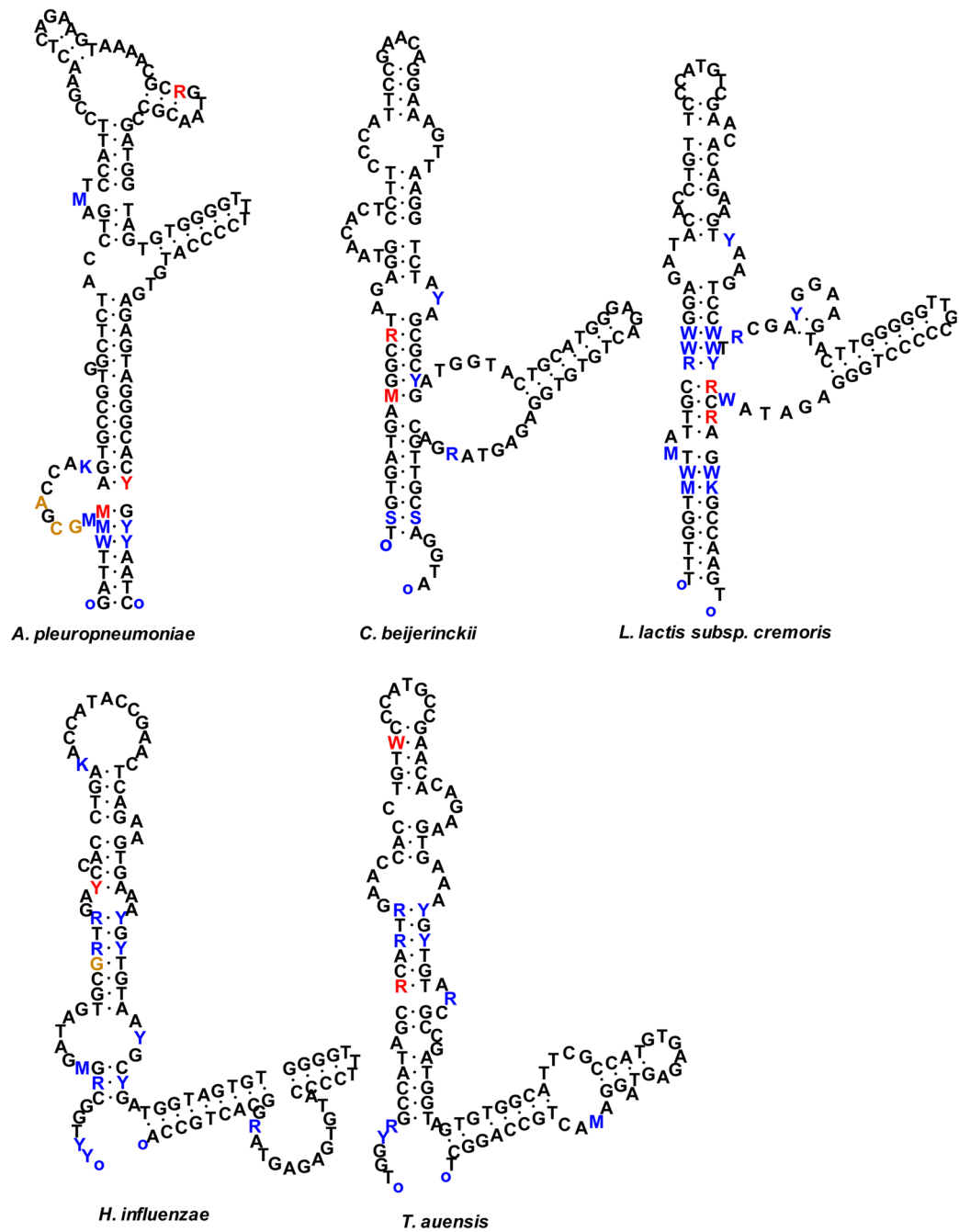
Fig. 2.
Species with alteration of secondary structures of 5S rRNA genes in two or more positions. Positions that differ between major and minor alleles of 5S rRNA are shown in colored letters. Conservative changes located in loops or compensatory changes due to co-variation in stems are shown in blue; changes that result in alteration of secondary structures are shown in red. Insertions/deletions are shown in brown. Substitutions are coded as follows: K = G or T, M = A or C, R = A or G, S = C or G, W = A or T, and Y = C or T.
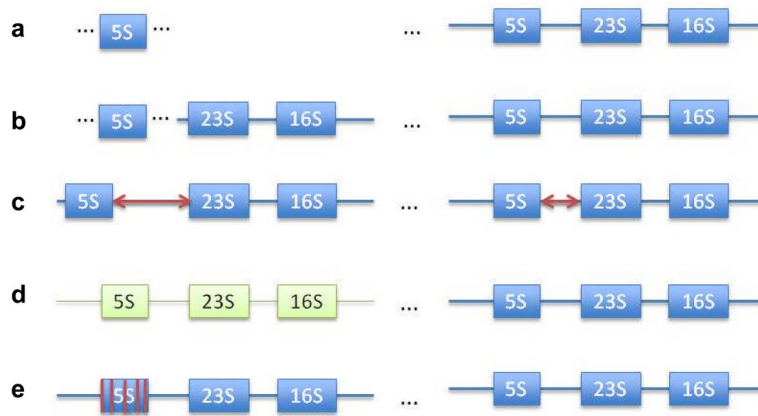
Fig. 3.
Classification of mechanisms responsible for divergence of paralogous 5S rRNA genes. (a) An orphan 5S gene copy is standalone and does not belong to a full operon; (b) a 5S gene in a split operon is separated by a far distance from its 16S and 23S counterparts; (c) spacer length divergence is related to length variation of the spacers between 5S and 23S rRNA genes within two complete operons; (d) Divergent operon is a complete operon that differs from other rrn operons in all three rRNA genes; (e) Unusual alteration of secondary structures occurs in a 5S rRNA gene whose secondary structure is drastically changed by mutations. In (a) and (b), all maximal divergences were with respect to a 5S gene copy belonging to a full operon.

Table 1

Impact of mutations on secondary structures

Species	Total mutations	Conserved			Non-conserved	Indels
		Loop:Loop	Stem:Stem		Stem:Loop	
			Co-variation	GU:GC		
Actinobacillus succinogenes	19	5	10	4	0	3
Actinobacillus pleuropneumoniae	13	3	4	3	3	3
Aggregatibacter actinomycetemcomitans	11	3	6	2	0	0
Aggregatibacter aphrophilus	14	5	8	0	1	0
Anaerococcus prevotii	12	0	6	5	1	0
Bacillus clausii	11	4	4	3	0	2
Bacillus megaterium	16	5	6	4	1	3
Bifidobacterium dentium	20	6	8	6	0	6
Clostridium bejerinckii	7	2	2	1	2	0
Clostridium perfringens	14	5	6	3	0	0
Desulfotomaculum acetoxidans	24	6	14	4	0	6
Francisella novicida	18	5	10	3	0	0
Francisella tularensis subsp. holarctica	17	5	10	2	0	1
Haemophilus ducreyi	14	5	8	1	0	1
Haemophilus influenzae	14	6	6	0	2	1
Haemophilus somnus	14	4	6	3	1	0
Jonesia denitrificans	11	3	6	2	0	2
Lactococcus lactis subsp. cremoris	17	4	10	1	2	0
Methanococcus vannielii	12	6	4	2	0	3
Methanosphaerula palustris	16	5	8	3	0	0
Psychromonas ingrahamii	17	4	8	4	1	1
Shewanella amazonensis	12	3	6	2	1	2
Staphylococcus saprophyticus subsp. saprophyticus	28	9	14	4	1	7
Symbiobacterium thermop hilum	16	5	8	2	1	1
Syntrophomonas wolfei su bsp. wolfei	17	6	6	5	0	4

Species	Total mutations	Conserved				Non-conserved		Indels
		Loop:Loop	Stem:Stem			Stem:Loop		
			Co-variation	GU:GC				
Thermoanaerobacter pseudethanolicus	27	7	14	5		1		0
Tolumonas auensis	10	4	4	0		2		0
Total	421	125	202	74		20		46

Table 2

Fifty two prokaryotic species with diversity >3% due to orphan 5S rRNA genes

Species	Copy of genes			5S diversity %
	5S	16S	23S	
Archea: Euryarchaeota (n=8)				
Methanosphaerula palustris	6	3	3	15.57
Methanococcus vannielii	5	4	4	10.53
Methanococcus maripaludis	4	3	3	9.65
Methanococcus voltae	3	2	2	8.77
Haloterrigena turkmenica	4	3	3	6.61
Methanothermobacter thermautotrophicus	3	2	2	6.11
Methanococcus aeolicus	4	2	2	4.96
Methanococcoides burtonii	6	3	3	4.13
Bacteria: Actinobacteria (n=3)				
Bifidobacterium dentium	5	4	4	18.80
Jonesia denitrificans	12	5	5	10.08
Bifidobacterium adolescentis	6	5	5	9.02
Bacteria: Firmicutes (n=10)				
Staphylococcus saprophyticus subsp. saprophyticus	8	6	6	26.15
Lactococcus lactis subsp. cremoris	7	6	6	18.10
Syntrophomonas wolfei subsp. wolfei	13	3	3	16.10
Bacillus megaterium	12	11	11	14.78
Symbiobacterium thermophilum	7	6	6	14.63
Bacillus clausii	8	7	7	10.34
Oceanobacillus iheyensis	8	7	7	5.98
Desulfotomaculum reducens	10	8	9	5.13
Bacillus selenitireducens	8	7	7	4.31
Bacillus halodurans	9	8	8	3.45
Bacteria: Fusobacteria (n=1)				
Streptobacillus moniliformis	6	5	5	6.48
Bacteria: Gammaproteobacteria (n=29)				
Actinobacillus pleuropneumoniae	7	3	6	25.20
Francisella novicida	4	3	3	17.89
Francisella tularensis subsp. holarctica	4	3	3	15.45
Psychromonas ingrahamii	11	10	10	15.45
Aggregatibacter aphrophilus	7	6	6	13.91
Haemophilus influenzae	7	6	6	13.68
Haemophilus ducreyi	7	6	6	13.04
Haemophilus somnus	6	5	10	12.61
Tolumonas auensis	9	8	8	11.40

Species	Copy of genes			5S diversity %
	5S	16S	23S	
Shewanella amazonensis	9	8	8	11.02
Aggregatibacter actinomycetemcomitans	7	6	6	10.11
Haemophilus parasuis	8	6	6	9.40
Vibrio parahaemolyticus	11	10	10	9.09
Mannheimia succiniciproducens	7	6	6	7.96
Aeromonas salmonicida subsp. salmonicida	10	9	9	7.81
Yersinia enterocolitica subsp. enterocolitica	8	7	7	7.00
Vibrio cholerae	8	7	7	6.78
Klebsiella variicola	9	8	8	6.09
Aliivibrio salmonicida	12	11	11	5.74
Vibrio harveyi	10	10	9	5.13
Klebsiella pneumoniae subsp. pneumoniae	9	8	8	4.88
Photorhabdus luminescens subsp. laumondii	8	7	7	3.94
Xenorhabdus bovienii	8	7	14	3.54
Marinomonas sp.	9	8	8	3.51
Citrobacter rodentium	8	7	7	3.45
Shewanella woodyi	11	10	10	3.45
Yersinia pestis Angola	8	7	7	3.33
Alteromonas macleodii	6	5	5	3.28
Enterobacter cloacae subsp. cloacae	9	8	8	3.05
Bacteria: Tenericutes n=(1)				
Mycoplasma synoviae	3	2	2	6.57

Table 3

Eight prokaryotic species with diversity >3% due to split ribosomal operon

Species	Copy of genes			5S diversity %
	5S	**16S**	**23S**	
Archea: Euryarchaeota (n=2)				
Methanococcus voltae	3	2	2	8.77
Methanococcus aeolicus	4	2	2	4.96
Bacteria: Actinobacteria (n=1)				
Rothia mucilaginosa	4	3	4	3.45
Bacteria: Firmicutes (n=5)				
Clostridium perfringens	10	10	10	12.61
Anaerococcus prevotii	4	4	4	11.21
Clostridium beijerinckii	15	14	14	11.21
Geobacillus sp.	9	8	8	8.55
Geobacillus kaustophilus	9	9	9	4.92

Table 4

Twenty four prokaryotic species with diversity >3% due to 5S-23S spacer length difference

Species	Copy of genes			5S diversity %
	5S	16S	23S	
Archea: Euryarchaeota (n=1)				
Methanosarcina acetivorans	3	3	3	4.44
Bacteria: Actinobacteria(n=3)				
Thermobispora bispora	3	4	3	6.84
Streptomyces avermitilis	6	6	6	6.25
Streptomyces griseus subsp. griseus	6	6	6	3.70
Bacteria: Firmicutes (n=16)				
Desulfotomaculum acetoxidans	9	10	11	22.86
Bacillus megaterium	12	11	11	14.78
Eubacterium rectale	5	5	5	8.55
Bacillus subtilis subsp. subtilis	10	10	10	6.90
Bacillus weihenstephanensis	14	14	14	6.03
Lactobacillus salivarius	7	7	7	6.03
Bacillus amyloliquefaciens	10	10	10	5.93
Bacillus pumilus	7	7	7	5.88
Geobacillus thermodenitrificans	10	10	10	5.51
Clostridium ljungdahlii	9	9	9	4.39
Bacillus anthracis	11	11	11	4.31
Bacillus thuringiensis	14	14	13	4.31
Clostridium novyi	10	10	10	4.17
Exiguobacterium sibiricum	9	9	9	3.48
Bacillus cereus	13	13	13	3.45
Clostridium botulinum	8	8	8	3.42
Bacteria: Gammaproteobacteria (n=4)				
Psychromonas ingrahamii	11	10	10	15.45
Shewanella loihica	9	8	8	4.41
Shewanella halifaxensis	10	10	10	3.45
Pseudomonas stutzeri	4	4	4	3.33