FEMS Microbiology Letters, 363, 2016, fnw171

doi: 10.1093/femsle/fnw171
Advance Access Publication Date: 12 July 2016
Minireview

MINIREVIEW – Physiology & Biochemistry

Thinking beside the box: Should we care about the non-coding strand of the 16S rRNA gene?

Jose F. Garcia-Mazcorro[1,*,†] and Jose R. Barcenas-Walls[2]

[1]Faculty of Veterinary Medicine, Universidad Autonoma de Nuevo Leon, Gral. Escobedo, Nuevo Leon 66050, Mexico and [2]Center for Research and Development in Health Sciences (CIDICS, Genomics Unit), Universidad Autonoma de Nuevo Leon, Monterrey, Nuevo Leon 64460, Mexico

*Corresponding author: Faculty of Veterinary Medicine, Universidad Autonoma de Nuevo Leon (UANL), Francisco Villa s/n Col. Ex-Hacienda el Canada. General Escobedo, Nuevo Leon 66050, Mexico. Tel: +011(52)83294000; E-mail: jose.garciamzc@uanl.edu.mx
One sentence summary: This minireview discusses the possibility that the non-coding strand of the 16S rRNA gene codes for something and shows preliminary data that it codes for small-interference RNA.
Editor: Olga Ozoline
†Jose F. Garcia-Mazcorro, http://orcid.org/0000-0001-6047-4361

ABSTRACT

The 16S rRNA gene (16S rDNA) codes for RNA that plays a fundamental role during translation in the ribosome and is used extensively as a marker gene to establish relationships among bacteria. However, the complementary non-coding 16S rDNA (nc16S rDNA) has been ignored. An idea emerged in the course of analyzing bacterial 16S rDNA sequences in search for nucleotide composition and substitution patterns: Does the nc16S rDNA code? If so, what does it code for? More importantly: Does 16S rDNA evolution reflect its own evolution or the evolution of its counterpart nc16S rDNA? The objective of this minireview is to discuss these thoughts. nc strands often encode small RNAs (sRNAs), ancient components of gene regulation. nc16S rDNA sequences from different bacterial groups were used to search for possible matches in the Bacterial Small Regulatory RNA Database. Intriguingly, the sequence of one published sRNA obtained from *Legionella pneumophila* (GenBank: AE017354.1) showed high non-random similarity with nc16S rDNA corresponding in part to the V5 region especially from *Legionella* and relatives. While the target(s) of this sRNA is unclear at the moment, its mere existence might open up a new chapter in the use of the 16S rDNA to study relationships among bacteria.

Keywords: 16S rDNA; non-coding strand; small RNAs; interference RNA; *Legionella*; evolution

INTRODUCTION

The genome of all life on Earth is composed by nucleic acids (DNA or RNA) containing information for survival, adaptation and preservation of the species. This genetic information is arranged in such a way that specific regions of the genome code for molecules that play an active role in the biology of the cell such as proteins, RNAs (e.g. ribosomal RNA or rRNA) and other small regulatory RNAs (sRNAs). Some regions serve as regulatory sequences for coding regions (e.g. promoters, enhancers) and the coding capacity of some other regions is simply unknown. All genomes evolve over time (albeit at different rates) and therefore today there is great variation among and within genomes from different species (Bustamante *et al*. 2005; Lukjancenko, Wassenaar and Ussery 2010), both in the way the genome is organized and in the way the genetic information is utilized to thrive.

16S rRNA gene (rDNA)

Genomes code for rRNA molecules that play a fundamental role during translation of messenger RNAs (mRNAs) in the ribosome. The 16S rDNA codes for the 16S rRNA that together with multiple

Received: 10 May 2016; **Accepted:** 8 July 2016

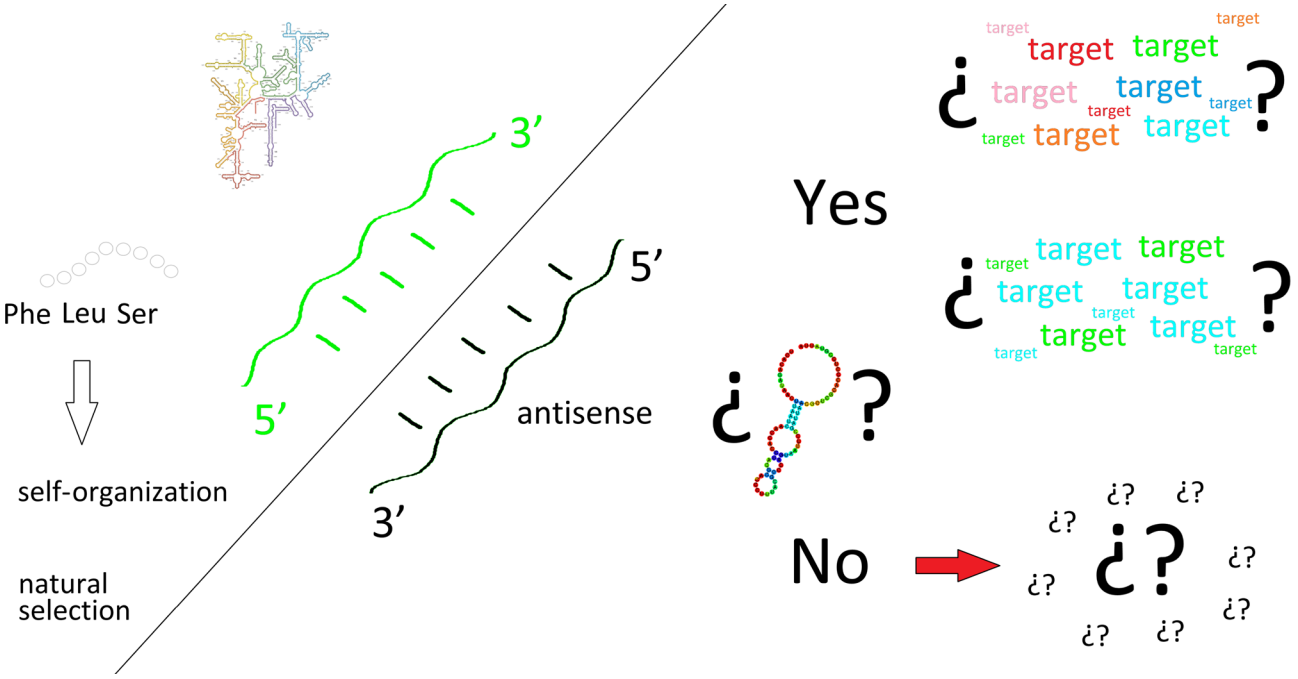

Figure 1. Thinking beside the box. On the left (green), the coding strand of the 16S rDNA codes for RNA that is crucial during translation of mRNAs in the ribosome. The translated proteins evolve accordingly to different theories. On the right (black), the nc 16S rDNA has been ignored by scientists in microbiology and related fields, including ourselves. This communication deals with the possibility that the nc16S rDNA codes for small regulatory RNAs and raises intriguing questions for future studies.

proteins compose the small 30S subunit of bacterial ribosomes (Wimberly *et al.* 2000). The 30S subunit is independent from the large 50S subunit right until translation is needed and its main function is to hold the mRNA during translation and select the correct transfer RNA for each codon (new evidence indicates the existence of a novel and frequent 70S-scanning mode of translation initiation, see Yamamoto *et al.* 2016).

The purpose of this minireview is to discuss whether the non-coding 16S rDNA (nc16S rDNA) codes something (Fig. 1). Here we will not discuss other ribosomal RNAs and will largely omit characteristics of the 16S rDNA such as its variable gene copies in bacterial genomes and intragenomic divergence (Acinas *et al.* 2004; Sun *et al.* 2013), the relationship between intragenomic divergence and environmental adaptation (Lopez-Lopez *et al.* 2007; Sanchez-Perez *et al.* 2008; Jensen, Frost and Torsvik 2009) and the fact that identical 16S rDNA sequences can be found in bacteria with highly divergent ecophysiologies (Jaspers and Overmann 2004). These topics are important in a context of 16S rDNA and ribosomal/bacterial evolution but can or have been discussed elsewhere.

Does the nc strand code?

Grassé (1977) was the very first who had the idea that proteins might be encoded in the opposite nc strands of protein-coding DNA sequences in overlapping fashion. Later, Casino *et al.* (1981) could not distinguish between coding and complementary DNA strands based on the presence of open reading frames, intron sequences and splicing points especially in the human ε-globin gene. This implies an extremely complex intrinsic organization of genomes (von Sternberg 1996), where both the coding and the nc strands have precise information about the specific order of amino acids in proteins and/or other products such as small RNAs (sRNAs). This phenomenon of symmetric transcription has also been reported in viruses (Spiegelman *et al.* 1972; Aloni 1973) and mitochondrial DNA (Aloni and Attardi 1971; Tabak, Grivell and Borst 1983; Villegas *et al.* 2007). Similarly, Alff-Steinberger (1984) suggested that there is evidence for a coding pattern on the nc strand of the *Escherichia coli* genome, but Sharp (1985) concluded (based on a limited number of sequences from one bacteriophage, *Bacillus* spp. and *Saccharomyces cerevisiae*) that nc strands do not encode proteins. Other early papers include the observation that coding sequences with in-phase, 100% overlapping antisense open-reading frames (ORFs) are present in every genome from bacteria to humans (Merino *et al.* 1994), suggesting that this phenomenon is common among many life forms and therefore very ancient, while others shed light into the relationship between absence of stop codons and G+C content in ncDNA (Forsdyke 1995). More recent work has confirmed the presence of overlapping protein-coding sequences in several organisms such as *Chlamydia* (Jensen *et al.* 2006) and *Streptomyces* (Tunca *et al.* 2009), while others have proposed interesting evolutionary models to explain this phenomenon (Mir and Schober 2014).

16S rDNA evolution

The 16S rDNA is ~1500 nucleotides long and it contains conserved regions, meaning the nucleotide sequences in those regions are the same over most bacteria, including organisms with extreme differences in genome composition and phenotypic behavior, and also contains nine regions that vary considerably in nucleotide composition among bacteria. Today, it is generally believed that evolution of the 16S rDNA (coding strand) and other ribosomal RNA genes is dictated primarily by its fitness within the ribosomal translational machinery. If a difference is observed in a particular position in a nucleotide sequence alignment (e.g. transition or transversion), it is assumed that those

differences in nucleotides yield different secondary and tertiary structures therefore affecting translation (Cheng *et al.* 2012). This phenomenon (along evolution of ribosomal proteins and overall ribosomal architecture) has been called functional specialization of ribosomes and is supported by a growing number of publications (Gilbert 2008; Filipovska and Rackham 2013).

While different methodologies can be employed to study 16S rDNA evolution, public 16S databases offer a unique opportunity to do so massively in hundreds of different types of microorganisms from many diverse environments. We have used >300 000 rDNA sequences from the Ribosomal Database Project (RDP) to determine nucleotide composition and evolution of the entire phylum of Firmicutes (Garcia-Mazcorro and Cabrera-Castillo 2016), the most represented phylum in RDP. The organization of entries as well as the ease of the tools to download the data makes RDP a good option to do this type of analysis (Garcia-Mazcorro 2013). Briefly, the results show that on average 16S rDNA from Firmicutes possess more guanines (G, 31%) followed by adenines (A, 27%), cytosines (C, 22%) and thymines (T, 20%). On average, transitions outnumbered transversions by 2.7-fold; there were slightly more G↔A transitions (77/sequence) compared to C↔T transitions (74/sequence); and A↔T transversions were higher followed by G↔T, A↔C and C↔G. Most sequence alignments were explained by models such as the HKY (Hasegawa-Kishino-Yano) with a proportion of invariable sites accordingly to the Bayesian Information Criterion. It was during this analysis on 16S evolution that an intriguing idea emerged: Does 16S rDNA evolution reflect its own evolution or the evolution of its counterpart nc16S rDNA? The implications of raising this question are wide and may open new avenues for classifying 16S sequences into groups (e.g. classification by the types and numbers of functional transcripts encoded in the ncDNA strand as opposed to or in addition to nucleotide similarity).

The nc16S rDNA

We and others have used the 16S rDNA extensively to describe bacterial communities and their relationships, even at the global scale (Gilbert, Jansson and Knight 2014). However, there is no indication in the literature as to believe that we have considered the nc16S rDNA (even the early publications from Carl Woese do not mention the nc16S, see for example Woese and Fox 1977). One database of bacterial sRNA targets verified by experiments (sRNATarBase, Cao *et al.* 2010) contains a list of rRNAs sequences, some of which match bacterial 16S rDNA, but their origin is unclear (i.e. individual sequences are not associated with specific publications and a revision of the literature in the website did not reveal research related to ncrRNA). The objective of this minireview is to discuss the possibility that the nc16S rDNA code something as well as the potential contribution of this possibility into the use, analysis and evolution of the 16S rDNA. Additionally, here we present data suggesting that the nc16S rDNA contain code for small regulatory RNAs.

sRNAs are extremely common in bacteria (Argaman *et al.* 2001; Rivas *et al.* 2001; Wassarman *et al.* 2001; Chen *et al.* 2002; Vogel *et al.* 2003) and other organisms. Therefore, it is possible that nc16S rDNA (and perhaps many other unexplored nc regions) contains sRNAs targeting, for example, mRNAs. To explore this possibility, we used good quality almost full-length 16S rDNA from various bacterial groups from RDP (previously filtered out for unpublished research). These sequences (reverse complements) were uploaded into the Bacterial Small Regulatory RNA database (BSRD, Li *et al.* 2012). The BSRD contains sRNAs collected from over 783 bacterial species and 957 strains from several large public databases. Importantly, the BSRD and other databases only contain previously described sRNAs (i.e. it cannot predict new sRNAs). Intriguingly, we found that several nc16S rDNA sequences matched one ncRNA (name: lpr0011, GenBank accession number: AE017354.1) found in *Legionella pneumophila* (Gammaproteobacteria, Weissenmayer *et al.* 2011). The authors examined the gene expression of the sRNAs using the Cluster of Orthologous Groups of Proteins (Tatusov *et al.* 2000) but could not assign any function to this particular sRNA (available as Table S7, Supporting Information, in their publication); therefore, confirmation of the expression of this sRNA is needed using techniques such as northern blot or quantitative real-time PCR. Interestingly, the authors detected two additional sRNA (lpr0060 and lpr0066) with rRNA as potential targets. These sequences were expectably found in the BSRD but did not match any 16S rDNA in RDP. Interestingly, a tool for global target prediction for bacterial small nc RNAs (sTarPicker, Ying *et al.* 2011) was able to find many potential targets (including a good proportion of ribosomal proteins but not rRNA) for the sRNA using four different genomes from *Legionella* (Table S1, Supporting Information).

The sRNA from *Legionella* (lpr0011, AE017354.1) corresponds to a relatively long 77 nucleotide region representing ~5% of the whole nc16S rDNA. This sRNA encompasses the V5 region (29 nucleotide long) plus a 42 nucleotide-long semiconserved region downstream and a short 6 nucleotide region upstream on the 16S rDNA. Although this sRNA is distributed among different bacterial groups (of course, a region of the 16S rDNA), *Legionella* and relatives showed the highest similarity (Table 1). The V5 region has been shown to be one of the least heterogeneous (less variable) within bacterial genomes (Sun *et al.* 2013) although it shows high variability among bacterial groups and is actually considered not to be a good target for development of genus-specific probes in pathogenic bacteria (Chakravorty *et al.* 2007). Interestingly, here we show that this V5 region is actually the most conserved throughout the sRNA sequence among *Legionella* and relatives (Table 2) and that the entire sRNA from all related sequences form a structure with few loops and bulges and a relatively long apical loop (Fig. 2), all key secondary structure determinants of micro RNA (miRNAs) precursors at least in eukaryotes (Ritchie, Legendre and Gautheret 2007). Firmicutes sequences showed higher variability in nucleotide composition (Table 3) and predicted secondary structures (data not shown). A more extensive analysis of 16S rDNA sequences may provide a more comprehensive overview of this preliminary data (work currently in progress).

The finding that the nc16S V5 rDNA codes for an sRNA invariably leads to the question of whether it is possible that small fragments of nc16S rDNA could also target the mature 16S rRNA. This can have a direct effect on translation (albeit unspecific if it does not target mRNAs). While researchers in the field of RNA interference have focused mostly in the mRNAs targets of sRNAs, the existence of antisense RNAs inhibiting translation by direct competition with ribosomes has also been described (Darfeuille *et al.* 2007). A recent review elegantly summarizes the evidence for RNA interference happening within the ribosome (Pircher, Gebetsberger and Polacek 2014) but to our knowledge there has been no sRNA associated with ncrRNA. RNA interference on the ribosome (as opposed to interference on mRNAs) makes sense when one thinks that the ribosome evolved in an RNA world but mRNA-based RNA-seq approaches often treat any ribosome-bound ncRNAs as contaminants (hopefully this minireview would help others to start thinking otherwise, especially in light of new research showing that ncRNAs

Table 1. RDP sequence IDs from several bacterial groups and their corresponding matches in the BSRD (E value cutoff: 0.1).

RDP sequence IDs	Sequence matches in the BSRD (GenBank Accession numbers)
Proteobacteria	
S000272704 (*Legionella*)	AE017354.1 (E: 6e-18)
S000661227 (*Pseudoalteromonas*)	AE017354.1 (E: 6e-18)
S000573711 (*Haemophilus*)	AE017354.1 (E: 1e-16)
S000710606 (*E. coli/Shigella*)	AE017354.1 (E: 1e-15)
S001034662 (*Comamonas*)	AE017354.1 (E: 5e-15)
S000340260 (*Coxiella*)	AE017354.1 (E: 3e-13)
S003097645 (*Desulfonema*)	AE017354.1 (E: 6e-12)
S000858396 (*Campylobacter*)	AE017354.1 (E: 5e-12)
S000626456 (*Bradyrhizobium*)	AE017354.1 (E: 1e-9)
S002054181 (*Haematobacter*)	AE017354.1 (E: 2e-8)
Firmicutes	
S000762693 (*Allobaculum*)	AE017354.1 (E: 8e-5)
S000510673 (*Staphylococcus*)	AE017354.1 (E: 5e-4)
S000837439 (*Gemmella*)	AE017354.1 (E: 0.019)
S000818970 (*Anaerococcus*)	AE017354.1 (E: 0.019)
S000516422 (*Faecalibacterium*)	AE017354.1 (E: 0.071)
S000346335 (*Bacillus*)	None
S001494467 (*Dolosigranulum*)	None
S000709472 (*Enterococcus*)	None
S000536394 (*Streptococcus*)	None
S001456122 (*Veillonella*)	None
Actinobacteria	
S000565825 (*Propionibacterium*)	AE017354.1 (E: 2e-8)
S002053830 (*Janibacter*)	AE017354.1 (E: 5e-6)
S000378605 (*Streptomyces*)	AE017354.1 (E: 0.001)
S000623860 (*Gaiella*)	None
Bacteroidetes	
S000684818 (*Haliscomenobacter*)	AE017354.1 (E: 1e-9)
S000973582 (*Porphyromonas*)	AE017354.1 (E: 8e-5)
S001392228 (*Cloacibacterium*)	AE017354.1 (E: 7e-5)
S000344103 (*Prevotella*)	None

are frequently bound to ribosomes, Carlevaro-Fita *et al.* 2016). Importantly, these ncRNA 'entities' are not passive hitchhikers of the translation apparatus but appear to be an emerging type of nc riboregulators of protein biosynthesis (Pircher, Gebetsberger and Polacek 2014). We think that it is also important to add that RNA interference at the ribosome level does not necessarily have to target the mature 30S subunit; in fact, this binding may be less likely once everything is assembled. Therefore, this potential phenomenon of RNA interference could also happen right after transcription of the 16S rDNA (i.e. just before getting packed with the proteins in the 30S subunit similarly to RNA interference of mRNAs before they get translated).

Other sRNA databases could also help shed light into the nature of the nc16S rDNA. To this end, we used the same 16S rDNA sequences from RDP used for searching within the BSRD (Table 1). These sequences (reverse complements) were uploaded into the miRNA database (miRBase, Kozomara and Griffiths-Jones 2014), which contains data of miRNAs from animals, plants and viruses. It is important to note that in animals and other organisms a functional interaction between miRNA and its target RNA is thought to require only partial base pairing (Broughton and Pasquinelli 2016), thus providing miRNAs with wide regulatory potential. Interestingly, we found several matches between nc16S rDNA and miRNAs although the associated E values were high, suggesting random matches (Table S2, Supporting Information, note that the E values also relate to the size of the database). Regardless, some sequences were found to be very similar (1 nucleotide difference) and/or have comparable lower E values (Table S2, Supporting Information) and therefore we can conclude that non-bacterial miRNAs share resemblance to nc16S rDNA. Whether this represents a common ancient origin needs to be investigated.

Although small-interference RNAs (siRNAs) and miRNAs were originally considered distinct (siRNAs were thought to be the defenders of genome integrity in response to foreign nucleic acids, while miRNAs were thought to be the main regulators of gene expression), it is becoming increasingly difficult to discern between the two (Carthew and Sontheimer 2009). Importantly, in all cases the identities of the genetic fragments to be silenced are specified by the nucleotide composition of the sRNA, which recognizes its target by perfect or imperfect Watson–Crick base pairing. Any match between eukaryotic siRNAs and the 16S rDNA can thus have important implications. To explore similarities between nc16S rDNA and siRNAs, we used a subset of randomly sequences from the 991 120 siRNA sequences (~20 nucleotides each) available at the siRNAdb (Chalk *et al.* 2005). This database consists of eukaryotic siRNAs only and siRNAs are thought to act by perfect matching with mRNAs; therefore, eukaryotic siRNAs should not share any resemblance to

Table 2. Similarities among representative sRNAs from Proteobacteria.

Nucleotide sequence (5′-3′, corresponding to the nc16S rDNA)	Sequence IDs
TCACAGATAACTTAATCAACCACCTACGCACCCTTTACGCCCAGTAATTCCGATTAACGCTCGCACCCTCCGTATTA	GenBank: AE017354.1
TCACATCTCGCTTAACAAACCGCCTGCGTACGCTTTACGCCCAGTAATTCCGATTAACGCTCGCACCCTCCGTATTA	S000661227 (*Pseudoalteromonas*)
TCACAACCAACTTACACAACAACGTACGTACCCTTTACGCCCAGTAATTCCGATTAACGCTTGCACCCTCCGTATTA	S000272704 (*Legionella*)
TCACACCTCACTTAAATAACCGCCTGCGTGCCCTTTACGCCCAGTTATTCCGATTAACGCTCGCACCCTCCGTATTA	S000573711 (*Haemophilus*)
TCACATCTGACTTAACAAACCGCCTGCGTGCGCTTTACGCCCAGTAATTCCGATTAACGCTCGCACCCTCCGTATTA	S000710606 (*E. coli/Shigella*)
TCACATCTGACTTAACAAACCGCCTGCGTGCGCTTTACGCCCAGTAATTCCGATTAACGCTCGCACCCTCCGTATTA	S001034662 (*Comamonas*)

Nucleotides in red represent differences compared to the sRNA from *Legionella* (GenBank: AE017354.1). Underlined sequence portions represent those portions that were detected in BSRD. The V5 region (based on the *E. coli* 16S, Baker, Smith and Cowan 2003) was highlighted (gray) for better visualization. These six sequences were used to predict secondary structures (Fig. 2).

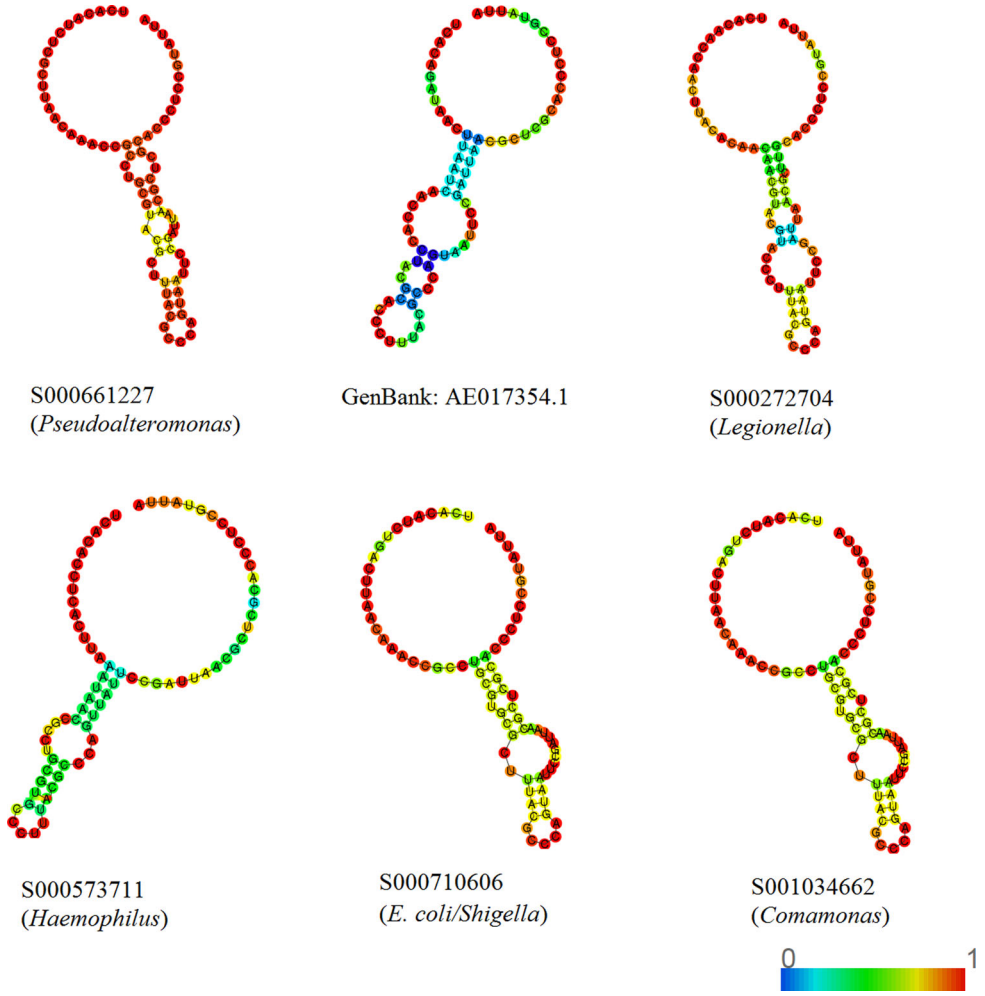

Figure 2. Secondary structures of a representative set of sRNAs from different bacterial groups using the Vienna RNA Websuite (Gruber *et al.* 2008). Please note that the structures are colored by base-pairing probabilities and that for unpaired regions the color denotes the probability of being unpaired.

Table 3. Similarities among representative sRNAs from Firmicutes.

Nucleotide sequence (5′-3′, corresponding to the nc16S rDNA)	Sequence IDs
TCACAGATAACTTAATCAACCACCTACGCACCCTTTACGCCCAGTAATTCCGATTAACGCTCGCACCCTCCGTATTA	GenBank: AE017354.1
TCACTTCCAACTTGTCTTCCCGCCTGCGCTCCCTTTACACCCAGTAATTCCGGACAACGCTTGTGACCTACGTTTTA	S000516422 (*Faecalibacterium*)
TAACTTCTAACTTGCTTACCCGCCTACGTACCCTTTACGCCCAATGATTCCGGACAACGCTCGGACCTTACGTATTA	S000818970 (*Anaerococcus*)
TCACATCAGACTTATTAAACCACCTGCGCGCGCTTTACGCCCAATAATTCCGGACAACGCTTGCCACCTACGTATTA	S000837439 (*Gemella*)
TCACATCAGACTTAAAAAACCGCCTACGCGCGCTTTACGCCCAATAATTCCGGATAACGCTTGCCACCTACGTATTA	S000510673 (*Staphylococcus*)
TCACTCCAGACTTGCAGGACCGCCTGCGCACCCTTTACGCCCAATCATTCCGGATAACGCTCGCCACCTACGTATTA	S000762693 (*Allobaculum*)

Nucleotides in red represent differences compared to the sRNA from *Legionella* (GenBank: AE017354.1). Underlined sequence portions represent those portions that were detected in BSRD. The V5 region (based on the *E. coli* 16S, Baker *et al.* 2003) was highlighted (gray) for better visualization.

bacterial ncrRNA (if they do, it would have important implications as this could imply ancient conservation of sRNAs). Expectably, no matches were found in RDP (with very few random exceptions) even though RDP allows to search imperfect matches (in other words, RDP would still be able to detect even weak matches). This means that, in contrast to miRNAs, some siRNAs are very specific to eukaryotes and do not pose any similarity whatsoever to bacterial ncrRNA.

Following the suggestions from one anonymous reviewer, we also used the same 16S rDNA sequences from RDP used for searching within the BSRD (Table 1) to search potential similarities within the Gene Expression Omnibus (GEO), a public repository of high-throughput functional genomic data from different organisms, including bacteria. A BLAST search within the GEO only revealed random matches with 16S sequences and a few transcripts from several plants.

Is the evolution of nc strands involved in the evolution of coding strands?

This question is hard to tackle. Mutations happen randomly but occur at different rates depending on a number of intertwined factors (Brison 2003). Generally, however, it is thought that only the coding regions get affected by natural selection (note that natural selection is not the only theory that explains evolution, see for example the paper by Hoelzer, Smith and Pepper 2006, on self-organization). Nonetheless, if a mutation occurs in the 16S rDNA and this mutation persists (i.e. get fixed) over time, then it would be fixed on the other complementary strand as well. This thinking also applies to the ncDNA. On the other hand, if only one strand would be affected by natural selection, then we would expect completely different rules applying to the two strands but this seems not to be the case. For example, it has been shown that both coding and nc regions of genes are governed by the same universal rule of TA/CG deficiency-TG/CT excess (Yomo and Ohno 1989). This symmetry is surprising because, again, it is thought that natural selection plays a role in the evolution of only one of the complementary strands. This and other early contributions (Ohno 1990; Zull and Smith 1990) indeed suggest that the initial acquisition of a function by a newborn gene (including regions for sRNAs) and entire genomes has to be, therefore, intrinsic in its construction (Yomo and Ohno 1989).

As discussed above, the possibility that ncrRNAs code something is intriguing but it poses complex issues to explore molecular evolution. The calculation of the relative number of mutations per site in both strands separately is particularly worrisome. Although this is in fact possible (Kaur and Makrigiorgos 2003; Tabone *et al*. 2006), the question of how molecular evolution happens *in vivo* (where both strands interact with each other, with other cellular components as well as any environmental condition) threat us to remain eternally. The bacterial genome duplicate by creating two copies of itself; therefore, each strand was thought to be equally exposed to random errors (Snyder and Champness 2007) although it is known that DNA replication does not occur with the same accuracy on both DNA strands, at least in *E. coli* (Fijalkowska *et al*. 1998). In fact, Schroeder *et al*. (2016) recently showed that local sequence context is the major determinant of mutagenesis in bacteria during DNA replication. A possible clue to the puzzle of simultaneous molecular evolution on both DNA strands may involve theoretical analysis based on the rate of transcription of each strand (e.g. is the transcription rate the same between rRNA and sRNAs? If not, does this relate to differences in the number of mutations and therefore to differences in mutation fixation and evolutionary rates?).

Final comments

The purpose of this work is to raise concerns about the nature, characteristics and implications of the nc16S rDNA, which is ignored by most scientists (including ourselves) working with the 16S rDNA. There is no reason to ignore sRNAs as those molecules are very common in all genomes (including bacterial) but to our knowledge no one has considered the possibility of having these sequences in the nc16S rDNA. Even recent throughout reviews on RNA interference in bacteria do not mention this possibility (Repoila and Darfeuille 2009; Pircher, Gebetsberger and Polacek 2014). Importantly, here we show that the nc16S rDNA from *Legionella* and relatives do show high non-random similarities with a published sRNA. The question of whether there are more sRNAs hiding in the nc16S (or other ncrRNA) needs further investigation. If the nc16S rDNA codes for sRNAs (ancient regulators of genetic expression), it would be interesting to experimentally investigate its potential targets. In particular, we are intrigued by two possibilities: that the sRNAs aim many (option 1) or just a few (option 2) targets (Fig. 1). Note that although the *Legionella* sRNA (lpr0011) matches a variable region (V5) of the 16S rDNA, this particular section is actually conserved but is accompanied by a variable region downstream on the 16S rDNA. It is likely that this variation gives this ancient sRNA the ability to target multiple mRNAs and other RNAs (generalist theory) as opposed to one or few specific types of RNA (specialist theory). This speculation is based on the fact that the so called trans-acting sRNAs are known to target multiple mRNAs via imperfect base pairing (Pircher, Gebetsberger and Polacek 2014) and is supported by the many potential targets predicted by sTarPicker (Table S1, Supporting Information).

Based on the results, discussions and speculations described above, the willingness of communicating microbiologists and other scientists about the nc16S rDNA is not unfounded and deserves closer consideration. Reports on antisense RNA expression are becoming more common in the literature, but validation and functional characterization is still lacking for most of them. This minireview shows preliminary data that the nc16S may code something but whether this is a general phenomenon among bacteria needs further scrutiny. It is also important for the reader to be aware that some sRNA molecules are the result of non-specific transcription with no physiological function (Thomason and Storz 2010). Nonetheless, we are convinced that researching the nc16S rDNA can help us all to better understand the nature of the 16S rDNA, its performance and functional specialization during translation, its use in molecular phylogeny and, ultimately, the beauty of bacterial biology.

SUPPLEMENTARY DATA

Supplementary data are available at FEMSLE online.

ACKNOWLEDGEMENTS

The authors thank deeply four anonymous reviewers for useful intellectual feedback on our original manuscript. We dedicate this manuscript to the memory of Carl Richard Woese.

Conflict of interest. None declared.

REFERENCES

Acinas SG, Marcelino LA, Klepac-Ceraj V *et al*. Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. *J Bacteriol* 2004;**186**:2629–35.

Alff-Steinberger C. Evidence for a coding pattern on the noncoding strand of the *E. coli* genome. *Nucleic Acids Res* 1984;**12**:2235–41.

Aloni Y. Poly A and symmetrical transcription of SV40 DNA. *Nat New Biol* 1973;**243**:2–6.

Aloni Y, Attardi G. Symmetrical *in vivo* transcription of mitochondrial DNA in HeLa cells. *P Natl Acad Sci USA* 1971;**68**:1757–61.

Argaman L, Hersberg R, Vogel J *et al*. Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol* 2001;**11**:941–50.

Baker GC, Smith JJ, Cowan DA. Review and re-analysis of domain-specific 16S primers. *J Microbiol Meth* 2003;**55**:541–55.
Brison D. The directed mutation controversy in an evolutionary context. *Crit Rev Microbiol* 2003;**29**:25–35.
Broughton JP, Pasquinelli AE. A tale of two sequences: microRNA-target chimeric reads. *Genet Sel Evol* 2016;**48**:31.
Bustamante CD, Fledel-Alon A, Williamson S *et al*. Natural selection on protein-coding genes in the human genome. *Nature* 2005;**437**:1153–7.
Cao Y, Wu J, Liu Q *et al*. sRNATarBase: A comprehensive database of bacterial sRNA targets verified by experiments. *RNA* 2010;**16**:2051–7.
Carlevaro-Fita J, Rahim A, Guigó R *et al*. Cytoplasmic long noncoding RNAs are frequently bound to an degraded at ribosomes in human cells. *RNA* 2016;**22**:867–82.
Carthew RW, Sontheimer EJ. Origins and Mechanisms of miRNAs and siRNAS. *Cell* 2009;**20**:642–55.
Casino A, Cipollaro M, Guerrini AM *et al*. Coding capacity of complementary DNA strands. *Nucleic Acids Res* 1981;**9**:1499–518.
Chakravorty S, Helb D, Burday M *et al*. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Meth* 2007;**69**:330–9.
Chalk AM, Warfinge RE, Georgii-Hemming P *et al*. siRNAdb: a database of siRNA sequences. *Nucleic Acids Res* 2005;**33**: D131–4.
Chen S, Lesnik EA, Hall TA *et al*. A bioinformatics based approach to discover small RNA genes in the *Escherichia coli* genome. *Biosystems* 2002;**65**:157–77.
Cheng N, Mao Y, Shi Y *et al*. Coevolution in RNA molecules driven by selective constraints: evidence from 5S rRNA. *PLoS One* 2012;**7**:e44376.
Darfeuille F, Unoson C, Vogel J *et al*. An antisense RNA inhibits translation by competing with standby ribosomes. *Mol Cell* 2007;**26**:381–92.
Fijalkowska IJ, Jonczyk P, Tkaczyk MM *et al*. Unequal fidelity of leading strand and lagging strand DNA replication on the Escherichia coli chromosome. *P Natl Acad Sci USA* 1998;**95**:10020–5.
Filipovska A, Rackham O. Specialization from synthesis: how ribosome diversity can customize protein function. *FEBS Lett* 2013;**587**:1189–97.
Forsdyke DR. Sense in antisense? *J Mol Evol* 1995;**41**:582–6.
Garcia-Mazcorro JF. Testing evolutionary models to explain the process of nucleotide substitution in gut bacterial 16S rRNA gene sequences. *FEMS Microbiol Lett* 2013;**346**:97–104.
Garcia-Mazcorro JF, Cabrera-Castillo D. *Nucleotide Composition and Substitution Patterns of 16S rRNA Gene Sequences Across The Entire Firmicutes Phylum*. ISME16, Montreal, Canada, 2016 (abstract accepted for publication).
Gilbert WV. Functional specialization of ribosomes? *Trends Biochem Sci* 2008;**36**:127–32.
Gilbert JA, Jansson JK, Knight R. The Earth Microbiome project: successes and aspirations. *BMC Biol* 2014;**12**:69.
Grassé PP. *Evolution of Living Organisms*. New York: Academic Press, 1977, 235.
Gruber AR, Lorenz R, Bernhart SH *et al*. The Vienna RNA websuite. *Nucleic Acids Res* 2008;**36**:W70–4.
Hoelzer GA, Smith E, Pepper JW. On the logical relationship between natural selection and self-organization. *J Evol Biol* 2006;**19**:1785–94.
Jaspers E, Overmann J. Ecological significance of microdiversity: identical 16S rRNA gene sequences can be found in bacteria with highly divergent genomes and ecophysiologies. *Appl Environ Microb* 2004;**70**:4831–9.
Jensen KT, Petersen L, Falk S *et al*. Novel overlapping coding sequences in *Chlamydia trachomatis*. *FEMS Microbiol Lett* 2006;**265**:106–17.
Jensen S, Frost P, Torsvik VL. The nonrandom microheterogeneity of 16S rRNA genes in *Vibrio splendidus* may reflect adaptation to versatile lifestyles. *FEMS Microbiol Lett* 2009;**294**: 207–15.
Kaur M, Makrigiorgos GM. Novel amplification of DNA in a hairpin structure: towards a radical elimination of PCR errors from amplified DNA. *Nucleic Acids Res* 2003;**31**:e26.
Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 2014;**42**:D68–73.
Li L, Huang D, Cheung MK *et al*. BSRD: A Repository for Bacterial Small Regulatory RNA. *Nucleic Acids Res* 2012;**41**: D233–8.
Lopez-Lopez A, Benlloch S, Bonfa M *et al*. Intragenomic 16S rDNA divergence in *Haloarcula marismortui* is an adaptation to different temperatures. *J Mol Evol* 2007;**65**:687–96.
Lukjancenko O, Wassenaar TM, Ussery DW. Comparison of 61 sequenced *Escherichia coli* genomes. *Microb Ecol* 2010;**60**: 708–20.
Mir K, Schober S. Selection pressure in alternative reading frames. *PLoS One* 2014;**9**:e108768.
Merino E, Balbas P, Puente JL *et al*. Antisense overlapping open Reading frames in genes from bacteria to humans. *Nucleic Acids Res* 1994;**22**:1903–8.
Ohno S. Grammatical analysis of DNA sequences provides a rationale for the regulatory control of an entire chromosome. *Genet Res* 1990;**56**:115–20.
Pircher A, Gebetsberger J, Polacek N. Ribosome-associated ncRNAs: An emerging class of translation regulators. *RNA Biol* 2014;**11**:1335–9.
Repoila F, Darfeuille F. Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. *Biol Cell* 2009;**101**:117–31.
Ritchie W, Legendre M, Gautheret D. RNA stem-loops: to be or not to be cleaved by RNAse III. *RNA* 2007;**13**:457–62.
Rivas E, Klein RJ, Jones TA *et al*. Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr Biol* 2001;**11**:1369–73.
Sanchez-Perez G, Mira A, Nyiro G *et al*. Adapting to environmental changes using specialized paralogs. *Trends Genet* 2008;**24**:154–8.
Schroeder JW, Hirst WG, Szewczyk GA *et al*. The effect of local sequence context on mutational bias of genes encoded on the leading and lagging strands. *Curr Biol* 2016;**26**: 692–7.
Sharp PM. Does the 'non-coding' strand code? *Nucleic Acids Res* 1985;**13**:1389–97.
Snyder L, Champness W (eds). The bacterial chromosome: DNA structure, replication, and segregation. In: *Molecular Genetics of Bacteria*, 3rd edn, Washington, DC: ASM Press, 2007.
Spiegelman WG, Reichardt LF, Yaniv M *et al*. Bidirectional transcription and the regulation of phage λ repressor synthesis. *P Natl Acad Sci USA* 1972;**69**:3156–60.
Sun DL, Jiang X, Wu QL *et al*. Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. *Appl Environ Microb* 2013;**79**:5962–9.
Tabak HF, Grivell LA, Borst P. Transcription of mitochondrial DNA. *CRC Cr Rev Bioch* 1983;**14**:297–317.
Tabone T, Sallmann G, Chiotis M *et al*. Chemical cleavage of mismatch (CCM) to locate base mismatches in heteroduplex DNA. *Nat Protoc* 2006;**1**:2297–304.

Tatusov RL, Galperin MY, Natale DA *et al*. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 2000;**28**:33–6.

Thomason MK, Storz G. Bacterial antisense RNAs: how many are there and what are they doing. *Annu Rev Genet* 2010;**44**: 167–88.

Tunca S, Barreiro C, Coque J-JR *et al*. Two overlapping antiparallel genes encoding the iron regulator DmdR1 and the Adm proteins control sidephore and antibiotic biosynthesis in *Streptomyces coelicolor* A3(2). *FEBS J* 2009;**276**:4814–27.

Villegas J, Burzio V, Villota C *et al*. Expression of a novel noncoding mitochondrial RNA in human proliferating cells. *Nucleic Acids Res* 2007;**35**:7336–47.

Vogel J, Bartels V, Tang TH *et al*. RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res* 2003;**31**:6435–43.

von Sternberg R. The role of constrained self-organization in genome structural evolution. *Acta Biotheor* 1996;**44**:95–118.

Wassarman KM, Repoila F, Rosenow C *et al*. Identification of novel small RNAs using comparative genomics and microarrays. *Gene Dev* 2001;**15**:1637–51.

Weissenmayer BA, Prendergast JGD, Lohan AJ *et al*. Sequencing illustrates the transcriptional response of *Legionella pneumophila* during infection and identifies seventy novel small non-coding RNAs. *PLoS One* 2011;**6**:e17570.

Wimberly BT, Brodersen DE, Clemons WM *et al*. Structure of the 30S ribosomal subunit. *Nature* 2000;**407**:327–39.

Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *P Natl Acad Sci USA* 1977;**74**:5088–90.

Yamamoto H, Wittek D, Gupta R *et al*. 70S-scanning initiation is a novel and frequent initiation mode of ribosomal translation in bacteria. *P Natl Acad Sci USA* 2016;**113**:E1180–9.

Ying X, Cao Y, Wu J *et al*. sTarPicker: a method for efficient prediction of bacterial sRNA targets based on a two-step model for hybridization. *PLoS One* 2011;**6**:e22705.

Yomo T, Ohno S. Concordant evolution of coding and noncoding regions of DNA made possible by the universal rule of TA/CG deficiency-TG/CT excess. *P Natl Acad Sci USA* 1989;**86**:8452–6.

Zull JE, Smith SK. Is genetic code redundancy related to retention of structural information in both DNA strands? *Trends Biochem Sci* 1990;**15**:257–61.