Identification of non-coding RNAs in environmental vibrios

Ana Cristina G. Silveira,1,2 Kelly L. Robertson,3 Baochuan Lin,3 Zheng Wang,3 Gary J. Vora,3 Ana Tereza R. Vasconcelos2 and Fabiano L. Thompson1

Correspondence
Fabiano L. Thompson
fabiano.thompson@biologia.ufrj.br

1Laboratory of Microbiology, Institute of Biology, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, CEP 21941-590, Brazil
2National Laboratory for Scientific Computing (LNCC), Petrópolis, RJ, Brazil
3Center for Bio/Molecular Science and Engineering, Naval Research Laboratory, Washington, DC 20375, USA

The discovery of non-coding RNA (ncRNA) has been mainly limited to laboratory model systems and human pathogenic bacteria. In this study, we begin to explore the ncRNA diversity in four recently sequenced environmental Vibrio species (Vibrio alginolyticus 40B, Vibrio communis 1DA3, Vibrio mimicus VM573 and Vibrio campbellii BAA-1116) by performing in silico searches using Infernal and Rfam for the identification of putative ncRNA-encoding genes. This search method resulted in the identification of 31–38 putative ncRNA genes per species and the total ncRNA catalogue spanned an assortment of regulatory mechanisms (riboswitches, cis-encoded ncRNAs, trans-encoded ncRNAs, modulators of protein activity, ribonucleoproteins, transcription termination ncRNAs and unknown). We chose to experimentally validate the identifications for V. campbellii BAA-1116 using a microarray-based expression profiling strategy. Transcript hybridization to tiled probes targeting annotated V. campbellii BAA-1116 intergenic regions revealed that 21 of the 38 predicted ncRNA genes were expressed in mid-exponential-phase cultures grown in nutrient-rich medium. The microarray findings were confirmed by testing a subset of three highly expressed (6S, tmRNA and TPP-2) and three moderately expressed (CsrB, GcvB and purine) ncRNAs via reverse transcription PCR. Our findings provide new information on the diversity of ncRNA in environmental vibrios while simultaneously promoting a more accurate annotation of genomic intergenic regions.

INTRODUCTION

One of the most recently elucidated molecular mechanisms for the regulation of bacterial gene expression has been the discovery of post-transcriptional regulation by means of non-coding RNAs (ncRNAs). These are small RNA transcripts (typically 50–250 nt) that are not translated into proteins. Bacterial ncRNAs were recognized more than two decades ago and were mainly associated with the regulation of replication of extrachromosomal elements (Wagner & Brantl, 1998). Only recently have ncRNAs been identified as key regulatory elements of bacterial chromosomes that are responsible for the regulation of critical processes such as stress response (Gottesman et al., 2006) and pathogenicity (Murphy & Payne, 2007). Some types of ncRNA (e.g. trans-encoded ncRNA) may regulate multiple genes and processes, thus having a profound effect on overall cell physiology.

Bacterial ncRNAs are a heterogeneous group of molecules that act by various mechanisms to modulate a wide range of cellular responses. Some ncRNAs have a structural function [e.g. transfer–messenger RNA (tmRNA)] whereas others have a regulatory function. Examples of known regulatory ncRNA types are (i) riboswitches, which are part of the mRNAs that they regulate, (ii) ncRNAs that bind to proteins, and (iii) two classes (cis and trans) that modulate the translation and stability of mRNAs. The first class is encoded in cis on the DNA strand opposite the target RNA, and the second type is encoded elsewhere (Waters & Storz, 2009). The first studies on ncRNAs were restricted to

Abbreviations: ncRNA, non-coding RNA; tmRNA, transfer–messenger RNA.

The microarray data described have been deposited in the NCBI Gene Expression Omnibus database with the accession number GSE20960.

Two supplementary tables, listing primers used for RT-PCR and ncRNAs found in the four Vibrio genomes examined, and text files with the DNA sequences of all ncRNAs found in this study are available with the online version of this paper.
**METHODS**

**In silico ncRNA identification.** The four nearly completely sequenced genomes of *V. mimicus* VM573, *V. alginolyticus* 40B, *V. communis* IDA3 (=LMG 25433) and *V. campbellii* ATCC BAA-1116 were analysed. Strain ATCC BAA-1116 (also known as *V. harveyi* strain BB120) has recently been identified as *V. communis* 1DA3 (Lin et al., 2010), whereas strain IDA3 has been characterized as the new species *V. campbellii* (Chimetto et al., 2010). The first three genomes were annotated using the System Automated Bacterial Integrated Annotation (Almeida et al., 2004; Thompson et al., 2009), whereas the last was obtained directly from the NCBI (GenBank; ftp://ncbi.nlm.nih.gov/genomes/Bacteria). The strains of *V. communis* and *V. alginolyticus* originated from Brazilian corals and represented the most frequently found species of vibrios, while the *V. mimicus* strain (which harbours the cholera toxin and toxin co-regulated pilus genes) was isolated from a patient in the USA in the 1990s. The search for ncRNAs was performed using the software Infernal (INFERence of RNA ALignment) (http://infernal.janelia.org) and the Rfam database (ftp://ftp.sanger.ac.uk/pub/databases/Rfam/CURRENT/). Rfam is a database of structure-annotated multiple sequence alignments, using covariance models and family annotation for a number of non-coding RNA families (Griffiths-Jones et al., 2003, 2005), whereas Infernal is based on the RNA secondary structure and sequence similarities to previously annotated ncRNAs. It is an implementation of a special case of profile stochastic context-free grammars called covariance models (CMs). A CM is like a sequence profile, but it scores a combination of sequence consensus and RNA secondary structure consensus (Eddy & Durbin, 1994). We used the cmsearch program within Infernal to find ncRNAs in the intergenic sequences (IGs) of the four vibrio genomes. Intergenic regions longer than 50 nt were extracted using a perl script. Sequences corresponding to tRNA and rRNA were excluded. In order to reduce the computational time for searching the ncRNAs with Rfam, we preprocessed the Rfam database in order to eliminate the CMs of eukaryotes. All of the IGs from each of the four genomes were compared with the preprocessed Rfam database using cmsearch. Subsequently, the putative ncRNAs were located in the genomes, and the upstream and downstream ORFs were identified. For stringency, the Infernal cmsearch score cutoff was 34, indicating that the candidate ncRNAs have a score >34 (Navrocki et al., 2009; http://infernal.janelia.org).

**Microarray-based expression profiling.** Five cultures of *V. campbellii* BAA-1116 were grown in Luria marine medium at 30 °C in a shaking incubator (200 r.p.m.) for 12 h. Total RNA extraction on 1 × 10^8 cells was carried out using the RiboPure-Bacteria kit (Ambion) according to the manufacturer’s recommended protocol with a minor modification (i.e. the use of 55% ethanol instead of 35% ethanol to ensure that small ncRNAs were not lost during the purification procedure). All extracted total RNA was treated with DNase I to remove trace amounts of genomic DNA. The quality and quantity of the RNA samples were determined by FlashGel RNA reagent (Affymetrix), 1 μg and diluted polyA RNA control (Affymetrix) using the RiboPure-Bacteria kit (Ambion) according to the manufacturer’s protocol.

mRNA labelling was carried out in 30 μl reaction volumes containing 1 μg mRNA and diluted polyA RNA control (Affymetrix) using the KREATech Amplified RNA (aRNA) Labelling kit (KREATech Biotechnology). Labelled mRNA was spin purified and fragmented in 33 μl reaction volumes containing 1 × fragmentation buffer (Ambion) and incubated at 70 °C for 15 min.

Non-coding RNA expression profiles were determined by hybridizing the fragmented and labelled mRNA to custom *V. campbellii* BAA-1116 whole-genome microarrays (520694F, Affymetrix) containing 25-mer probes every 15–25 nt for all annotated intergenic regions. Each targeted 25-mer sequence is represented by a sense-strand probe and its reverse complement antisense-strand probe. Briefly, the 520694F arrays were incubated with 120 μl pre-hybridization buffer containing 10 mM Tris, pH 7.8, and 0.01% Tween-20 for 10 min at 49 °C. Then 117 μl hybridization cocktail master mix [100 mM MES, 1 M NaCl, 20 mM EDTA, 0.01% Tween-20, 0.5 mg BSA ml^-1, 0.1 mg herring sperm DNA ml^-1, 0.5 μg herring sperm RNA ml^-1] (Promega), 1 × Oligo control reagent (Affymetrix), 1 × hybridization control and biotin-labelled mRNA fragments were equilibrated for 5 min at 49 °C and added to the 520694F arrays. All hybridizations were allowed to incubate for
16 h at 49 °C in a GeneChip Hybridization Oven 640 (Affymetrix) at 60 r.p.m. Post-hybridization, the arrays were washed and stained with the GeneChip Fluidics Station 450 (Affymetrix) and scanned using the GeneChip Scanner 7G (Affymetrix). The hybridization intensities were analysed with the GeneChip Operating Software (GCOS) to generate raw image files (.DAT) and simplified image files (.CEL) with intensities assigned to each of the corresponding probe positions. All described microarray data have been deposited in the NCBI Gene Expression Omnibus database (accession no. GSE20960).

RT-PCR. Overnight cultures of V. campbellii BAA-1116 were diluted 1:5000 in Luria marine medium (time 0) and grown in a 30 °C shaking incubator for 24 h at 200 r.p.m. Aliquots were withdrawn every 6 h for small ncRNA extraction using the mirVana miRNA Isolation kit (Ambion). The purified and enriched ncRNA samples were subsequently treated with DNase I to ensure the absence of contaminating genomic DNA. RT-PCR assays were conducted on an iCycler (Bio-Rad) using the OneStep RT-PCR kit (Qiagen). RT-PCR mixtures consisted of 1× Qiagen OneStep RT-PCR buffer, 1× OneStep RT-PCR enzyme mix, 200 nM primers (see Supplementary Table S1, available with the online version of this paper) and 0.1–25 ng ncRNA from each extracted sample, and were subjected to the following cycling conditions: one cycle at 50 °C for 30 min and 95 °C for 15 min, followed by 40 cycles of 94 °C for 30 s, 54 °C for 30 s and 72 °C for 60 s and a final elongation step for 10 min at 72 °C. All amplicons were electro-phoresed on a 1.5% agarose gel in TAE buffer and stained with ethidium bromide.

RESULTS AND DISCUSSION

Diversity of ncRNAs in vibrio genomes

The in silico search identified 140 putative ncRNAs in the four Vibrio genomes (31 in V. alginolyticus 40B, 38 in V. campbellii BAA-1116, 38 in V. communis 1DA3 and 33 in V. mimicus VM573; see Supplementary Table S2). The size of the ncRNAs varied between 56 and 402 nt. Eighty-seven per cent of the identified ncRNAs (n=122) have a size between 50 and 250 nt. The number of ncRNAs in the smaller chromosome of these vibrios (n=9–12) is significantly lower (P<0.01) than in the larger chromosome (n=20–27). The ncRNAs identified were distributed into seven types of regulatory mechanisms. In total, 46 ncRNAs were classified as riboswitches, 30 as trans-encoded, 22 as modulators of protein activity, nine as cis regulators, four as ribonucleoproteins and nine as transcription terminators; 20 were of an unknown mechanism(s). The DNA sequences of all ncRNAs found in this study are listed in the supplementary material.

Expression of predicted ncRNAs in V. campbellii BAA-1116

The in silico analysis identified 38 putative ncRNAs in V. campbellii BAA-1116. We attempted to detect the expression of these ncRNAs using a V. campbellii BAA-1116 whole-genome microarray (520694F, Affymetrix) that contains sense and anti-sense strand probe pairs that are spaced at regular intervals within the intergenic regions of the BAA-1116 genome. Of the 38 putative ncRNAs, eight were identified as the ncRNA intron_gpII. Due to the sequence identity amongst the eight intron_gpII genes, all eight genes were represented by only one tiled unique probe set and as such we would be able to detect a maximum of 31 ncRNAs using this method. Five of the 31 ncRNAs were the previously described quorum-sensing regulatory RNAs (Qrr1–5) (Lenz et al., 2004; Tu & Bassler, 2007). The hybridization of RNA samples from mid-exponential-phase cultures revealed that 21 of the 31 predicted ncRNA genes were expressed during this physiological state (Fig. 1). Five of the predicted ncRNAs (6S, tmRNA, TPP-2, S15 and Alpha_RBS) demonstrated relatively high expression levels (Fig. 1a). Of particular note was the 6S ncRNA, which not only demonstrated the highest expression level among the targeted ncRNAs but was also one of the most highly expressed transcripts in the entire genome (data not shown). Based on the calculated expression threshold and sense strand to anti-sense strand hybridization differential, 14 other ncRNA genes were also considered to be expressed during mid-exponential-phase growth (Fig. 1b) while seven ncRNA genes (encoding the predicted glycine, lysine-1, lysine-2, Spot-42, FMN, MOCO and Yybp_ykoY ncRNAs) were not expressed (Fig. 1c).

Interestingly, of the five quorum-sensing regulatory RNAs (Qrr1–5), Qrr4 and Qrr2 were also found to be expressed during mid-exponential-phase growth (Fig. 1d). While this result coincides with previous findings that Qrr4 and Qrr2 are the two most highly expressed of the five qrr genes in low-cell-density states (with Qrr4>Qrr2) (Tu & Bassler, 2007), it also indicates that a basal level of transcription for both qrr4 and qrr2 is maintained in mid-exponential-phase cultures – a high-cell-density state in which inactivated (dephosphorylated) LuxO is unable to initiate the transcription of the qrr genes.

In order to confirm these findings, we generated enriched small ncRNA preparations from V. campbellii BAA-1116 cells harvested during lag phase (6 h), mid-exponential phase (12 h), late-exponential phase (18 h) and stationary phase (24 h) and tested for the presence of a subset of the predicted ncRNAs via RT-PCR (Fig. 2; primers listed in Supplementary Table S1). All six of the targeted ncRNA transcripts (6S, tmRNA, TPP-2, CsrB, GcvB and purine) that were detected via microarray expression profiling were also found to be expressed at the 12 h time point via RT-PCR. In addition, all six targeted ncRNA transcripts were found in all four phases of growth tested. Not only was the negative control Spot-42 ncRNA not expressed at the 12 h time point (as would be expected from the microarray data), but it was also not found to be expressed at any other tested time point. This was expected considering that Luria marine medium does not contain galactose and none of the galactose-utilization genes appeared to be modulated under the conditions tested (data not shown).
Biological significance of ncRNAs

A riboswitch is one of the simplest and most frequently found regulatory elements in bacteria. As part of an mRNA, it promotes its own regulation. The binding of small molecules to this leader sequence results in conformational changes in the mRNA. These changes may also be a response to cell and environmental stimuli (e.g., increased environmental temperature, nutrient depletion, stalled ribosomes) (Wassarman et al., 1999). The two *V. mimicus* TPP riboswitches found in this study were classified as phosphomethylpyrimidine kinases, which participate in the metabolism of thiamine, and a thiamine biosynthesis protein, ThiC. The TPP riboswitch, also called the Thi-box or THI element, binds directly to thiamine pyrophosphate (TPP; the active form of thiamine or vitamin B1) in bacteria and eukaryotes (Kubodera et al., 2003; Sudarsan et al., 2003; Winkler et al., 2002). The purine riboswitch regulates purine biosynthesis and transport by binding to guanine and adenine (Mandal et al., 2001).

---

**Fig. 1.** Mid-exponential-phase microarray expression profiles of 31 of the 38 identified ncRNAs in *V. campbellii* BAA-1116. (a) Highly expressed ncRNAs TPP-2, 6S, tmRNA, S15 and Alpha_RBS. Grey bars, sense-strand probes; white bars, anti-sense-strand probes. Asterisks denote signal intensity values <150 that cannot be seen on the plot. (b) Transcribed ncRNA species whose expression was above 2x the average background signal intensity threshold (dashed line) and demonstrated a signal intensity ratio >2.5 when compared with the non-hybridizing complementary probe. (c) ncRNAs that are not expressed during mid-exponential-phase growth in Luria marine medium. (d) Detected expression of quorum-sensing regulatory RNAs Qrr2 and Qrr4. Data shown represent means ± SD of five independent biological replicates.

**Fig. 2.** Time-course analysis of *V. campbellii* BAA-1116 ncRNA expression via RT-PCR. ncRNA preparations from lag phase (6 h), mid-exponential-phase (12 h), late-exponential-phase (18 h) and stationary-phase (24 h) cultures in Luria marine medium were subjected to RT-PCR to confirm the expression of six ncRNAs (6S, tmRNA, TPP-2, CsrB, GcvB and Purine). RT-PCRs were performed using 0.1 ng (6S, tmRNA and TPP-2), 1.0 ng (CsrB, GcvB and Purine) or 25 ng of each total small ncRNA preparation. Based on the microarray expression profiles, the Spot-42 ncRNA was chosen as a negative control.
et al., 2003). V. campbellii BAA-1116 and V. communis 1DA3 had a purine riboswitch located in the smaller chromosome in proximity to an adenosine deaminase gene. This gene encodes an enzyme that promotes the deamination of adenosine to form inosine (found in tRNAs), but it is not known if the expression of this enzyme is regulated by the purine riboswitch. Further work will be required to determine the relationship between these two genes.

trans-encoded ncRNAs were one of the most commonly found types of ncRNAs in the Vibrio genomes; particularly the Qrr subtype. Qrrs are involved in the regulation of quorum sensing along with Hfq (Lenz et al., 2004). Qrr and Hfq appear to interfere directly with the mRNA that encodes the regulators LuxR/HapR. In general, trans-encoded ncRNAs do not overlap the genes they regulate. As the region of complementarity between trans-encoded ncRNAs and their cognate mRNA is short and may lack specificity, trans-encoded ncRNAs may regulate more than one type of mRNA. The trans-encoded ncRNAs need the RNA chaperone Hfq to help in the base pairing with their target mRNAs. trans-encoded ncRNAs appear to be crucial for a wide range of cellular processes beyond quorum sensing, such as the regulation of rpoS, ompF and iron dependent proteins (mediated by the ncRNAs DsrA, MicF and Ryh, respectively). trans-encoded ncRNAs may inhibit or stimulate translation, depending on how and where they ligate with these target mRNAs (Wagner et al., 2002). For instance, by binding to the mRNAs, these ncRNAs may inhibit or block the access of ribosomes or even promote mRNA degradation by RNases. On the other hand, when they bind close to a translation initiation region (TIR) of target mRNA, they promote the opening of the mRNA and expose the TIR, allowing it to interact with ribosomes. The trans-encoded ncRNAs found in vibrios (GcvB) are upstream of thll (thiamine biosynthesis protein Thl) and downstream of the transcriptional regulator LysR family. In E. coli, GcvB is involved in the expression of transport systems for dipeptides and oligopeptides (Urbanowski et al., 2000). It was also shown to be a regulator of resistance to acid in E. coli, increasing the ability of the bacterium to grow in low pH (Jin et al., 2009). GcvB expression is activated by the adjacent gene gcvA (Urbanowski et al., 2000).

Non-coding RNAs that modulate protein activity also appear to be pivotal for a variety of cellular processes in vibrios. They may be involved in the control of carbon metabolism, the production of extracellular products, cell motility, biofilm formation, quorum sensing and pathogenesis (Mizuno et al., 1984). CsrB, 6S and tmRNA were expressed during different periods of the growth curve. The first two appeared highly expressed throughout the growth cycle. CsrB is a global regulatory molecule. In E. coli, CsrB modulates the activity of CsrA (Majdalani et al., 1998). This protein binds to RNA and regulates carbon utilization and motility. CsrB forms a complex with CsrA and RNA, inhibiting the degradation of glg mRNAs (coding for genes of glycogen metabolism). Another type of ncRNA that modulates protein activity is the 6S ncRNA. It associates with RNA polymerase and sigma factor 70 (σ^70) to repress the expression of σ^70-dependent genes during the stationary phase (Tetart & Bouche, 1992). tmRNA is a protein activity modulator with dual functions, acting as tRNA and mRNA. tmRNA forms a complex with the small protein B (SmpB), EF-Tu and the ribosomal protein S1. It recycles stalled ribosomes and facilitates the degradation of anomalous mRNAs (Moller et al., 2002). tmRNA was found in each of the Vibrio genomes in this study and always downstream of smpB. This gene encodes an essential protein for the process of peptide-tagging by tmRNA (Karzai et al., 1999).

The fourth type of ncRNA found in the Vibrio genomes is the cis-encoded ncRNA (Yyb-p-ykoY). These types of ncRNAs are encoded in the same location as their targets, on the opposite strand (Wagner et al., 2002). They have regions of perfect-match base pairing (>75 nt) with their targets and one specific target for each ncRNA. The first cis-encoded ncRNAs were described in plasmids and little was known about their distribution in chromosomes until recently.

**Concluding remarks**

The abundance and diversity of ncRNAs in the genomes of vibrios suggest the importance of these molecules in cellular function and response to environmental stimuli (Liu et al., 2009). The present study demonstrates that environmental vibrios possess several different types of ncRNAs and that these ncRNAs are possibly responsible for the fine regulation of gene expression associated with different cellular responses and processes in vibrios.

**ACKNOWLEDGEMENTS**

A. C. G. S., A. T. R. V. and F. L. T. acknowledge the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), US Embassy, Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), and International Foundation for Science (IFS) for funding. G. J. V. acknowledges funding from the Office of Naval Research via US Naval Research Laboratory core funds.

**REFERENCES**


Vibri non-coding RNAs


Edited by: D. W. Ussery