Meta-Omics Approaches Reveal Unique Small RNAs Exhibited by the Uncultured Microorganisms Dwelling Deep-Sea Hydrothermal Sediment in Guaymas Basin

Muhammad Zohaib Nawaz

Shanghai Jiaotong University: Shanghai Jiao Tong University <https://orcid.org/0000-0003-2277-0113>

Fengping Wang (✉ fengpingw@sjtu.edu.cn)

Shanghai Jiao Tong University School of Life Sciences and Biotechnology

Research Article

Keywords: sRNA, deep-sea, uncultured microbes, ANME, mcrA, metagenome, metatranscriptome

Posted Date: December 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1185944/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Small regulatory RNAs (sRNAs) are present in almost all investigated microbes, regarded as modulators and regulators of gene expression and also known to play their regulatory role in the environmentally significant process. It has been estimated that less than 1% of the microbes in nature are culturable in the laboratory, hindering our understanding of their physiology, and living strategies. However, recent big advancing of DNA sequencing and omics-related data analysis makes the understanding of the genetics, metabolic potentials, even ecological roles of uncultivated microbes possible. In this study, we used a metagenome and metatranscriptome based integrated approach to identifying small RNAs in the microbiome of Guaymas Basin sediments. Hundreds of environmental sRNAs comprising of 228 groups were identified based on their homology, 82% of which displayed high similarity with previously known small RNAs in Rfam database, whereas, “18%” are putative novel sRNA motifs. A putative *cis*-acting sRNA potentially binding to methyl coenzyme M reductase, a key enzyme in methanogenesis or anaerobic oxidation of methane (AOM), was discovered in the genome of ANaerobic MEthane oxidizing archaea group 1 (ANME-1), which were the dominate microbe in the sample. These sRNAs were actively expressed in local Guaymas Basin hydrothermal environment, suggesting important roles of sRNAs in regulating microbial activity in natural environments.

Introduction

Small non-coding RNAs (sRNAs) are short untranslated transcripts (50-500 nucleotide [nt]) that are generally encoded in intergenic regions (IGRs) of the genomes. Most of the sRNAs act as regulators of the gene expression and could modulate expression both at the transcriptional and post-transcriptional level (Gottesman and Storz, 2011). Many of them appeared to regulate environmentally significant processes including quorum sensing (Fu et al., 2019), iron homeostasis (Chareyre and Mandin, 2018), amino acid and vitamin biosynthesis (Yang et al., 2017; Marinho et al., 2019), stress responses and photosynthesis (Brenes-Alvarez et al., 2016). Previous sRNA identification studies have mainly been carried out on and remain limited to a few model microorganisms, so little is known about their diversity and ecological relevance in environmental microbial communities of which more than 99% remain uncultivated (Lewis et al., 2010; Puspita et al., 2012; Solden et al., 2016). As most of the previous studies on sRNA discovery were limited to cultured microbes, environmental sRNAs derived from uncultured microbes largely remain unknown. Previous studies have reported sRNA identification in metagenomic data, deciphering that sRNAs are the abundantly existing type of RNA in almost all microorganisms. For example, Weinberg et al. reported that metagenomic samples contain diverse classes of environmental RNAs which previously remain unknown (Weinberg et al., 2009). In another study, they discovered more than hundred structured sRNA using comparative genomics based approaches from metagenomic and genomic data, including twelve metabolite sensing sRNAs, three of which were experimentally validated (Weinberg et al., 2010). Metatranscriptomic data has also been used to identify the sRNAs, for example, Murakami et al. analyzed the metatranscriptomic data of oceanfront deep-subsurface hot spring and identified more than 2000 novel sRNA (Murakami et al., 2012). Similarly, Shi et al. reveal unique sRNAs in

Ocean 's water column through metatranscriptomic analysis (Shi et al., 2009). The discovery of the sRNAs from the various environments containing uncultured microbes is indispensable to get insights into their occurrence and functional diversity. To study the phylogenies of uncultured microorganisms and establishing their population dynamics, metagenomics has come forward as a powerful technique overcoming the need for culturing the microbes. Metagenomic research could reveal genomic sequences and provide diversity analysis of microbial species. Previously, metagenome analysis has been performed in different environments ranging from deep-sea hydrothermal vents to human gut (Almeida et al., 2019; Forster et al., 2019; Gojobori et al., 2019; Haro-Moreno et al., 2019). A limitation associated with the metagenome research includes it doesn't provide any information regarding which genes are expressed under which conditions, therefore, could not reveal information about the functional activity of microorganisms and their adaptation mechanisms. Thanks to metatranscriptomic research, which could reveal gene expression profiles under different environmental conditions. Therefore, metagenomics and metatranscriptomics based integrated approaches could provide more insights into the microbial diversity, dynamics of their gene expression and the potential role of sRNAs in adaptation strategies of microbial species.

In this study, we use both the techniques for identifying active small regulatory RNAs in the Guaymas Basin hydrothermal environment. The Guaymas Basin is a semiclosed basin in the Gulf of California (Byrne and Emery, 1960) characterized by the hydrothermally active environment that includes anoxic sediments seeps and vent plumes, each exhibiting a wide range of temperatures. High biological productivity in the water and terrigenous input from the Mexican mainland and Baja California lead to rapid sulfide- and hydrocarbon-rich, anaerobic sediments accumulation of approximately 400 m in thickness (Simoneit and Lonsdale, 1982). Guaymas Basin represents unique features as compared to other hydrothermal vent sites as huge amounts of petroleum-like products, such as gasoline-range aliphatic and aromatic hydrocarbons, short-chain fatty acids, ammonia and residual polar asphaltic material are produced using pyrolysis. Based on high carbonate content, the vent fluid also exhibits unique chemistry, i.e. nearly neutral pH and release a vast quantity of methane (Welhan, 1988). Anaerobic methane-oxidizing archaea (ANME) which are the known active group therein (Wang et al., 2019), undergo sulfate-dependent, anaerobic oxidation of methane (AOM) to maintain the global flux of methane (Meyerdierks et al., 2005; Knittel and Boetius, 2009; Meyerdierks et al., 2010).

Here, we applied a comparative genomics approach to detecting sRNAs in uncultured deep-sea microbes using their metagenome and metatranscriptome, and also focused the methane metabolizing archaeal group belonging to phylum Euryarchaeota, ANMEs. Almost all known anaerobic methane-metabolizing archaea utilize methyl coenzyme M reductase (MCR) as the key enzyme in methanogenesis and AOM processes (Wongnate et al., 2016). MCR is highly conserved, and in particular, the gene encoding the alpha subunit of the enzyme complex (*mcrA*) has been utilized as a molecular marker gene for the detection and phylogenetic investigation of methane/alkane metabolizing Archaea (Wang et al., 2020). A plethora of novel sRNAs in deep-sea microbes and their metagenome and metatranscriptome were identified in this study, particularly we reported the presence of a putative *cis*-regulatory RNA at the 5'-end of *mcrA* in ANME-1, which is likely regulating the expression of this significant gene, suggesting its

important role in methane metabolism. To our knowledge, this is the first discovery of sRNA related to methane metabolism. Furthermore, nearly complete genomes of ANME-2a and ANME-2d were scanned to detect sRNAs using two independent comparative genomics-based approaches.

Materials And Methods

Sample Description, Metagenome and metatranscriptome data collection

The sediment sample used in this study was collected by push cores at a deep-sea vent site in Guaymas Basin (Gulf of California, Mexico) in October 2008 during cruise AT15-25 on *Alvin* dive 4460 (latitude 27° 0.71238 N, longitude 111° 24.3237 W, depth 2016 m). The sediment core 10 was used in this study, it was oily and covered with a white microbial mat. The temperature was measured with the *Alvin* temperature probe and increased from ~7-10°C at the surface to 31°C at 6cm sediment depth. Sediment layers at every 2 cm depth interval were sectioned and stored at -70°C and the sediment layer at 4-6 cm below the surface was used in this study for DNA isolation (Wang et al., 2019).

SDS based method was used for the DNA extraction from the sample and descriptions of metagenome library construction and sequencing has been reported earlier (He et al., 2016). Details of bioinformatics analysis performed on raw Hiseq metagenomic reads is given below. Descriptions of metatranscriptomic data including methods of RNA extraction, complementary DNA Synthesis, and Sequencing has been reported in the previous study (Wang et al., 2019). In detail, total RNA was extracted by an SDS-based method. Briefly, 0.5 g of subsampled sediment with six replicates was mixed with an equal weight of 0.1 mm diameter glass beads and 670 µl of extraction buffer (100 mM Tris-HCl, 100 mM Sodium EDTA, 100 mM Sodium Phosphate and 1.5 M NaCl, pH 8.0), and then homogenized with a tissue lyzer (Tissuelyser-48, Shanghai Jingxin, China) at 30 Hz for 30 s, with a 120 s interval for two cycles (4 °C). Fifty µl lysozyme (20 mg/ml) and 10 µl proteinase K (20 mg/ml) were added and incubated for 30 min in a 37 °C water bath. After incubation, 70 µl of 10% SDS was added and incubated at 55 °C for 30 min

with gentle mixing every 10 min. The supernatant was collected after 10 000 × g centrifugation for 10 min at 4 °C and transferred into a 2 ml tube. The supernatant was mixed with an equal volume of Phenol: chloroform: Iso-amyl alcohol (P:C:I, 25:24:1, vol/vol) pH 5.7. The aqueous phase was retained by centrifugation and precipitated with 0.6 vol of isopropyl alcohol and 0.3 M sodium acetate (pH 5.2) at 4 °C for 30 min. The tube was centrifuged at 16 000 × g for 20 min at 4 °C and the supernatant was removed carefully to avoid loss. Finally, the nucleic acid pellets were washed with pre-cooled 70% ethanol and resuspended in RNase free water. DNA

was removed by TURBO DNA-free Kit (Ambion, Thermo Fisher Scientific). RNA was purified with the RNeasy MinElute Kit (Qiagen). DNA removal was confirmed by negative amplification of archaeal and bacterial 16S rRNA genes. Extracted total RNA was quantified by Qubit 3.0 fluorometer with RNA HS Assay Kit (Thermo Fisher Scientific). A total of 0.5 µg RNA were

converted to double-stranded cDNA using random primers and the Maxima H Minus Double-Stranded cDNA Synthesis Kit (Thermo Fisher Scientific). Then, the purified DNA (4 µg in total) and reverse transcript cDNA from RNA (500 ng in total) samples were sent to the Beijing Genomic Institute (BGI) for paired-end sequencing by the Illumina HiSeq 2000 sequencing platform.

***De novo* assembly and binning of metagenomic sequences**

The raw shotgun sequencing metagenomic reads were trimmed with Sickle (<https://github.com/najoshi/sickle>) using the default parameters. Trimmed, paired-end DNA reads were assembled using IDBA-UD (Peng et al., 2012) with the following parameters: *mink* 52, *maxk* 92, *step* 8. Binning of assembled metagenomic sequences was initially performed using tetra-nucleotide frequencies signatures in ESOM using 5Kb as fragment length cutoffs (Dick et al., 2009). Ribosomal RNA (rRNA) genes were called with EMIRGE individually.

Metagenome Annotation

Protein coding sequences (CDS) were determined using Prodigal (Hyatt et al., 2010) with *-meta* parameter. For each predicted CDS, functional information was collected by comparing against NCBI NR, KEGG (Kanehisa and Goto, 2000), COG (Tatusov et al., 2003) and Pfam (Finn et al., 2014) with E-values $<10^{-5}$ and inspected manually.

Metatranscriptome Analysis

After filtering the low-quality and rRNA reads, the rest of the high quality reads were mapped to the metagenome-assembled contigs with TopHat v2.1.1 (Kim et al., 2013), and FPKM (expected fragments per kilobase of transcript per million fragments mapped) values were used to estimate the expression level of each gene by using Cufflinks 2.2.1 scripts (<http://cole-trapnell-lab.github.io/cufflinks/>). The transcriptome data used in this study has been uploaded to the public database NODE (<http://www.biosino.org/node>) under the ID “OED111522”, at this moment it is kept private.

Metagenome and metatranscriptome-based integrated approach to detect sRNA from Guaymas Basin hydrothermal sediment

Retrieving Non-redundant, highly expressed, non-rRNA, cDNA reads mapped to intergenic regions of the metagenome.

Low-quality cDNA reads were removed with sickle (<https://github.com/najoshi/sickle>) using the default parameters (Joshi and Fass, 2011) and cleaned cDNA reads were further scanned for the ribosomal RNA (rRNA) in Silva database (Quast et al., 2013) using SortMeRNA (Kopylova et al., 2012). Reads identified as rRNA were removed from the cDNA datasets. Resulting cDNA reads were mapped to metagenomic assembled contigs from the same sample using Bowtie2 (Langmead and Salzberg, 2012) using default parameters. Reads mapped to protein-coding genes were rejected, keeping those mapped to the intergenic regions of the metagenome. Duplicated cDNA reads were counted using CD-HIT (Fu et al.,

2012) and were subsequently removed. cDNA reads having more than a hundred copies were selected and these non-redundant, highly expressed, non-ribosomal, non-protein coding, cDNA reads mapped to intergenic region of metagenome from the same sample were analyzed further.

Clustering of non-rRNA and non-protein coding, cDNA reads and secondary structure prediction.

The selected non-ribosomal cDNA reads mapped to IGR (4157571 reads out of 145587140 reads) of the metagenome were clustered to form distinct groups based on their sequence identity. Reads were clustered using CD-HIT (Fu et al., 2012) and all matches that met a minimum cutoff of 90% sequence identity over 90% average sequence length were clustered into a separate group. In total, 550 clusters were obtained.

To find sRNAs in methane metabolizing archaea ANME belonging to phylum Euryarchaeota, a Perl script was used to generate only those clusters having at least one representative sequence from ANME, where 10 such clusters were obtained. Twenty random representative sequences from each cluster were picked and aligned using CLUSTALW (Larkin et al., 2007) and were subsequently subjected to RNA secondary structure prediction tool RNAz (Washietl et al., 2005). Positively predicted clusters by RNAz were analyzed and discussed further. The strategy used to identify sRNAs using metagenomic and metatranscriptomic data from Guaymas Basin has been demonstrated in Figure 1.

Molecular simulation

Molecular dynamics simulations were carried out with ff03 force field (Duan et al., 2003). The sRNA model was solvated within octahedron water box of TIP3P (Jorgensen et al., 1983) with a water thickness above 10 Å. Sodium ions were added for the charge neutralization. The solvated systems were subjected to 10,000-step minimization with the steepest descent method for the first 1000 cycles and then conjugate gradient method. After that these systems were gradually heated from 0 to 300 K for 50 ps controlled by the Langevin constrain with a collision frequency of 2 ps⁻¹. Next, the constant pressure and temperature (NPT) ensemble were applied for 50 ps equilibrium. Finally, 50ns production simulations were carried out under NPT ensemble, without any restraint. The Particle Mesh Ewald method (Darden et al., 1993) was employed to calculate long-range electrostatic interactions, and structural snapshots were flushed every 500 steps. All MD simulations were performed using the parallel version of pmemd implemented in AMBER 12 (Case et al., 2012). Root-mean-square deviation (RMSD)-based clustering was performed by average linkage clustering algorithms. The representative structures were extracted to describe the linear and looping regions of sRNA. The clustering and the routine RMSD were calculated by cpptraj in AMBER 12.

Interaction of common linear region on sRNA structure with mRNA

To decipher the RNA-RNA interaction, IntaRNA tool (Wright et al., 2014) was used and interaction of linear regions (LR) on sRNA with the *mcrA* mRNA was investigated.

Comparative genomics-based approaches for identifying sRNAs in ANME

The nearly complete genome of important uncultured methane metabolizing archaea ANME-2a (99.02% completed with 1.31% contamination using checkM) and ANME-2d (99.67% completed with 4.41% contamination using checkM) belonging to phylum Euryarchaeota have assembled in various metagenomic studies (Haroon et al., 2013; Wang et al., 2014). Nearly complete genomes of ANME-2a and ANME-2d were utilized in two independent comparative genomics-based approaches to find small regulatory RNAs in the present study. As most of the sRNAs identified up to date are located in the intergenic region, two different approaches were utilized to create the homologous intergenic dataset for sRNA identification. The combination of RNAz and RNA Infernal was applied to datasets created by both the approaches. RNAz detects evolutionarily conserved and thermodynamically stable RNA secondary structures in multiple sequence alignments, which is not only accurate as compared to other available tools but also efficient as well. Knowledge-based approaches, with homologues of identified sRNAs for profiling, can be used as complementary to *de novo* ones. RNA Infernal, one of the knowledge-based sRNA identification tools, together with a *de novo* tool RNAz were utilized in this study. Methods for creating datasets are discussed below.

Method I. Intergenic regions from both ANME-2a and ANME-2d genomes and other 26 genomes belonging to family Methanosarcinaceae (as reference genomes) were extracted and were aligned using genome comparison method named progressive MAUVE (Darling et al., 2009), which identifies conserved genomic regions, rearrangements and inversions in conserved regions, and the exact sequence breakpoints of such rearrangements across multiple genomes. Thousands of datasets of aligned conserved intergenic regions were obtained. However, datasets that were found to contain regions of either ANME-2a or ANME-2d or both were selected and aligned using CLUSTALW. RNAz was applied on these aligned datasets to find if they represent a stable secondary structure. Datasets that were predicted as sRNA by RNAz were divided into smaller fragments and were subjected to motif searching tool RNA Infernal. Sequences giving an infernal score higher than 10 were considered as sRNA, criteria used by Xu *et al.* (Xu et al., 2009). However, only 4 sequences in ANME-2a and one in ANME-2d having an infernal score equal to or greater than '10', were obtained and considered as high confidence sRNAs. To extend our sRNA scanning across the two genomes an alternate method (discussed below) was applied.

Method II. As regulatory motifs were mainly found to be encoded in intergenic regions (IGR) of the genome, IGR dataset was generated by extracting sequences both from 5' and 3'-untranslated regions (UTRs) of each gene in ANME-2a and ANME-2d, together with its orthologous IGR regions in all the other 26 genomes belonging to family Methanosarcinaceae used in this study. The method is explained in details in a study by Nawaz *et al.* 2017 (Nawaz et al., 2017). In line with the method I, a combination of RNAz and RNA infernal was applied to datasets generated using method II for sRNA identification.

Results And Discussion

sRNAs identified in deep-sea Guaymas Basin hydrothermal sediment

Use of metagenomic and metatranscriptomic data from the same sample provided an advantage for sRNA identification over identifying in any of this individual data. To our knowledge, this is the first metagenome and metatranscriptome-based integrated approach used for sRNA identification in deep-sea uncultured microbes. As illustrated in Fig. 1, the integrated approach for sRNA identification in metagenomic and metatranscriptomics for the Guaymas Basin sample resulted in 228 groups of sRNAs detected from metatranscriptomic reads (Supplementary Table S1). About 82% of them were found to have hits in the Rfam database (Nawrocki et al., 2015) while the rest of 18% were considered as having novel motifs which previously remained undiscovered. Many of the sequences belonging to different groups were predicted to have hits with microRNAs and CRISPERS. The 10 most abundant sRNAs detected in the metatranscriptome are listed in Table 1. Sequences from five groups including group number 100, 238, 240, 525 and 643 were found to have hits with well-known sRNA called carbon storage regulatory C (CsrC). CsrC is a post-transcriptional regulator of an important protein CsrA which have an important role in the regulation of central carbon flux, motility and biofilm formation (Argaman et al., 2001). Although, the length of sRNA reads here resembling with CsrC sRNA in Rfam is much shorter, i.e, 150 nt compared to 245 nt of the full length of sRNA. However, the predicted structure of this read was closely related to the structure of full-length CsrC (Figure S1). Results show that CsrC is highly expressed and therefore, likely playing an active role in Guaymas Basin microbes. Besides, Rfam hits of another sRNA involved in the regulation of iron-containing proteins 'FsrA' (Smaldone et al., 2012) was found with reads from two groups. Similarly, ydaO-yuaA which is known to act as cyclic-di-AMP sensing riboswitch (Nelson et al., 2013) was found to have hits with sequences belonging to four different groups in this study. In addition, cDNA reads from two groups were found to have hits with cyclic-di-GMP which is known to play its role in important mechanism such as biofilm formation (Valentini and Filloux, 2016). Therefore, cyclic-di-AMP/GMP riboswitches are also active RNA regulatory structures in Guaymas Basin. Rfam hits of reads belonging to 6 groups were found as SmY, which are thought to have a role in mRNA splicing (Jones et al., 2009). Reads from 5 groups were found having hits with RsmY with a possible role in quorum sensing (Nakatsu et al., 2019) while two groups were found having hits with RsmZ which act as protein sponge (Duss et al., 2014). Similarly, group-II-D1D4-2 which is known to exist in high salt conditions with a possible role in mRNA splicing (Seetharaman et al., 2006), was found as hits in Rfam for reads belonging to four different groups. Group 10 was found to have a read with Rfam hit as isrJ which have a role in host cell invasion (Gong et al., 2011) whereas reads in group 116 and 799 have hit as sraL, which have a role in the prevention of premature transcription termination (Silva et al., 2019), both isrJ and sraL are well-known post-transcriptional Hfq-binding sRNA (Zeng and Sundin, 2014) (Table 1). Details of Rfam hits for reads belonging to 228 groups of identified sRNAs are available in Supplementary Table S1.

Table 1
Abundant sRNAs detected in the metatranscriptome

Sr. #	sRNA	Groups having sRNA	Function	References
1	CsrC	100, 238, 240, 525, 643	Regulation of central carbon flux, motility and biofilm formation	(Argaman et al., 2001)
2	FsrA	50, 306	Regulation of iron-containing proteins	(Smaldone et al., 2012)
3	ydaO-yuaA	28, 102, 355, 808	Cyclic-di-AMP sensing riboswitch	(Nelson et al., 2013)
4	cyclic-di-GMP	839, 1050	Biofilm formation	(Valentini and Filloux, 2016)
5	SmY	74, 141, 246, 254, 274, 341,	mRNA splicing	(Jones et al., 2009)
6	RsmY	36, 45, 63, 817, 915	Quorum sensing	(Nakatsu et al., 2019)
7	RsmZ	110, 169	Protein sponge	(Duss et al., 2014)
8	group-II-D1D4-2	2, 71, 193, 1003	mRNA splicing	(Seetharaman et al., 2006)
9	isrJ	10	Host cell invasion	(Gong et al., 2011)
10	sraL	116, 799	Prevention of premature transcription termination	(Silva et al., 2019)

Putative *cis*-sRNA of *mcrA* in ANME-1

The sRNA identification from Guaymas Basin was also specifically targeted to ANMEs where only one group of sRNAs was identified to have a representative sequence from ANME-1. The identified sRNAs groups were analyzed manually. Here, we report the identification of cDNA read that overlap the 5'-end of methyl coenzyme M reductase subunit A (*mcrA*) in ANME-1, intergenic region (1 nt) and methyl coenzyme M reductase subunit G (*mcrG*). It is not surprising that sRNA overlap with the coding regions of the genes, as they are known to evolve from mRNA, random transcriptions and tRNAs (Gottesman and Storz, 2011). The stable secondary structure of this sequence determined using molecular dynamic tool Amber (Case et al., 2012) indicated that it is a stable and functional RNA. As *cis*-sRNAs are usually located at the 5'-end of their target gene (Weinberg et al., 2010), therefore, we suggest that this sRNA may regulate the *mcrA* gene in ANME-1. As transcription and translation occur simultaneously in bacteria, therefore, a high number of copies are not necessarily required for an sRNA to work at the level of transcription. High copy

number of this *cis*-sRNA (131 copies) indicates that it might act as a post-transcriptional regulator of *mcrA* gene expression. A functional RNA binds to other molecular structure through its linear region (Sharma et al., 2007; Peer and Margalit, 2011; Richter and Backofen, 2012), we select the common linear regions on all the possible structures of *cis*-sRNA to find its binding with mRNA. The linear region commonly exhibited by all the possible structures of sRNA was “CGUCCGCUUC” near the 3'-end of the sRNA (Position 129-138) (Figure S2).

To decipher the RNA-RNA interaction, IntaRNA tool (Wright et al., 2014) was used to investigate the interaction of linear regions (LR) on sRNA with the *mcrA* mRNA. *Cis*-sRNA was found to interact with its linear region (position 129-138) with the seed region of 7 bp to its mRNA. A region of 6–8 contiguous base pairs are generally involved in the sRNA-mRNA interaction (Gottesman and Storz, 2011). Description of the interaction of linear common region on sRNA structure with mRNA is shown in Supplementary Figure S3, which is also demonstrated in Supplementary Table S2.

Putative *cis*-sRNA which overlaps with the 5'-end of the *mcrA* gene and transcribed in the opposite direction of its target mRNA, it might act as a riboswitch. Previous results have shown that methane metabolism archaea were abundant in the samples studied here (Wang et al., 2019), suggesting a high methanotrophic activity herein. We suspect that this sRNA might act as methane sensing riboswitch and under high concentration of methane, it binds to mRNA (*mcrA*), increasing the stability of mRNA which ultimately leads towards enhanced translation of *mcrA* protein and therefore, entailed in elevated methanotrophic activity and vice versa. However, more experimental evidences are needed to verify the anti-*mcrA* function of the sRNA in the future.

sRNA in ANMEs

To investigate if this type of *cis*-sRNA is widely present in methanogens and methanotrophs, we blasted *cis*-sRNA against methanogens and methanotrophic genome, and nucleotide database at NCBI. One hit with the same region on ANME-1 recovered previously from Eel River sediments, was found and its position was exactly the same as identified *cis*-sRNA, i.e, upstream of *mcrA*, with 93% sequence identity. Its associated *mcrA* gene in our sample has also best hit with *mcrA* gene recovered in Hallam's study (Hallam et al., 2003) with 93% sequence identity. However, this sRNA was not found in any of the other methane metabolizing archaea. The possible case might exist that other methanogens and methanotrophs could use sRNA for the same purpose but don't share sequence identity.

To investigate further whether similar sRNA structure exists in other methane metabolizing archaea, we extracted the same region in almost all the sequenced genomes of the methane metabolizers and predict sRNA using same methods as used in this study. However, sRNA was not found in any of the methanogen and methanotrophic group. As small RNAs are appeared to evolve rapidly (Gottesman and Storz, 2011) and strain-specific sRNAs are also known to exist widely (Toffano-Nioche et al., 2012; Dugar et al., 2013), therefore, this *cis*-sRNA appeared as strain-specific sRNA regulating *mcrA*. ANMEs are not pure cultured yet in the laboratory. To test further if other sRNAs are present in ANMEs, the nearly complete genomes of ANME-2a and ANME-2d were screened as described in the Materials and Methods.

sRNAs predicted in ANMEs through Method I

Only 4 sequences in ANME-2a and one in ANME-2d having an infernal score equal to or greater than '10', were obtained and considered as high confidence sRNAs among all the datasets created.

sRNAs predicted in ANMEs through Method II

Initially, a total of 2051 sets of orthologous UTRs, of which 1175 from 5'-UTR and 876 from 3'-UTR, respectively, were extracted for sRNA prediction in ANME-2a. RNAz predicted 33 sets (1.6%, out of 2051) as sRNAs, with 25 and 8 from 5' and 3'-UTRs, respectively. After searching against Rfam database using Infernal, only 4 (0.2%, out of total datasets) RNA motifs in ANME-2a were predicted with an Infernal score higher than 10, thus considered as reliable sRNA candidates. Similarly, in total 1436 sets of orthologous UTRs, of which 872 from 5'-UTR and 564 from 3'-UTR, respectively, were extracted for sRNA prediction in ANME-2d. RNAz predicted 19 sets (1.3%, out of 3673) as sRNAs, with 13 and 6 from 5' and 3'-UTRs, respectively. After searching against the Rfam database using Infernal, none of the RNA motifs in ANME-2d was predicted with an Infernal score higher than 10.

Using both of the methods we obtained 6 non-redundant sRNA in ANME-2a and only one in ANME-2d. Identified sRNAs and their flanking genes in ANME-2a and ANME-2d genomes are shown in Figure 2. These sRNAs are flanked by the genes encoding archaeal flagellin N-terminal like domain, Aminopeptidase T, DisA (DNA integrity scanning protein), Epimerases, DUF772 (Domain of unknown function), 4Fe-4S dicluster domain, Superfamily II helicase, Orotate phosphoribosyltransferase, Cas3 (Crisper associated), Nucleic acid binding protein, and hypothetical proteins. Flagellin N-terminal like domain functions in recognition of flagellar export substrates (Chilcott and Hughes, 1998; Hirano et al., 2003; Vegh et al., 2006; Weber-Sparenberg et al., 2006; Stafford et al., 2007). Aminopeptidases are a group of metalloprotease which are involved in the cleavage of free amino acids from peptides and are considered as essential enzyme because they perform various vital functions in the cell including hydrolysis of regulatory peptides, protein turnover, protein maturation, nitrogen nutrition, and modulation of gene expression (Taylor, 1993; Matsui et al., 2006; Luckett et al., 2012). DisA functions in response to DNA-damage and stress homeostasis (Gandara and Alonso, 2015). Epimerases are involved in the synthesis of complex polymers of carbohydrate which constitutes the prokaryotic cell walls and envelopes (Allard et al., 2001). 4Fe-4S dicluster domain is present in electron transferring proteins in prokaryotes where they play their role in important biological processes including nitrogen fixation, oxidative phosphorylation, and photosynthesis (Beinert, 1990). Superfamily II helicase is involved in nucleic acid metabolism including DNA replication, DNA repair, transcription, translation, splicing and other related processes (Singleton et al., 2007; Lohman et al., 2008; Pyle, 2008; Fairman-Williams et al., 2010). Orotate phosphoribosyltransferase is key enzyme involved in the synthesis of pyrimidine synthesis (Ozturk et al., 1995). Cas3 is the associated protein of CRISPR/Cas system that provides acquired immunity against viruses and foreign nucleic acids (Sinkunas et al., 2011). As the *cis*-encoded sRNAs are involved in regulating the gene expression of their nearby genes (Brantl, 2007). The nearby genes of identified sRNAs in ANMEs suggest that these sRNAs might play their regulatory role in the

important biological process including bacterial movement, nucleic acid and amino acid metabolism, stress tolerance, electron transportation, and bacterial defence system. For most of the predicted sRNA, their best Rfam hits are microRNAs that functions in RNA silencing and post-transcriptional regulation of gene expression. sRNAs are present in prokaryotes and have miRNAs in their eukaryotic counterparts. As we annotated the identified sRNAs against Rfam database which contain RNA families of both the prokaryotes and eukaryotes. Identified sRNAs having hits as Splicing RNA or mir-11 shows these sRNAs exhibit functional or sequence similarity with their eukaryotic counterparts and previously remain undetermined in prokaryotes. Rfam hits of identified sRNAs in ANME-2a and ANME-2d are enlisted in Table 2.

Table 2
Rfam analysis of identified sRNAs in ANME-2a and ANME-2d, detected using their nearly complete genome

Sr. #	sRNA	Organism	Best hit	Texa	Function	Rfam
1	A2a1	ANME-2a	CRISPR-DR47	Prokaryotes	Immune system	RF01360
2	A2a2	ANME-2a	mir-944	Plants, animals, viruses	RNA silencing	RF01024
3	A2a3	ANME-2a	SNORA8	Eukaryotes	Modification of uridines to pseudouridines	RF00393
4	A2a4	ANME-2a	mir-1803	Plants, animals, viruses	RNA silencing	RF02094
5	A2a5	ANME-2a	mir-11	<i>Drosophila</i>	Apoptosis	RF00813
6	A2a6	ANME-2a	TtnuCD3	Eukaryotes	Chemical modifications of other RNAs	RF02280
7	A2d1	ANME-2d	CRISPR-DR2	Prokaryotes	Immune system	RF01315

Conclusions

Analysis of metagenomic and metatranscriptomic data lead to the identification of a massive number of sRNAs in Guaymas Basin (Gulf of California), demonstrating that microbes dwelling there exhibit diverse sRNAs which might have role adaptation. We identified 228 groups comprising of hundreds of sRNAs. Rfam analysis demonstrated that most of them (82%) were predicted to have similarity with already discovered sRNAs. However, 18% of these identified sRNAs were more likely to be novel. The strategy applied to detect sRNAs here was designed to discover sRNAs with high expression, i.e, with more than a

hundred copies of cDNA. High expression is required for sRNA only if it regulates the gene expression at the post-transcriptional level. Therefore, post-transcriptional regulators are the abundantly existing class of sRNA. The approach used here was particularly applied to focusing sRNAs in ANMEs. Interestingly, *cis*-acting sRNA of *mcrA* gene, a key enzyme in methanogenesis, and AOM processes were identified in ANME-1 that have been recovered from Guaymas Basin hydrothermal sediment sample. Small RNAs have previously been discovered to regulate almost every global response in prokaryotes. However, this is the first ever discovery of sRNA related to methanogen or methanotrophy. Furthermore, two independent comparative genomics approaches were used to identify sRNAs in nearly complete genomes of ANME-2a and ANME-2d (Haroon et al., 2013; Wang et al., 2014) where six and only one sRNA was detected respectively. Archaeal and bacterial sRNAs are almost structurally and functionally similar, however, the prediction of fewer sRNAs using comparative genomics-based approaches suggests that sRNAs in archaea are not evolutionary conserved as those in bacteria. Although functions of these sRNAs need to be investigated further using experimental approaches, this study provides the pioneering work on sRNA identification in ANME. Furthermore, our study investigated the active sRNAs in Guaymas Basin hydrothermal sediment sample and these sRNAs could be studied further to reveal their function therein.

Declarations

Acknowledgements

We are thankful to National Natural Science Foundation of China (Grant No. 31290232), China Ocean Mineral Resources R & D Association (Grant No. DY125-22-04) and National Natural Science Foundation of China (Grant No. 41576129, 91228201, 91428308) for supporting our research. Moreover, Higher Education Commission (HEC), Pakistan is gratefully acknowledged for providing financial support to M.Z.N. We are thankful to Lanxuan Liu for helping us in molecular dynamics simulations. We are also thankful to Yinzhaio Wang for useful discussions and valuable comments to improve the manuscript.

Author information

State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, People's Republic of China.

Muhammad Zohaib Nawaz, Fengping Wang

School of Oceanography, Shanghai JiaoTong University, Shanghai, 200240, People's Republic of China.

Fengping Wang

Author's contribution

F.W., and M.Z.N. designed the experiment and analysis. M.Z.N. designed the pipeline and performed the analysis. M.Z.N. and F.W. wrote the manuscript, in consultation with all M.Z.N.

Competing interests

The authors declare no competing financial interests.

Corresponding author

Correspondence to F. Wang.

References

1. Allard, S.T., Giraud, M.F., and Naismith, J.H. (2001). Epimerases: structure, function and mechanism. *Cell Mol Life Sci* 58(11), 1650-1665. doi: 10.1007/PL00000803.
2. Almeida, A., Mitchell, A.L., Boland, M., Forster, S.C., Gloor, G.B., Tarkowska, A., et al. (2019). A new genomic blueprint of the human gut microbiota. *Nature* 568(7753), 499-504. doi: 10.1038/s41586-019-0965-1.
3. Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., Wagner, E.G., Margalit, H., et al. (2001). Novel small RNA-encoding genes in the intergenic regions of Escherichia coli. *Curr Biol* 11(12), 941-950. doi: 10.1016/s0960-9822(01)00270-6.
4. Beinert, H. (1990). Recent developments in the field of iron-sulfur proteins. *FASEB J* 4(8), 2483-2491. doi: 10.1096/fasebj.4.8.2185975.
5. Brantl, S. (2007). Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr Opin Microbiol* 10(2), 102-109. doi: 10.1016/j.mib.2007.03.012.
6. Brenes-Alvarez, M., Olmedo-Verd, E., Vioque, A., and Muro-Pastor, A.M. (2016). Identification of Conserved and Potentially Regulatory Small RNAs in Heterocystous Cyanobacteria. *Front Microbiol* 7, 48. doi: 10.3389/fmicb.2016.00048.
7. Byrne, J.V., and Emery, K.O. (1960). Sediments of the Gulf of California. *Geological Society of America Bulletin* 71(7). doi: 10.1130/0016-7606(1960)71[983:Sotgoc]2.0.Co;2.
8. Case, D., Darden, T., Cheatham, T., Simmerling, C., Wang, J., Duke, R., et al. (2012). Amber 12 reference manual. *University of California, San Francisco*.
9. Chareyre, S., and Mandin, P. (2018). Bacterial Iron Homeostasis Regulation by sRNAs. *Microbiol Spectr* 6(2), 267-281. doi: 10.1128/microbiolspec.RWR-0010-2017.
10. Chilcott, G.S., and Hughes, K.T. (1998). The type III secretion determinants of the flagellar anti-transcription factor, FlgM, extend from the amino-terminus into the anti-sigma28 domain. *Mol Microbiol* 30(5), 1029-1040. doi: 10.1046/j.1365-2958.1998.01131.x.
11. Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: AnN-log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* 98(12), 10089-10092. doi: 10.1063/1.464397.
12. Darling, A.E., Mau, B., and Perna, N.T. (2009). Progressive Mauve: Multiple alignment of genomes with gene flux and rearrangement. *arXiv preprint arXiv:0910.5780*.

13. Dick, G.J., Andersson, A.F., Baker, B.J., Simmons, S.L., Thomas, B.C., Yelton, A.P., et al. (2009). Community-wide analysis of microbial genome sequence signatures. *Genome Biol* 10(8), R85. doi: 10.1186/gb-2009-10-8-r85.
14. Duan, Y., Wu, C., Chowdhury, S., Lee, M.C., Xiong, G., Zhang, W., et al. (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem* 24(16), 1999-2012. doi: 10.1002/jcc.10349.
15. Dugar, G., Herbig, A., Forstner, K.U., Heidrich, N., Reinhardt, R., Nieselt, K., et al. (2013). High-resolution transcriptome maps reveal strain-specific regulatory features of multiple *Campylobacter jejuni* isolates. *PLoS Genet* 9(5), e1003495. doi: 10.1371/journal.pgen.1003495.
16. Duss, O., Michel, E., Yulikov, M., Schubert, M., Jeschke, G., and Allain, F.H. (2014). Structural basis of the non-coding RNA RsmZ acting as a protein sponge. *Nature* 509(7502), 588-592. doi: 10.1038/nature13271.
17. Fairman-Williams, M.E., Guenther, U.P., and Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol* 20(3), 313-324. doi: 10.1016/j.sbi.2010.03.011.
18. Finn, R.D., Bateman, A., Clements, J., Coghill, P., Eberhardt, R.Y., Eddy, S.R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res* 42(Database issue), D222-230. doi: 10.1093/nar/gkt1223.
19. Forster, S.C., Kumar, N., Anonye, B.O., Almeida, A., Viciani, E., Stares, M.D., et al. (2019). A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nat Biotechnol* 37(2), 186-192. doi: 10.1038/s41587-018-0009-7.
20. Fu, H., Elena, R.C., and Marquez, P.H. (2019). The roles of small RNAs: insights from bacterial quorum sensing. *ExRNA* 1(1), 1-8. doi: 10.1186/s41544-019-0027-8.
21. Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28(23), 3150-3152. doi: 10.1093/bioinformatics/bts565.
22. Gandara, C., and Alonso, J.C. (2015). DisA and c-di-AMP act at the intersection between DNA-damage response and stress homeostasis in exponentially growing *Bacillus subtilis* cells. *DNA Repair (Amst)* 27, 1-8. doi: 10.1016/j.dnarep.2014.12.007.
23. Gojobori, T., Wada, T., Kobayashi, T., and Mineta, K. (2019). *Marine Metagenomics: Technological Aspects and Applications*. Springer.
24. Gong, H., Vu, G.P., Bai, Y., Chan, E., Wu, R., Yang, E., et al. (2011). A *Salmonella* small non-coding RNA facilitates bacterial invasion and intracellular replication by modulating the expression of virulence factors. *PLoS Pathog* 7(9), e1002120. doi: 10.1371/journal.ppat.1002120.
25. Gottesman, S., and Storz, G. (2011). Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harb Perspect Biol* 3(12). doi: 10.1101/cshperspect.a003798.
26. Hallam, S.J., Girguis, P.R., Preston, C.M., Richardson, P.M., and DeLong, E.F. (2003). Identification of methyl coenzyme M reductase A (*mcrA*) genes associated with methane-oxidizing archaea. *Appl Environ Microbiol* 69(9), 5483-5491. doi: 10.1128/aem.69.9.5483-5491.2003.

27. Haro-Moreno, J.M., Rodriguez-Valera, F., and Lopez-Perez, M. (2019). Prokaryotic Population Dynamics and Viral Predation in a Marine Succession Experiment Using Metagenomics. *Front Microbiol* 10, 2926. doi: 10.3389/fmicb.2019.02926.
28. Haroon, M.F., Hu, S., Shi, Y., Imelfort, M., Keller, J., Hugenholtz, P., et al. (2013). Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* 500(7464), 567-570. doi: 10.1038/nature12375.
29. He, Y., Li, M., Perumal, V., Feng, X., Fang, J., Xie, J., et al. (2016). Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments. *Nat Microbiol* 1(6), 16035. doi: 10.1038/nmicrobiol.2016.35.
30. Hirano, T., Minamino, T., Namba, K., and Macnab, R.M. (2003). Substrate specificity classes and the recognition signal for Salmonella type III flagellar export. *J Bacteriol* 185(8), 2485-2492. doi: 10.1128/jb.185.8.2485-2492.2003.
31. Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 119. doi: 10.1186/1471-2105-11-119.
32. Jones, T.A., Otto, W., Marz, M., Eddy, S.R., and Stadler, P.F. (2009). A survey of nematode SmY RNAs. *RNA Biol* 6(1), 5-8. doi: 10.4161/rna.6.1.7634.
33. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* 79(2), 926-935. doi: 10.1063/1.445869.
34. Joshi, N., and Fass, J. (2011). Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. Available from: github.com/najoshi/sickle.
35. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28(1), 27-30. doi: 10.1093/nar/28.1.27.
36. Knittel, K., and Boetius, A. (2009). Anaerobic oxidation of methane: progress with an unknown process. *Annu Rev Microbiol* 63, 311-334. doi: 10.1146/annurev.micro.61.080706.093130.
37. Kopylova, E., Noe, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28(24), 3211-3217. doi: 10.1093/bioinformatics/bts611.
38. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4), 357-359. doi: 10.1038/nmeth.1923.
39. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21), 2947-2948. doi: 10.1093/bioinformatics/btm404.
40. Lewis, K., Epstein, S., D'Onofrio, A., and Ling, L.L. (2010). Uncultured microorganisms as a source of secondary metabolites. *J Antibiot (Tokyo)* 63(8), 468-476. doi: 10.1038/ja.2010.87.
41. Lohman, T.M., Tomko, E.J., and Wu, C.G. (2008). Non-hexameric DNA helicases and translocases: mechanisms and regulation. *Nat Rev Mol Cell Biol* 9(5), 391-401. doi: 10.1038/nrm2394.

42. Luckett, J.C., Darch, O., Watters, C., Abuoun, M., Wright, V., Paredes-Osses, E., et al. (2012). A novel virulence strategy for *Pseudomonas aeruginosa* mediated by an autotransporter with arginine-specific aminopeptidase activity. *PLoS Pathog* 8(8), e1002854. doi: 10.1371/journal.ppat.1002854.
43. Marinho, C.M., Dos Santos, P.T., Kallipolitis, B.H., Johansson, J., Ignatov, D., Guerreiro, D.N., et al. (2019). The sigma(B)-dependent regulatory sRNA Rli47 represses isoleucine biosynthesis in *Listeria monocytogenes* through a direct interaction with the *ilvA* transcript. *RNA Biol* 16(10), 1424-1437. doi: 10.1080/15476286.2019.1632776.
44. Matsui, M., Fowler, J.H., and Walling, L.L. (2006). Leucine aminopeptidases: diversity in structure and function. *Biol Chem* 387(12), 1535-1544. doi: 10.1515/BC.2006.191.
45. Meyerdierks, A., Kube, M., Kostadinov, I., Teeling, H., Glockner, F.O., Reinhardt, R., et al. (2010). Metagenome and mRNA expression analyses of anaerobic methanotrophic archaea of the ANME-1 group. *Environ Microbiol* 12(2), 422-439. doi: 10.1111/j.1462-2920.2009.02083.x.
46. Meyerdierks, A., Kube, M., Lombardot, T., Knittel, K., Bauer, M., Glockner, F.O., et al. (2005). Insights into the genomes of archaea mediating the anaerobic oxidation of methane. *Environ Microbiol* 7(12), 1937-1951. doi: 10.1111/j.1462-2920.2005.00844.x.
47. Murakami, S., Fujishima, K., Tomita, M., and Kanai, A. (2012). Metatranscriptomic analysis of microbes in an Oceanfront deep-subsurface hot spring reveals novel small RNAs and type-specific tRNA degradation. *Appl Environ Microbiol* 78(4), 1015-1022. doi: 10.1128/AEM.06811-11.
48. Nakatsu, Y., Matsui, H., Yamamoto, M., Noutoshi, Y., Toyoda, K., and Ichinose, Y. (2019). Quorum-dependent expression of *rsmX* and *rsmY*, small non-coding RNAs, in *Pseudomonas syringae*. *Microbiol Res* 223-225, 72-78. doi: 10.1016/j.micres.2019.04.004.
49. Nawaz, M.Z., Jian, H., He, Y., Xiong, L., Xiao, X., and Wang, F. (2017). Genome-Wide Detection of Small Regulatory RNAs in Deep-Sea Bacterium *Shewanella piezotolerans* WP3. *Front Microbiol* 8, 1093. doi: 10.3389/fmicb.2017.01093.
50. Nawrocki, E.P., Burge, S.W., Bateman, A., Daub, J., Eberhardt, R.Y., Eddy, S.R., et al. (2015). Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res* 43(Database issue), D130-137. doi: 10.1093/nar/gku1063.
51. Nelson, J.W., Sudarsan, N., Furukawa, K., Weinberg, Z., Wang, J.X., and Breaker, R.R. (2013). Riboswitches in eubacteria sense the second messenger c-di-AMP. *Nature Chemical Biology* 9(12), 834-839. doi: 10.1038/nchembio.1363.
52. Ozturk, D.H., Dorfman, R.H., Scapin, G., Sacchettini, J.C., and Grubmeyer, C. (1995). Locations and functional roles of conserved lysine residues in *Salmonella typhimurium* orotate phosphoribosyltransferase. *Biochemistry* 34(34), 10755-10763. doi: 10.1021/bi00034a007.
53. Peer, A., and Margalit, H. (2011). Accessibility and evolutionary conservation mark bacterial small-rna target-binding regions. *J Bacteriol* 193(7), 1690-1701. doi: 10.1128/JB.01419-10.
54. Peng, Y., Leung, H.C., Yiu, S.M., and Chin, F.Y. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28(11), 1420-1428. doi: 10.1093/bioinformatics/bts174.

55. Puspita, I.D., Kamagata, Y., Tanaka, M., Asano, K., and Nakatsu, C.H. (2012). Are uncultivated bacteria really uncultivable? *Microbes Environ* 27(4), 356-366. doi: 10.1264/jsme2.me12092.
56. Pyle, A.M. (2008). Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu Rev Biophys* 37, 317-336. doi: 10.1146/annurev.biophys.37.032807.125908.
57. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41(Database issue), D590-596. doi: 10.1093/nar/gks1219.
58. Richter, A.S., and Backofen, R. (2012). Accessibility and conservation: general features of bacterial small RNA-mRNA interactions? *RNA Biol* 9(7), 954-965. doi: 10.4161/rna.20294.
59. Seetharaman, M., Eldho, N.V., Padgett, R.A., and Dayie, K.T. (2006). Structure of a self-splicing group II intron catalytic effector domain 5: parallels with spliceosomal U6 RNA. *RNA* 12(2), 235-247. doi: 10.1261/rna.2237806.
60. Sharma, C.M., Darfeuille, F., Plantinga, T.H., and Vogel, J. (2007). A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev* 21(21), 2804-2817. doi: 10.1101/gad.447207.
61. Shi, Y., Tyson, G.W., and DeLong, E.F. (2009). Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* 459(7244), 266-269. doi: 10.1038/nature08055.
62. Silva, I.J., Barahona, S., Eyraud, A., Lalaouna, D., Figueroa-Bossi, N., Masse, E., et al. (2019). SraL sRNA interaction regulates the terminator by preventing premature transcription termination of rho mRNA. *Proc Natl Acad Sci U S A* 116(8), 3042-3051. doi: 10.1073/pnas.1811589116.
63. Simoneit, B.R.T., and Lonsdale, P.F. (1982). Hydrothermal petroleum in mineralized mounds at the seabed of Guaymas Basin. *Nature* 295(5846), 198-202. doi: 10.1038/295198a0.
64. Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem* 76, 23-50. doi: 10.1146/annurev.biochem.76.052305.115300.
65. Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J* 30(7), 1335-1342. doi: 10.1038/emboj.2011.41.
66. Smaldone, G.T., Antelmann, H., Gaballa, A., and Helmann, J.D. (2012). The FsrA sRNA and FbpB protein mediate the iron-dependent induction of the *Bacillus subtilis* lutABC iron-sulfur-containing oxidases. *J Bacteriol* 194(10), 2586-2593. doi: 10.1128/JB.05567-11.
67. Solden, L., Lloyd, K., and Wrighton, K. (2016). The bright side of microbial dark matter: lessons learned from the uncultivated majority. *Curr Opin Microbiol* 31, 217-226. doi: 10.1016/j.mib.2016.04.020.
68. Stafford, G.P., Evans, L.D., Krumscheid, R., Dhillon, P., Fraser, G.M., and Hughes, C. (2007). Sorting of early and late flagellar subunits after docking at the membrane ATPase of the type III export pathway. *J Mol Biol* 374(4), 877-882. doi: 10.1016/j.jmb.2007.09.080.

69. Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4, 41. doi: 10.1186/1471-2105-4-41.
70. Taylor, A. (1993). Aminopeptidases: structure and function. *FASEB J* 7(2), 290-298. doi: 10.1096/fasebj.7.2.8440407.
71. Toffano-Nioche, C., Nguyen, A.N., Kuchly, C., Ott, A., Gautheret, D., Bouloc, P., et al. (2012). Transcriptomic profiling of the oyster pathogen *Vibrio splendidus* opens a window on the evolutionary dynamics of the small RNA repertoire in the *Vibrio* genus. *RNA* 18(12), 2201-2219. doi: 10.1261/rna.033324.112.
72. Valentini, M., and Filloux, A. (2016). Biofilms and Cyclic di-GMP (c-di-GMP) Signaling: Lessons from *Pseudomonas aeruginosa* and Other Bacteria. *J Biol Chem* 291(24), 12547-12555. doi: 10.1074/jbc.R115.711507.
73. Vegh, B.M., Gal, P., Dobo, J., Zavodszky, P., and Vonderviszt, F. (2006). Localization of the flagellum-specific secretion signal in *Salmonella* flagellin. *Biochem Biophys Res Commun* 345(1), 93-98. doi: 10.1016/j.bbrc.2006.04.055.
74. Wang, F.P., Zhang, Y., Chen, Y., He, Y., Qi, J., Hinrichs, K.U., et al. (2014). Methanotrophic archaea possessing diverging methane-oxidizing and electron-transporting pathways. *ISME J* 8(5), 1069-1078. doi: 10.1038/ismej.2013.212.
75. Wang, Y., Feng, X., Natarajan, V.P., Xiao, X., and Wang, F. (2019). Diverse anaerobic methane- and multi-carbon alkane-metabolizing archaea coexist and show activity in Guaymas Basin hydrothermal sediment. *Environ Microbiol* 21(4), 1344-1355. doi: 10.1111/1462-2920.14568.
76. Wang, Y., Wegener, G., Ruff, S.E., and Wang, F. (2020). Methyl/alkyl-coenzyme M reductase-based anaerobic alkane oxidation in archaea. *Environ Microbiol*. doi: 10.1111/1462-2920.15057.
77. Washietl, S., Hofacker, I.L., and Stadler, P.F. (2005). Fast and reliable prediction of noncoding RNAs. *Proc Natl Acad Sci U S A* 102(7), 2454-2459. doi: 10.1073/pnas.0409169102.
78. Weber-Sparenberg, C., Poplau, P., Brookman, H., Rochon, M., Mockel, C., Nietschke, M., et al. (2006). Characterization of the type III export signal of the flagellar hook scaffolding protein FlgD of *Escherichia coli*. *Arch Microbiol* 186(4), 307-316. doi: 10.1007/s00203-006-0146-0.
79. Weinberg, Z., Perreault, J., Meyer, M.M., and Breaker, R.R. (2009). Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis. *Nature* 462(7273), 656-659. doi: 10.1038/nature08586.
80. Weinberg, Z., Wang, J.X., Bogue, J., Yang, J., Corbino, K., Moy, R.H., et al. (2010). Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol* 11(3), R31. doi: 10.1186/gb-2010-11-3-r31.
81. Welhan, J.A. (1988). Origins of methane in hydrothermal systems. *Chemical Geology* 71(1-3), 183-198. doi: 10.1016/0009-2541(88)90114-3.
82. Wongnate, T., Sliwa, D., Ginovska, B., Smith, D., Wolf, M.W., Lehnert, N., et al. (2016). The radical mechanism of biological methane synthesis by methyl-coenzyme M reductase. *Science* 352(6288),

953-958. doi: 10.1126/science.aaf0616.

83. Wright, P.R., Georg, J., Mann, M., Sorescu, D.A., Richter, A.S., Lott, S., et al. (2014). CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. *Nucleic Acids Res* 42(Web Server issue), W119-123. doi: 10.1093/nar/gku359.
84. Xu, X., Ji, Y., and Stormo, G.D. (2009). Discovering cis-regulatory RNAs in Shewanella genomes by Support Vector Machines. *PLoS Comput Biol* 5(4), e1000338. doi: 10.1371/journal.pcbi.1000338.
85. Yang, Y.Z., Ding, S., Wang, Y., Li, C.L., Shen, Y., Meeley, R., et al. (2017). Small kernel2 Encodes a Glutaminase in Vitamin B6 Biosynthesis Essential for Maize Seed Development. *Plant Physiol* 174(2), 1127-1138. doi: 10.1104/pp.16.01295.
86. Zeng, Q., and Sundin, G.W. (2014). Genome-wide identification of Hfq-regulated small RNAs in the fire blight pathogen *Erwinia amylovora* discovered small RNAs with virulence regulatory function. *BMC Genomics* 15(1), 414. doi: 10.1186/1471-2164-15-414.

Figures

Figure 1

Schematic diagram showing the metagenome and metatranscriptome based integrated approach used to detect small regulatory RNAs in Guaymas Basin.

Figure 2

Position of sRNAs identified in ANME-2a and ANME-2d and their flanking genes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarydata.docx](#)