

## Discovery of Small Regulatory RNAs Extends Our Understanding of Gene Regulation in the *Acidithiobacillus* Genus

Amir Shmaryahu<sup>a</sup> and David S. Holmes<sup>b</sup>

Center for Bioinformatics and Genome Biology, Life Science Foundation,  
MIFAB and Andrés Bello University, Santiago, Chile

<sup>a</sup>amir.shmaryahu@gmail.com, <sup>b</sup>dsholmes2000@yahoo.com

**Keywords:** Biomining, small regulatory RNAs, noncoding RNA, *Acidithiobacillus*

**Abstract.** Small regulatory RNAs (srRNAs) control gene expression in Bacteria, usually at the post-transcriptional level, by acting as antisense RNAs that bind targeted mRNAs or by interacting with regulatory proteins. srRNAs are involved in the regulation of a large variety of processes such as plasmid replication, transposition and global genetic circuits that respond to environmental changes. Since their discovery a few years ago, it has become apparent that they are prolific and widespread. In this study, we describe bioinformatic approaches to srRNA discovery in the biomining microorganisms *Acidithiobacillus ferrooxidans*, *A. caldus* and *A. thiooxidans*. Intergenic regions of the annotated genomes were extracted and computationally searched for srRNAs. Candidate srRNAs that were associated with predicted sigma 70 promoters and/or rho-independent terminators were chosen for further study. The resulting potential srRNAs include known examples from other microorganisms and some novel candidates and reveal interesting underlying biology of the *Acidithiobacillus* genus.

### Introduction

All Bacteria contain small regulatory RNAs (srRNA), ranging in size from 70 to 500 nucleotides, that control gene expression. Despite their widespread occurrence and important functional roles, current automatic genome annotation programs fail to locate srRNA genes. The computational discovery of srRNAs is particularly challenging because they do not encode protein products so bioinformatic tools such as open reading frame (ORF) identification, codon usage and Hidden Markov Model (HMM) searches based on conserved protein motifs cannot be used. In addition, srRNAs can be poorly conserved at the nucleotide sequence level and only some exhibit conserved secondary structure. Currently, most genomic studies of bacteria focus on annotation and functional assignment of unknown proteins identified. However, genome wide discovery of srRNA genes should not be omitted in any functional genomic analysis.

SrRNAs are proving to be multifunctional and have provided explanations for a number of previously mysterious regulatory effects. Phages and plasmids have long been recognized to use antisense RNA regulators [1, 2] but now srRNAs are being discovered in all bacterial genomes, including pathogens. Quorum sensing in *Vibrio* species has now been shown to depend upon srRNAs [3]. In eukaryotic cells, microRNAs and RNAi parallel in many ways the bacterial srRNAs, confirming that this level of regulation is widespread [4, 5].

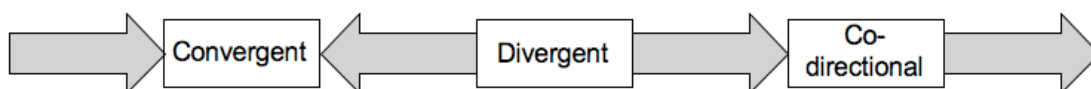
The intent of this study is to use what has been learned in analyzing srRNAs in other organisms, principally *Escherichia coli* [6], to predict their presence and function in the biomining microorganisms *Acidithiobacillus ferrooxidans*, *A. caldus* and *A. thiooxidans*, hereafter referred to as “acidithiobacilli”. Four groups have recently published genome wide searches for srRNA genes in *E. coli*. Two of these studies are primarily based on stringent experimental criteria [7, 8], and the other two groups used bioinformatic algorithms [9,10].

## Materials and Methods

A database of DNA sequences of all *A. ferrooxidans* intergenic regions (IGRs) was created. An IGR sequence is defined as the region between annotated protein, ribosomal and tRNA encoding genes. The IGR data base was searched for (i) putative  $\sigma_{70}$  promoters using a bioinformatic approach based on HMMs of the -35 and -10  $\sigma_{70}$  boxes and the size of the region separating the two boxes and (ii) Rho-independent terminators using TransTermHP [11]. Rfam ([www.sanger.ac.uk/Software/Rfam](http://www.sanger.ac.uk/Software/Rfam)) was used to predict conserved RNA motifs via multiple sequence alignment comparisons and secondary structure was predicted using M-fold ([www.bioinfo.rpi.edu/applications/mfold](http://www.bioinfo.rpi.edu/applications/mfold)). Candidate srRNA sequences were used to search the annotated genomes of *A. thiooxidans* and *A. caldus* for regions that exhibited sequence similarity and conserved genetic context.

## Results and Discussion

Nucleotide sequences of 30–500 base pairs of three classes of intergenic region (IGR) (convergent, divergent and co-directional as shown in Figure 1) were extracted from the *A. ferrooxidans* annotated genome and analyzed for the presence of (i) predicted sigma-70 promoters, (ii) rho-independent transcriptional terminators, (iii) potential similarity with known srRNAs, (iv) the ability to form secondary structures characteristic of known srRNAs and (v) conservation of sequence and genetic context in the three acidithiobacilli genomes.



**Figure 1.** Three classes of intergenic spaces (boxes) of the acidithiobacilli were searched for srRNAs. Arrows represent protein encoding genes and the direction of their transcription.

This screen resulted in the identification of eight candidate srRNA encoding genes (Table 1). Three of these have previously been described in other organisms and five are novel. It is not known if the novel examples are specific to the acidithiobacilli or remain to be discovered in other microorganisms.

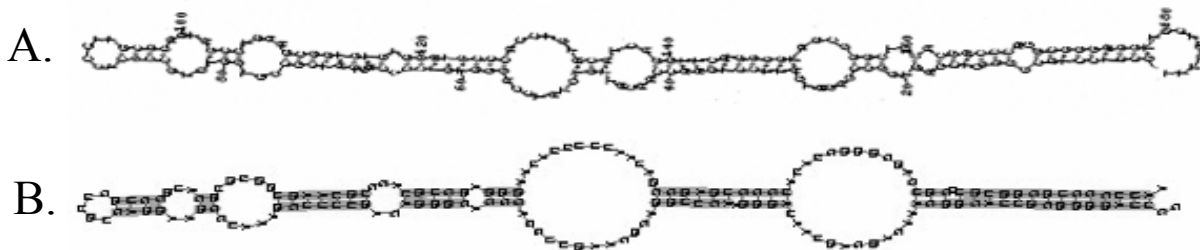
**Table 1.** Candidate srRNAs with conserved sequences in *Acidithiobacillus ferrooxidans*, *A. caldus* and *A. thiooxidans*.

Name	Predicted size (nt)	Found in: (class of IGR)	Predicted function
rnpB	293	Divergent	tRNA maturation
ssrA	360	Divergent	Translational regulation
6S	187	Co-directional	Growth regulation
srRNA1	89	Co-directional	Unknown
srRNA2	87	Co-directional	Unknown
srRNA3	244	Divergent	Unknown
srRNA4	169	Co-directional	Unknown
srRNA5	89	Co-directional	Unknown

The three known srRNAs tentatively identified in the acidithiobacilli are rnpB, ssrA and 6SRNA. RnpB is a highly conserved enzymatic RNA that is found in RNaseP, a riboenzyme involved in processing tRNAs and rRNAs [12]. The gene potentially encoding the protein component of RNaseP was also detected in the genomes of the acidithiobacilli.

SsrA, also known as TmRNA, is involved in the process termed *trans*-translation [13]. According to the well-accepted model, when a ribosome stalls on a problematic mRNA, typically at the 3' end of a truncated mRNA without an in-frame stop codon, SsrA RNA charged with alanine is recruited to the A site of the ribosome together with an associated protein, SmpB, and donates an alanine to the growing polypeptide chain by acting as a tRNA. The ribosome then switches from the mRNA to a short coding sequence in SsrA RNA. Translation continues and terminates normally at the stop codon that follows the SsrA RNA reading frame. The final translation product is a chimeric polypeptide in which the specific 11-residue tag (AANDENYALAA) is attached at the C terminus. The tag sequence targets the product of *trans*-translation for degradation by ATP-dependent proteases [14].

In *E. coli*, 6S RNA is 184 nucleotides long and folds into an extended double stranded hairpin structure with various single stranded bulges that are important for its function. 6S RNA specifically associates with RNA polymerase holoenzyme containing the  $\sigma_{70}$  transcription factor. This interaction is thought to aid in the transition from exponential to stationary phase transcription [15]. 6S RNA genes have recently been identified in many bacterial genomes [16,17]. An RNA of the about the right length (187 nucleotides) and with a predicted secondary structure that contained key interacting single stranded bulges was identified in the genomes of the acidithiobacilli (Figure 2).



**Figure 2.** (A) Predicted structure of 6S RNA in *A. ferrooxidans*, *A. caldus* and *A. thiooxidans*. (B) Known structure of 6S RNA in *E. coli* K12.

The roles of the five candidate srRNAs predicted in the acidithiobacilli are unknown and represent an important challenge for the experimental biologist. Given their versatile roles in transcriptional and translational control of gene expression and in quality control of macromolecular products, it is suggested that the study of these predicted srRNAs will yield important clues in the future as to how the acidithiobacilli fine-tuning cell processes in response to changing environments. Studies are underway to experimentally validate the existence of the predicted srRNAs, their potential target genes and their role in gene regulation.

### Acknowledgements

Work supported by Fondecyt 1050063, DI-UNAB 34-06, DI-UNAB 24-06/I and a Microsoft Sponsored Research Award.

### References

- [1] N. Delihias: Mol. Microbiol. Vol. 15 (1995), p. 411.
- [2] E.G.H Wagner and S. Brantl: Trends Biochem. Sci. Vol. 23 (1998), p. 451
- [3] D.H. Lenz, K.C Mok, B.N Lilley, R.V Kulkarni, N.S Wingreen and B.L. Bassler: Cell Vol 118 (2004), p. 69
- [4] J.C. Carrington and V. Ambros: Science Vol. 301 (2003), p. 336
- [5] M.T McManus and P.A Sharp: Nat. Rev. Genet. Vol. 3 (2003), p. 737
- [6] G. Storz: Science Vol. 296 (2002), p. 1260

- [7] L. Argaman, R. Hershberg, J. Vogel, G. Bejerano, E.G.H. Wagner, H. Margalit and S. Altuvia: *Current Biol.* Vol. 11 (2001), p. 941
- [8] K.M. Wassarman, F. Repoila, C. Rosenow, G. Storz and S. Gottesman: *Genes Dev.* Vol. 15 (2001), p. 1637
- [9] R.J. Carter, I. Dubchak and S.R. Holbrook: *Nucl. Acids Res.* Vol. 29 (2001), p. 3928
- [10] E. Rivas, R.J. Klein, T.A. Jones and S.R. Eddy: *Current Biol.* Vol. 11 (2001), p. 1369
- [11] C. Kingsford, K. Ayanbule and S.L. Salzberg: *Genome Biology* Vol. 8 (2007), R22
- [12] V. Gopalan, A. Vioque and S. Altman: *J. Biol. Chem.* Vol. 277 (2002), p. 6759
- [13] A.W. Karzai, E.D. Roche and R.T Sauer: *Nat. Struct. Biol.* Vol. 7 (2000), p. 449
- [14] S. Gottesman, E. Roche, Y.N. Zhou and R.T. Sauer: *Genes Dev.* Vol. 12 (1998), p. 1338
- [15] K.M. Wassarman, G. Storz: *Cell* Vol. 101 (2002), p. 613
- [16] A.E. Trotochaud and K.M. Wassarman: *Nat. Struct. Mol. Biol.* Vol. 12 (2005), p. 313
- [17] J.E. Barrick, N. Sudarsan, Z. Weinberg, W.L. Ruzzo and R.R. Breaker: *RNA* Vol. 11 (2005), p. 774