H-2645

Identification of Small RNAs in *Desulfovibrio vulgaris* Hildenborough

A. S. Burns[1,4], M. P. Joachimiak[2,4], A. M. Deutschbauer[2,4], A. P. Arkin[2,3,4], and K. S. Bender[1,4]

[1]Southern Illinois University, Carbondale, IL, [2]Lawrence Berkeley National Laboratory, Berkeley, CA, [3]University of California, Berkeley, CA, [4]VIMSS (Virtual Inst. of Microbial Stress and Survival) http://vimss.lbl.gov/

ENIGMA — Ecosystems and Networks Integrated with Genes and Molecular Assemblies

DOE GENOMICS:GTL — ACCELERATING DISCOVERY FOR ENERGY AND ENVIRONMENT — OFFICE OF SCIENCE U.S. DEPARTMENT OF ENERGY

VIMSS Virtual Institute for Microbial Stress and Survival | UCSF University of California San Francisco | Southern Illinois University Carbondale | Berkeley University of California | http://vimss.lbl.gov/

Abstract

Desulfovibrio vulgaris is an anaerobic sulfate-reducing bacterium capable of facilitating the removal of toxic metals such as uranium from contaminated sites via reduction. As such, it is essential to understand the intricate regulatory cascades involved in how *D. vulgaris* and its relatives respond to stressors in such sites. One approach is the identification and analysis of small non-coding RNAs (sRNAs); molecules ranging in size from 20-200 nucleotides that predominantly affect gene regulation by binding to complementary mRNA in an anti-sense fashion and therefore provide an immediate regulatory response. To identify sRNAs in *D. vulgaris*, a bacterium that does not possess an annotated *hfq* gene, RNA was pooled from stationary and exponential phases, nitrate exposure, and biofilm conditions. The subsequent RNA was size fractionated, modified, and converted to cDNA for high throughput transcriptomic deep sequencing. A computational approach to identify sRNAs via the alignment of seven separate *Desulfovibrio* genomes was also performed. From the deep sequencing analysis, 2,296 reads between 20 and 250 nt were identified with expression above genome background. Analysis of those reads limited the number of candidates to ~87 intergenic, while ~140 appeared to be antisense to annotated open reading frames (ORFs). Further BLAST analysis of the intergenic candidates and other *Desulfovibrio* genomes indicated that eight candidates were likely portions of ORFs not previously annotated in the *D. vulgaris* genome. Comparison of the intergenic and antisense data sets to the bioinformatical predicted candidates, resulted in ~54 common candidates. Current approaches using Northern analysis and qRT-PCR are being used to verify expression of the candidates and to further develop the role these sRNAs play in *D. vulgaris* regulation.

Materials and Methods

- mVISTA, a genome browser, was used to align and compare seven *Desulfovibrio* genomes (1,3).
- The program XRate was used to re-estimate branch length and conserved secondary structure was determined by windowlicker, a scanning algorithm, resulting in a list of putative sRNA candidates (2).
- High throughput transcriptomic deep sequencing was also used to identify novel sRNAs (Fig. 1).

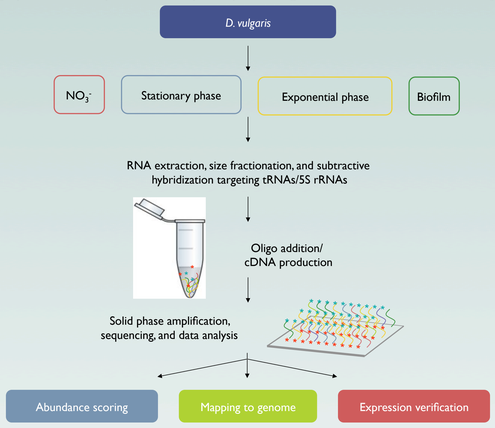

Figure 1. High throughput deep sequencing via Solexa (Illumina).

- Sequenced cDNA was scored based on abundance and mapped to the genome of *Desulfovibrio vulgaris* Hildenborough.
- mRNA fragments, possible mRNA tail or leader sequences, and candidates smaller than 40 nt were disregarded.
- Remaining candidates were separated into intergenic putative sRNAs and those antisense to existing open reading frames (ORFs) and ordered by abundance and presence in previous computational approach (Fig. 2).
- Northern analysis was performed on 10 µg of *D. vulgaris* Hildenborough exponential and stationary RNA separated on a 8% polyacrylamide/ 7M Urea gel and transferred to a nylon membrane by electroblotting.
- Expression of candidates was verified with hybridization of membrane with [γ-^{32}P]-ATP radiolabeled oligo probes (Fig. 3).

Results

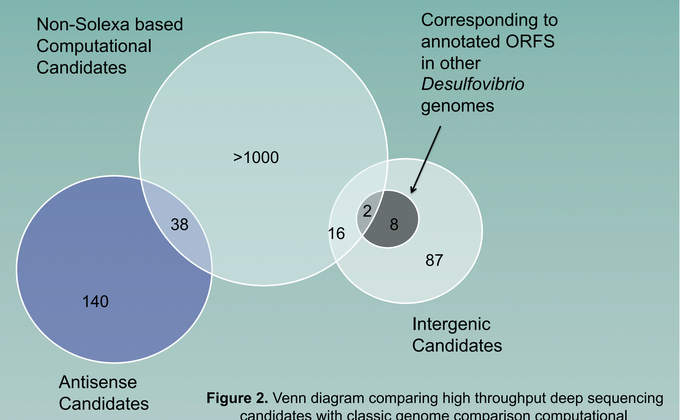

Figure 2. Venn diagram comparing high throughput deep sequencing candidates with classic genome comparison computational techniques.

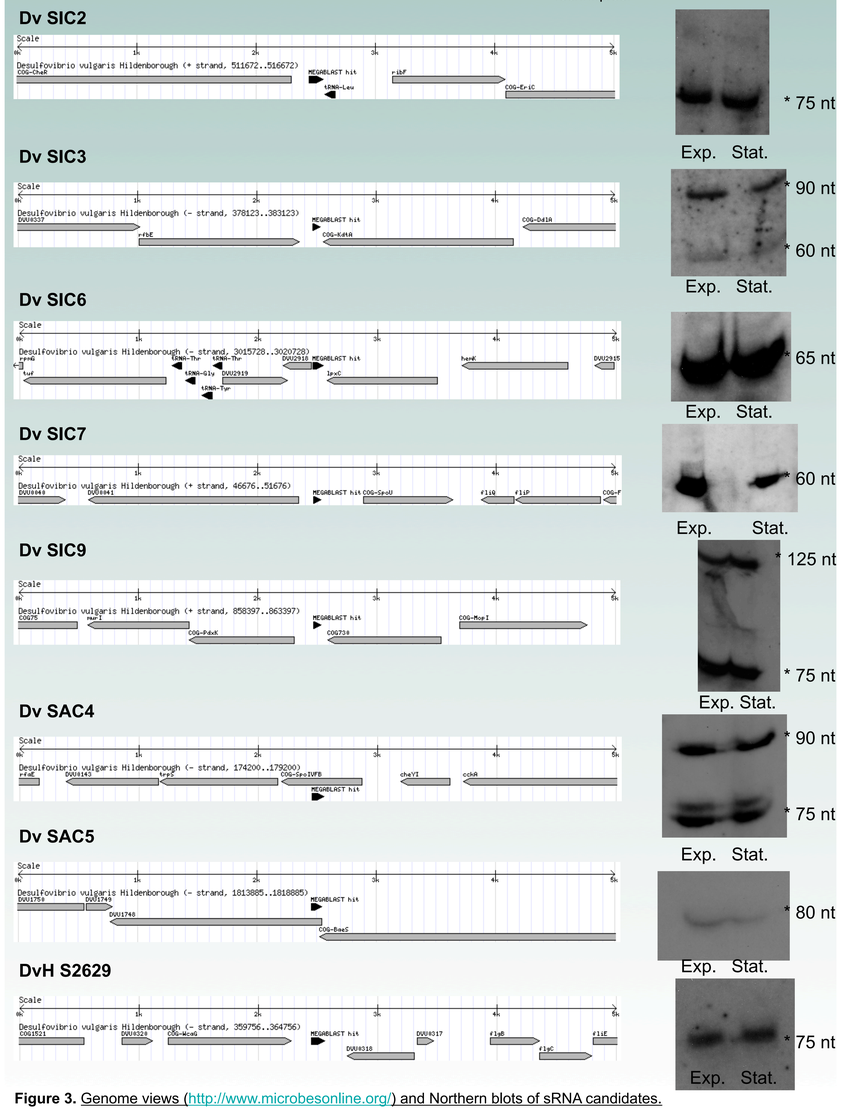

Figure 3. Genome views (http://www.microbesonline.org/) and Northern blots of sRNA candidates.

Results

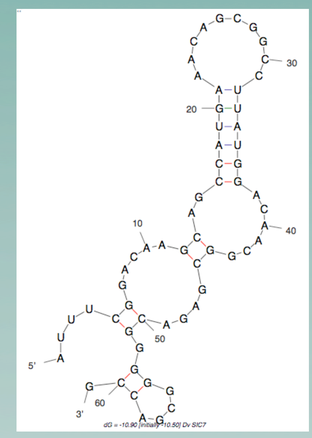

Figure 4. Preliminary RNA secondary structure for putative sRNA Dv SIC7 created using mfold **http://mfold.bioinfo.rpi.edu/**

Conclusions

- High throughput transcriptomic deep sequencing as well as classic genome conservation computational approaches yields a plethora of novel putative sRNAs throughout the genome of *Desulfovibrio vulgaris* Hildenborough.
- Roughly 50% of the candidates tested for expression in exponential or stationary phase RNA yielded bands.
  - Attempting Northern blots with RNA from nitrate stressed and biofilm cultures may provide results for so far unresolved candidates.
- Results show that some of the sRNAs such as Dv SIC3, Dv SIC 9, and Dv SAC4 contain multiple processed forms.
- One sRNA– Dv SIC7– exhibits differing expression level under exponential phase as compared to stationary phase which was confirmed by probing with a 5S loading control.
  - Quantitative reverse transcription PCR (qRT-PCR) experiments are undergoing to determine how expression of Dv SIC7 changes from exponential to stationary phase.
- 3' and 5' RACE experiments are underway to determine the full sequence of the putative sRNAs and target mRNAs identification will begin shortly after.
- Northern analysis and qRT-PCR under various stressors will be utilized to determine conditions in which the sRNA is up or down regulated.
  - These conditions will then be targets for experiments dealing in over-expression or knockout of the putative sRNA to determine phenotypic effects.

References

1. **Frazer, K. A., L. Pachter, A. Poliakov, E. M. Rubin, and I. Dubchak.** 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Research **32:**W273-W279.
2. **Klosterman, P. S., A. V. Uzilov, Y. R. Bendaña, R. K. Bradley, S. Chao, C. Kosiol, N. Goldman, and I. Holmes.** 2006. XRate: a fast prototyping, training and annotation tool for phylo-grammars. BMC Bioinformatics **7:**428.
3. **Mayor, C., M. Brudno, J. R. Schwartz, A. Poliakov, E. M. Rubin, K. A. Frazer, L. S. Pachter, and I. Dubchak.** 2000. VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics Applications **16:**1046-1047.

ACKNOWLEDGEMENTS

ENIGMA is a Scientific Focus Area Program supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL Foundational Science through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy. Travel supported by the Graduate Student Travel Award of Molecular Biology, Microbiology, and Biochemistry Graduate Program at Southern Illinois University.