Title
Small Antisense Regulatory RNA Genes in Bacterial Genomes

Authors
Mavromatis, K.
Kyrpides, N.C.

Publication Date
2006-05-15

Small Antisense Regulatory RNA Genes in Bacterial Genomes

K. Mavromatis[1], N. C. Kyrpides[1]
[1] DOE Joint Genome Institute, Walnut Creek, CA

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Introduction

Small RNAs (sRNA) can act as regulators of the cell functions, mainly in two ways(1,2). First, they bind to specific proteins and change their activity, an example case is 6S that binds to RNA polymerase and alters its activity. Second, sRNA molecules affect mRNA translation through base pairing interactions near the RBS. These interactions can alter mRNA structure and/or stability resulting either to inhibition or promotion of ribosome binding.

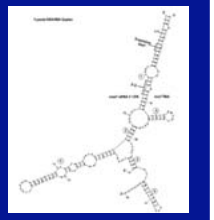
Antisense sRNA are small RNA molecules that have a small region which is complementary to the target mRNA (3). Thus, they can inhibit translation by occluding the ribosome binding site, or activate translation by preventing the formation of inhibitory mRNA structures. The rest of the molecule folds creating secondary structure that is required for its function. The specificity of these molecules is based on the complementarity with the target mRNA (figure). Mutations can accumulate in these molecules provided that they do not affect this pairing and the overall structure of the molecule. As a result these sRNAs can become more diverse between distant phylogenetically species.

The RNA-binding protein Hfq appears to play important role in the regulation of gene translation through the antisense RNA fashion (4). Hfq is a conserved, abundant protein that has been implicated in a number of RNA-mediated events. This interaction frequently results to the degradation of the mRNA.

We developed a method based on the above mentioned observations for the identification of putative antisense RNAs in the currently public genomes. Our method identifies homologous intergenic regions that exhibit complementarity with homologous genes in different organisms. Further criteria for conservation of the RNA complementarity pattern (complement bases relative to the start of the gene), and predicted loops in the putative RNA gene are used to filter results.

Enterobacterial organisms were used for the evaluation of the method and the results were compared to information known from the literature (5-8). Predictions made for Escherichia coli (K12) are currently experimentally studied.

1. Wassarman, K.M., Zhang, A. & Storz, G. Small RNAs in Escherichia coli. *Trends Microbiol* **7**. 37-45 (1999).
2. Eddy, S.R. Non-coding RNA genes and the modern RNA world. *Nat Rev Genet* **2**. 919-929 (2001).
3. Wassarman, K.M. Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes. *Cell* **109**, 141-144 (2002).
4. Klein, R.J., Misulovin, Z. & Eddy, S.R. Noncoding RNA genes identified in AT-rich hyperthermophiles. *Proc Natl Acad Sci U S A* **99**. 7542-7547 (2002).
5. Wassarman, K.M., Repoila, F., Rosenow, C., Storz, G. & Gottesman, S. Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev* **15**. 1637-1651 (2001).
6. Rivas, E., Klein, R.J., Jones, T.A. & Eddy, S.R. Computational identification of noncoding RNAs in E. coli by comparative genomics. *Curr Biol* **11**. 1369-1373 (2001).
7. Carter, R.J., Dubchak, I. & Holbrook, S.R. A computational approach to identify genes for functional RNAs in genomic sequences. *Nucleic Acids Res* **29**. 3928-3938 (2001).
8. Zhang, A. et al. Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol* **50**. 1111-1124 (2003).



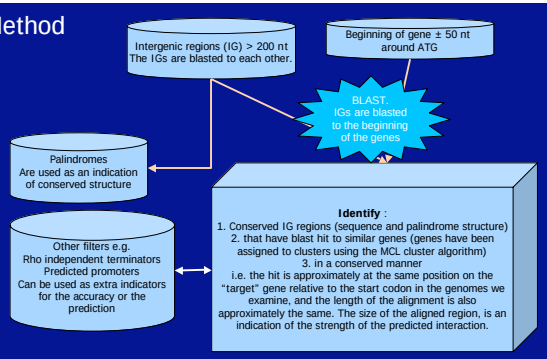
from: Delihas, BMC Microbiology 2003:3:13.

Conclusions

- A significant number of conserved IG regions with consistent complementarity with genes can be identified in Enterobacteria
  - These genomes are phylogenetically closed to each other
- In Bacilli the number of predicted interactions is low
  - These genomes are more distant phylogenetically
- Chlamydia (intracellular pathogens) have a very low number of predicted interactions implying either loss of this mechanism, or replacement with another
  - Absence of Hfq homolog is in favor of loss of this function
  - The small number of transcriptional regulators in these genomes implies that these genomes do not need a large number of regulatory elements.
- Surpiringly, Rickettsias exhibit significant number of putative interactions although they seem not to have Hfq homolog.
- Genes that are predicted to interact with IG elements belong to several functional categories.
  - There is a preference for transport systems, and core functions.

- More genomes will be examined in the future, and experimental verification will be pursued.

Method



The method is based on the comparisons between ONE genome with a group of others

Prediction statistics

Rickettsias/ *Rickettsia conorri* str.Malish 7

Common genomes	Alignment between IG and gene(nt)	Number of IG	Hit genes
5		63	132
4	7	142	376
5		54	95
4	12	126	315
5		34	60
4	15	106	231

Enterobacteria / *Escherichia coli* K12

Common genomes	Alignment between IG and gene(nt)	Number of IG	Hit genes
9		95	183
8		173	447
7	7	221	630
9		89	147
8		160	365
7	12	208	514
9		62	75
8		115	185
7	15	161	267

Chlamydiaceae / *Chlamydophila pneumoniae* AR39

Common genomes	Alignment between IG and gene(nt)	Number of IG	Hit genes
7		0	0
6		3	6
5	7	14	75
7		0	0
6		3	4
5	12	12	39
7		0	0
6		1	1
5	15	6	15

Bacilli / *Bacillus subtilis*

Common genomes	Alignment between IG and gene(nt)	Number of IG	Hit genes
7		2	3
6		7	10
5	7	8	14
7		1	2
6		1	3
5	12	3	5
7		1	1
6		1	2
5	15	1	2

Common genomes: The number of genomes that the predicted interaction is present.
Alignment btw IG and gene: The least number of nucleotides that are aligned between the IG and the "target" gene.

Target genes

Escherichia coli K12

GENES	FUNCTION GROUP	DEFINITION
15	R	General function prediction only
11	C	Energy production and conversion
10	G	Carbohydrate transport and metabolism
10	J	Translation, ribosomal structure and biogenesis
8	E	Amino acid transport and metabolism
8	M	Cell wall/membrane/envelope biogenesis
6	O	Posttranslational modification, protein turnover, chaperones
5	S	Function unknown
4	F	Nucleotide transport and metabolism
4	I	Lipid transport and metabolism
3	H	Coenzyme transport and metabolism
3	K	Transcription
3	L	Replication, recombination and repair
3	P	Inorganic ion transport and metabolism
2	T	Signal transduction mechanisms
2	U	Intracellular trafficking, secretion, and vesicular transport
1	D	Cell cycle control, cell division, chromosome partitioning
1	Q	Secondary metabolites biosynthesis, transport and catabolism
1	V	Defense mechanisms

Bacillus thuringiensis

GENES	FUNCTION GROUP	DEFINITION
3	R	General function prediction only
2	J	Translation, ribosomal structure and biogenesis
1	C	Energy production and conversion
1	E	Amino acid transport and metabolism
1	H	Coenzyme transport and metabolism

Rickettsia conorii

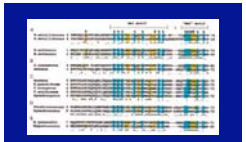
GENES	FUNCTION GROUP	DEFINITION
12	J	Translation, ribosomal structure and biogenesis
9	L	Replication, recombination and repair
6	H	Coenzyme transport and metabolism
6	O	Posttranslational modification, protein turnover, chaperones
5	M	Cell wall/membrane/envelope biogenesis
5	R	General function prediction only
4	D	Cell cycle control, cell division, chromosome partitioning
4	S	Function unknown
4	U	Intracellular trafficking, secretion, and vesicular transport
3	K	Transcription
2	C	Energy production and conversion
2	E	Amino acid transport and metabolism
1	I	Lipid transport and metabolism
1	P	Inorganic ion transport and metabolism

Requirements

The method we use is heavily dependent on the good quality of annotation and gene prediction in the related genomes. Furthermore it requires:

- Genomes of relatively close organisms. However, a degree of divergence is necessary in order to avoid random hits coming from syntenic regions.
- Presence of correct gene models. Common leaders, promoters, that are not included in the gene models can give false results.
- Presence of a region of sequence similarity between gene and intergenic region detectable by blast. Small alignment region, or complicated patterns of recognition cannot be identified.

Organisms and Hfq



Sequence alignment of different Hfq homologs. The sequence similarity is restricted to small motifs, making the indentification of Hfq homologs a difficult task.
From Valentin-Hansen etal. Mol.Microb. 2004. **51**:1525

	Organism	Presence of Hfq homolog
Group I Enterobacteria	*Escherichia coli O157:H7 EDL933*	+
	Escherichia coli K12	+
	Escherichia coli Sakai O157:H7	+
	Escherichia coli UTI89	+
	Salmonella enterica Typhi Ty2	+
	Salmonella enterica Typhi CT18	+
	Salmonella typhimurium LT2	+
	Shigella flexneri 2a 2457T	+
	Shigella flexneri 2a 301	+
Group II Bacilli	*Bacillus anthracis Ames 0581*	+
	Bacillus anthracis Ames	+
	Bacillus anthracis Sterne	+
	Bacillus cereus ATCC 10987	+
	Bacillus cereus ATCC 14579	+
	Bacillus cereus E33L	+
	Bacillus clausii KSM-K16	+
	Bacillus halodurans C-125	+
	Bacillus licheniformis Goettingen	+
	Bacillus licheniformis Novozymes	+
	Bacillus subtilis 168	+
	Bacillus thuringiensis konkukian 97-27	+
Group III Chlamydiaceae	*Chlamydia muridarum Nigg*	-
	Chlamydophila pneumoniae AR39	-
	Chlamydophila pneumoniae CWL029	-
	Chlamydophila pneumoniae J138	-
	Chlamydophila pneumoniae TW-183	-
	Chlamydia trachomatis D/UW-3/CX	-
	Chlamydophila caviae GPIC	-
Group IV Rickettsias	*Rickettsia akari Hartford*	-
	Rickettsia conorii Malish 7	-
	Rickettsia prowazekii Madrid E	-
	Rickettsia sibirica 246	-
	Rickettsia typhi Wilmington	-

Groups of organisms compared.
The presence of identified Hfq homolog is shown with a (+) sign

Hfq has been shown to participate in the degradation of the sRNA-mRNA complex. Absence of the protein from some organisms could be an indication that:

- An alternative form of the protein exists that is not identifiable by simple BLAST search.
- They use a different mechanism.
- They do not have this mechanism.

Results in *E.coli (K12)*

Total predicted RNAs: 62
Known in the literature: 14
Known antisense RNAs: 7
Correctly predicted (RNA & interaction): 5

Experimentally checked: 10
Experimentally verified: 6

Predictions made in Enterobacteria, were compared to other known sRNA and predictions from the literature made for *E.coli* K12.
5 out of 7 known antisense RNA were predicted correctly.
Preliminary experimental results verify that 6 out of 10 are transcribed (data not shown).

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE:AC02-05CH11231 and Los Alamos National Laboratory under contract No. W-7405-ENG-36.