An Assessment of Bacterial Small RNA Target Prediction Programs

### Adrien Pain, Alban Ott, Hamza Amine,

### Tatiana Rochat, Philippe Bouloc, Daniel Gautheret

## Supplemental Files

(Available at: http://rna.igmors.u-psud.fr/suppl\_data/Targetbench)

**Table S1.** (Excel file TableS1-trusted-pairs.xlsx). List of trusted (experimentally validated) sRNA/target pairs used as benchmark. Includes coordinates of published interactions, references and evidence level.

**NC\_000913\_CDS\_-200nt\_100nt.fa**

5' regions of E. coli K12 genes obtained by automatic extraction of

-200/+100 fragments around each start codon (4317 regions).

Strain: Escherichia coli str. K-12 substr. MG1655.

Coordinates are given for assembly #2: NC\_000913.2

**NC\_000913\_RNA-seq\_TSS.fa**

5' regions of E. coli K12 genes extracted based on actual

RNAseq-derived TSS (1). For each TSS in Table S5 and S9 of Li et

al.'s paper (1), we extracted the region from the TSS to 100 nt past

the start codon. For genes not expressed in the RNAseq data or genes

preceeded by other genes in an operon, we extracted the -200/+100

region around ATG as above. (total: 4317 regions)

Strain: Escherichia coli str. K-12 substr. MG1655

**coli\_sRNA\_vx.fa**

All coli sRNA sequences to be used for target prediction.

**coli\_targetpairs\_Vx.tsv**

**coli\_targetpairs\_Vx\_direct.tsv**

Flat file extracts from table of true target pairs, for programmatic

usage. The "direct" version includes only those pairs supported by

direct experimental evidence. The "direct" version includes only those

pairs supported by direct experimental evidence.

**coli\_pairs\_Vx\_with\_compMut\_realUTR.tsv**

**coli\_pairs\_Vx\_with\_compMut\_defaultUTR.tsv**

Files containing the list of base pairs with experimental compensatory

mutation support (same references as in coli\_targetpairs above).

Relative coordinates of each individual base pair are provided, in the

order: sRNA position, base, mRNA position, base.

As mRNA coordinates differ when using real or default

(-200/+100) UTRs, a specific file is provided for each system.

**IstR.fa, RseX.fa, RydC.fa**

Set of homologous sequences for sRNAs not available in the copraRNA

server.

**References:**

 (1) Li S et al. Directional RNA-seq reveals highly complex condition-dependent transcriptomes in E. coli K12 through accurate full-length transcripts assembling. BMC Genomics 2013, 14:520.

## Supplemental Figures



**Figure S1**. Localization of all known target sites in *E. coli* mRNA 5’ regions. The histogram shows the number of trusted mRNA-sRNA pairs that overlap the corresponding mRNA positions, shown around the ATG start codon (position 1 on X axis). Trusted pairs are taken from Table S1.



**Figure S2**: Representations of target prediction performances, using only sRNA/target pairs with a direct or near-direct evidence for interaction (88 base pairs, see table S1). A: Rank distribution of trusted targets. For each program, the distribution shows the ranks of the best ranking target of each sRNA (24 sRNAs). Horizontal lines and numbers indicate median ranks, red dots indicate mean ranks. B: ROC-like curves. For each program, the curve shows the number of trusted pairs predicted by the program (Y axis) among the X best ranking predictions (X-axis).



**Figure S3**: Rank distribution of all 102 sRNA/target pairs. For each program, the distribution shows the ranks of trusted sRNA/target pairs among all predicted pairs for this sRNA. Horizontal lines and numbers indicate median ranks and red dots average ranks.

## Supplemental Methods

**Programs, parameters and command lines**

# Dataset :

# see all files in http://rna.igmors.u-psud.fr/suppl\_data/Targetbench

# sRNA : coli\_sRNA\_v5.fa

# mRNA : NC\_000913\_CDS\_-200nt\_100nt.fa

# True targets (102 pairs): coli\_targetpairs\_V5.tsv

# ============================================================================

# TargetRNA2 (online version at http://cs.wellesley.edu/~btjaden/TargetRNA2/advanced.html)

# Authors : Tjaden B (btjaden@wellesley.edu)

# (btjaden@wellesley.edu)

# Parameters :

# Replicon name : Escherichia coli str. K-12 substr. MG1655

# sRNA conservation and accessibiliy : on

# NTs upstream : 200

# NTs downstream : 100

# Hybridization seed : 7

# NTs in interaction region : 20

# mRNA structural accessibility : on

# P-value threshold : 1

# Filter size : 5000

# ============================================================================

# RNAplex (Version : 0.2)

# Authors : Tafer H, Hofacker IL

# [ivo@tbi.univie.ac.at]

# Accessibility computed with RNAplfold (V1.8)

RNAplfold -u 45 -O < coli\_sRNA\_v5.fa

RNAplfold -W 240 -L 160 -u 45 -O < NC\_000913\_CDS\_-200nt\_100nt.fa

# Command line :

RNAplex -t NC\_000913\_CDS\_-200nt\_100nt.fa -q coli\_sRNA\_v5.fa -l 40 -z 20 -a ./access/ > results\_targetRNA.txt

# ============================================================================

# IntaRNA (Version : IntaRNAa)

# Authors : Busch A., Richter A.S., Backofen R

# [backofen@informatik.uni-freiburg.de]

# Command line :

IntaRNA -m coli\_sRNA\_v5.fa -t NC\_000913\_CDS\_-200nt\_100nt.fa -o -p 7 -s 0 > IntaRNA\_results.txt

# ============================================================================

# CopraRNA V1.2.7 (online version at http://rna.informatik.uni-freiburg.de/CopraRNA/Input.jsp)

# Authors : Wright PR, Richter AS, Papenfort K, Mann M, Vogel J, Hess WR, Backofen R and Georg J

# [jens.georg@biologie.uni-freiburg.de] [backofen@informatik.uni-freiburg.de]

# sRNA alignments :

# - when possible we use precomputed results (ArcZ, ChiX, CyaR, DsrA, FnrS, GcvB, GlmZ, MicA, MicC, MicF, OmrA, OmrB, OxyS, RprA, RybB, RyhB, SgrS, Spot42 )

# - IstR : align\_IstR.fa

# - RseX : align\_RseX.fa

# - RydC : align\_RydC.fa

# - McaS : not enough sequences found

# Other parameters :

# organism of interest : NC\_000913 (Escherichia coli K 12 substr MG1655 uid57779)

# Extract sequences around: start codon

# nt up : 200

# nt down : 100

# ============================================================================

# RNAhybrid (Version : 2.1.2)

# Command line :

RNAhybrid -t NC\_000913\_CDS\_-200nt\_100nt.fasta -q sRNA\_v5.fna -g all -b 1 -s 3utr\_fly -n 30000 -m 30000 > rnahybrid\_results.txt

# ============================================================================

# RNAup (Vienna Package Version 2.1.8)

# Command line :

RNAup -b -d2 --noLP -c 'S' -o < input\_file.fa > RNAup\_results.txt

# Input file should be in multi-fasta format with one sRNA followed by all mRNAs

# ============================================================================

# Pairfold (package MultiRNAFold-2.0)

# Command line :

pairfold sRNA.fa mRNA.fa -m RNA > result\_sRNA\_mRNA.txt

# (program takes 1 sRNA and 1 mRNA at a time)

# ============================================================================

# RNAduplex (Vienna Package Version 2.1.8)

# Command line : RNAduplex < input\_file.fa > RNAduplex\_results.txt

# Input file should be in multi-fasta format with alternate sRNAs and mRNAs: {sRNA,mRNA,sRNA,mRNA..}

# ============================================================================

# RNACofold (Vienna Package Version 2.1.8)

# Command line : RNAcofold < input\_file.fa > RNAcofold\_results.txt

# Input file should be in multi-fasta format with alternate sRNAs and mRNAs: {sRNA,mRNA,sRNA,mRNA..}

# ============================================================================

# Risearch:

# RIsearch (Version : 1.0)

# Authors : Wenzel A (wenzel@rth.dk), Akbasli E, Gorodkin J

# [gorodkin@rth.dk]

# Command line :

RIsearch -q coli\_sRNA\_v5.fa -t NC\_000913\_CDS\_-200nt\_100nt.fa -d 30 > Risearch\_results.txt