An assessment of bacterial small RNA target prediction programs.

Adrien Pain, Alban Ott, Hamza Amine, Tatiana Rochat, Philippe Bouloc, Daniel Gautheret

REVIEW

RNA Biology 12:5, 509–513; May 2015; Published with license by Taylor & Francis Group, LLC

An assessment of bacterial small RNA target prediction programs

Adrien Pain[†,‡], Alban Ott[†], Hamza Amine, Tatiana Rochat[§], Philippe Bouloc, and Daniel Gautheret*

Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS; Université Paris-Sud; Orsay Cedex, France

[†]These authors equally contributed to this work.

Present affiliations: [‡]Department of Parasites and Insect Vectors; Unit of Insect Vector Genetics and Genomics; CNRS Unit of Hosts, Vectors, and Pathogens (URA3012); Institut Pasteur; Paris, France; [§]Molecular Virology and Immunology (UR892); INRA; Jouy en Josas, France

Keywords: bacteria, sRNA, sRNA target prediction

Most bacterial regulatory RNAs exert their function through base-pairing with target RNAs. Computational prediction of targets is a busy research field that offers biologists a variety of web sites and software. However, it is difficult for a non-expert to evaluate how reliable those programs are. Here, we provide a simple benchmark for bacterial sRNA target prediction based on trusted *E. coli* sRNA/target pairs. We use this benchmark to assess the most recent RNA target predictors as well as earlier programs for RNA-RNA hybrid prediction. Moreover, we consider how the definition of mRNA boundaries can impact overall predictions. Recent algorithms that exploit both conservation of targets and accessibility information offer improved accuracy over previous software. However, even with the best predictors, the number of true biological targets with low scores and non-targets with high scores remains puzzling.

Introduction

Identification of regulatory small RNAs (sRNAs) in bacterial species is rapidly advancing owing to recent improvement of bioinformatics and sequencing technologies.[1] Bacterial sRNAs act usually through sequence-specific binding to target mRNAs, leading to altered stability or translational repression/activation of their targets.[2] In the order of 100 to 200 sRNAs are present in a typical bacterial genome and each sRNA may regulate a dozen or more targets, making the sRNA/mRNA interaction network a significant part of the overall gene regulation network. Biological functions controlled by sRNAs are diverse, related to stress responses as well as carbon and amino acid metabolism, virulence and cellular development.[3]

As sRNA collections build up, the main bottleneck in understanding bacterial sRNA function has now shifted to the question of target prediction. Experimental validation of specific sRNA/mRNA pairs using reporter gene assays or compensatory mutations of binding sites is labor intensive,[4] and high throughput *in vitro* screens are subject to high rates of false positives.[5] On the other hand, computational target prediction is challenging since sRNAs and their target often form imperfect and relatively short RNA-RNA hybrids that are not easily distinguishable from many other hybrids formed by random pairs of transcripts. Yet, computational methods have recently benefited from new algorithms and the introduction of RNA accessibility and conservation information that improved their performances.[6-8]

Here, we evaluate computational target prediction methods using a collection of trusted, experimentally validated sRNA/mRNA pairs in *E. coli*. We designed a benchmark system so that wet lab scientists can easily interpret it for practical purposes. We also evaluated program speeds and abilities to locate the actual duplex region in each sRNA/mRNA pair.

*Correspondence to: Daniel Gautheret; Email: daniel.gautheret@u-psud.fr

Submitted: 09/11/2014; Revised: 02/04/2015; Accepted: 02/13/2015

http://dx.doi.org/10.1080/15476286.2015.1020269

Results

Target predictors

Programs for the identification of RNA-RNA interactions can be grouped in 3 main categories that we term "alignment-like," "inter-RNA" and "independent fold." The "alignment-like" category includes programs such as Guugle[9] and RIsearch[10] that can quickly scan large sequence files for reverse complements of a given RNA sequence. While Guugle looks only for stretches of Watson-Crick or G:U matches, RIsearch implements a Turner-like energy model[11] and allows gaps in the base paired segments. The "inter-RNA" category includes predictors that use a nearest neighbor thermodynamic model restricted to interactions between sRNA and mRNA, neglecting intramolecular base-pairs. Representative programs in this class are Pairfold,[12,13] RNAcofold,[14] TargetRNA,[15] RNAplex,[6] RNAhybrid,[16,17] and RNAduplex.[18] These programs identify intermolecular contacts and compute a binding free energy for the interaction. RNAcofold

and Pairfold achieve this by considering the joint structure predicted after concatenating the 2 RNA sequences. The "independent fold" approach analyzes the secondary structure folding landscape of each RNA independently and models the total binding energy as a sum of 2 contributions. RNAup[19] and IntaRNA[7] are the major programs in this class. Finally, a recent generation of programs including CopraRNA[20,21] and TargetRNA2[22] combine the benefits of the "inter-RNA" class of programs to the use of conservation information. While TargetRNA2 requires conservation of the sRNA region involved in target binding, based on observations by Peer and Margalit,[23] CopraRNA uses the conservation of the interaction itself.

Another, more practical, categorization of RNA target identification software has on one hand actual target predictors, *i.e.* programs that identify and score the most favorable local interaction (CopraRNA, IntaRNA, RNAplex, RNAup, TargetRNA2) and on the other hand programs that model the full hybrid formed between the small RNA and the longer RNA (RNAhybrid, RNAduplex, RNAcofold, Pairfold). While the latter category can be suitable to microRNA/mRNA binding analysis, where the 21 nt RNA can be bound almost entirely to its target, results are more difficult to interpret when dealing with bacterial sRNAs that are generally over 100 nt long and are bound to the target over a small fraction of their size.

The pros and cons of sRNA target prediction algorithms have been reviewed.[24] The goal we set here is to evaluate the ability of programs to differentiate a *bona fide* sRNA/mRNA regulatory pair from a random duplex that may form between any transcripts in a cell. To this aim, we need to rank RNA-RNA duplexes and thus to rely on scores or free energy values provided by the programs. This requirement excludes programs whose energy model is too simplistic, such as Guugle or RIsearch. Although RIsearch does computes free energy values, the authors recommend the program to be used for fast screening of potential pairs, but not to evaluate specific interactions.

We included in our comparison programs from the major categories above, including the most recent bacterial sRNA target predictors CopraRNA, TargetRNA (in its last implementation TargetRNA2[22]) and IntaRNA, programs for local hybrid prediction that can deal with both bacterial sRNAs and miRNAs (RNAup, RNAplex also known as its web server implementation RNApredator[25]), and programs predicting full RNA-RNA hybrids, including Pairfold, RNAcofold, RNAhybrid and RNAduplex.

We ran IntaRNA, RNAplex, RNAup, RNAhybrid, RNAduplex, RNAcofold and Pairfold using Unix command line versions and CopraRNA and TargetRNA2 from their respective web servers (see Suppl. Methods).

Benchmark sets and search domains

Escherichia coli is the organism with the highest number of experimentally validated sRNA/target pairs. We compiled a list of 102 "trusted" sRNA/target pairs involving 22 different sRNAs (**Table S1**), each one bing supported by at least one published experimental evidence from the list described in Methods. This includes *E. coli* pairs from the datasets of Peer and Margalit[23] and Wright et al.,[20] plus a few additions from the literature.[26,27]

Bacterial small RNAs are known to interact preferentially with mRNA 5′ regions with few exceptions to this rule (e.g., see ref.[28,29]). Within our data set, all pairs, except one, are found to occur in the mRNA region located −150/+100 nucleotides (nt) around the start codon (**Fig. S1**). A common practice that consists in restricting sRNA target searches to the −200/+100 region (e.g., see ref.[20]) around the start codon of all mRNAs is thus justified at least for enteric bacteria. Indeed, restricting the search space to the 5′ region efficiently reduces false positive occurrences and computing time. However, bacterial mRNAs often have short 5′ UTRs that do not extend much upstream of the Shine-Dalgarno sequence. Therefore, since the long 200 nt upstream region may be a source of false positives, this artifact can be avoided when exact transcription starts are known. Such information requires massive, strand-specific sequencing of transcripts and is still rarely available. However, this situation is quickly changing and, for certain bacteria including *E. coli*, accurate UTR information is already available.[30] Thus, we asked whether a precise delineation of UTRs could improve target identification. When possible, we performed predictions on 2 target data sets: one made of arbitrary −200 / +100 nt regions around each start codon (hereafter named "default UTRs") and one made of regions predicted from publicly available strand-specific RNA-seq data (hereafter named "real UTRs"). Note that we also included in our analysis 5′ regions of open reading frames located in polycistronic genes, as these regions are also subject to regulation as exemplified by the *manXYZ* operon mRNA targeted at 2 locations by the sRNA SgrS,[31] or the *Staphylococcus aureus opp3* operon mRNA containing at least 2 RsaE-binding sites.[32] Details of the 5′ UTR definition procedures are provided in Methods and Suppl. Methods.

Rank-based prediction performance

Our first performance measure aimed at addressing the following question: how many putative targets for a given sRNA should be experimentally tested (for instance using a reporter assay) before finding the first true target? Programs were run using the best options suggested by authors. RNA accessibility information, which is known to improve target prediction significantly, was selected when possible.[19,33] Programs from websites (*i.e.*, CopraRNA, TargetRNA2) were run using preselected mRNA target sequences that correspond to "default UTRs" as defined above. These servers cannot take arbitrary lists of target sequences as input. Therefore tests with "real UTR" targets were limited to programs ran from the command line. Parameter details are presented in Suppl. Methods.

Putative targets were ranked by scores or free energy values provided by each program. The median rank is the median position of the highest ranking validated target among predicted targets for all sRNAs in the dataset (**Fig. 1A–B**). Median ranks range from 4 (CopraRNA) to 1340 (RNAcofold). Expectedly, programs designed for evaluating local interactions (CopraRNA, intaRNA, RNAplex, RNAup, TargetRNA2) perform far better than programs computing complete hybrids. In this latter category, RNAhybrid, RNAduplex, perform significantly better due to their free energy scores restricted to intermolecular base-pairs.

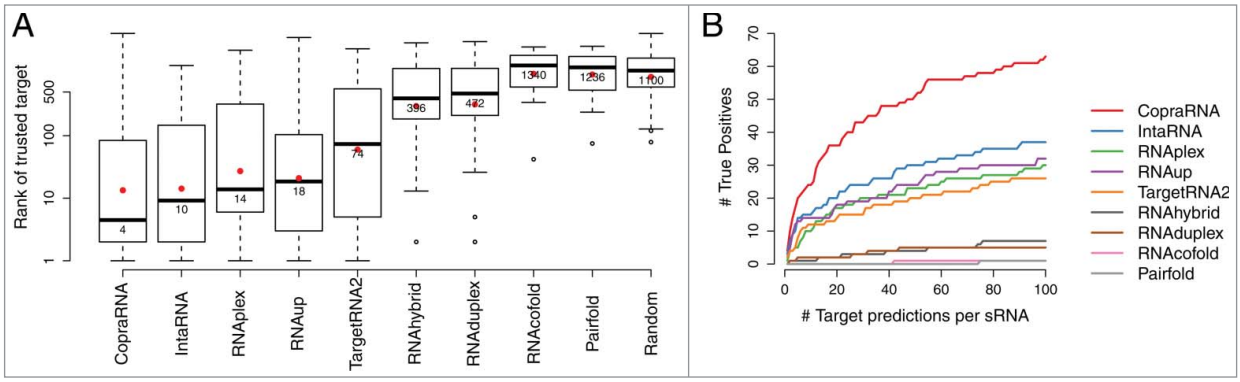

Figure 1. Two representations of target predictor performances. (**A**) Rank distribution of trusted targets (lower=better). For each program, the distribution shows the ranks of the best ranking target of each sRNA (22 sRNAs). Horizontal lines and numbers indicate median ranks, red dots indicate mean 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).

The least performing programs RNAcofold and Pairfold are both "concatenation" algorithms that are not intended to evaluate the binding energy of the complex and therefore cannot perform any better than a random ranking.

Our test probably underestimates the performance of programs since some of the pairs ranking before the first confirmed target are also possibly true targets. Nevertheless, some sRNA/target pairs rank very poorly, even by the best performing predictors (**Fig. 1A**), which questions their presence in our benchmark set. Indeed, some sRNA/target pairs are part of the trusted pair list because target expression is affected by the amount of the corresponding sRNA. However, this effect may occur indirectly in the absence of any contact between the 2 RNAs. To test whether putative indirect pairs may have impacted our benchmark, we considered only pairs for which we had direct evidence of sRNA/target interaction (82 pairs, see **Table S1** and Methods). There was no improvement of prediction accuracy over the full data set (**Fig. S2A**), suggesting the presence of indirect pairs in our dataset is not a reason for the low ranking of some actual targets. Further inspection of poorly ranked trusted sRNA/target pairs shows that the same pairs tend to be missed by all predictors. Typical mispredictions involve pairs with target sites located 5′ of upstream genes such as in RyhB/*fur* mRNA[34] and hybrid structures formed by double kissing complexes such as in OxyS/*fhlA* mRNA.[35]

We tested the effect of using the "real" UTR sequences (RNA-seq-based UTRs) of mRNAs instead of arbitrary upstream regions (−200 nt before AUG codon) on prediction performance (**Table 1**). Using "real" UTRs improved results moderately (AUC gains of 4.9% to 17.5%) albeit not significantly for most predictors. We observed a significant AUC improvement (Wilcoxon P = 0.03) only for concatenation methods RNAcofold and Pairfold, however this should be put into perspective as these methods had a very small AUC to start with. Yet, we may hypothesize that more accurate mRNA boundaries can improve free energy estimations of the hybrid in some cases.

The best prediction accuracy overall was obtained using CopraRNA, suggesting an advantage to a method using both RNA accessibility and conservation information, in line with the authors own estimation.[20] Although CopraRNA performed best overall, it should be noted that its use is limited compared to other programs since it requires an homolog set for each sRNA of interest and targets are evaluated only in a predefined list of genomes.

Table 1. Area under ROC-like curve (AUC) for target prediction with 9 programs, using default UTR or "real" RNA-seq-derived UTRs

Program	AUC default UTR	AUC real UTR	AUC gain (%)
CopraRNA	46.41	na*	na*
IntaRNA	27.02	28.83	6.3
RNAplex	20.81	21.88	4.9
RNAup	22.75	27.59	17.5
TargetRNA2	18.52	na*	na*
RNAhybrid	4.14	4.38	5.5
RNAduplex	3.86	3.55	−8.7
RNAcofold	0.57	1.42	59.9
Pairfold	0.25	1.22	79.5

*Programs CopraRNA and TargetRNA2 can be run only using default UTRs.

Assessing other predictor qualities

While the above comparison used the ranks of the best scoring targets of each sRNA, we found interesting to analyze ranks of all trusted targets combined. When all 102 trusted targets are considered, median ranks for all programs increase mathematically (**Fig. S3**). However, considering that our test set contains only 4-5 targets per sRNA, these new median ranks ranging from 50 to

Table 2. Recall of experimentally demonstrated base pairs (%)

IntaRNA/CopraRNA*	76.7
RNAplex	73.6
TargetRNA2	55.9
RNAup	78.9

*IntaRNA and CopraRNA predict the same hybrid region.

Table 3. Average run time for matching 1 sRNA to all 4317 *E. coli* mRNAs

TargetRNA2 (web site)	1 min
RNAplex (incl. accessibility calculation)	1 min
RNAhybrid	2 min
RNAcofold	23 min
RNAup	55 min
IntaRNA	66 min
RNAduplex	75 min
CopraRNA (website)	105 min
Pairfold	120 min

600 for the best predictors are a good indication of how predicted target sets remain populated mostly by unconfirmed targets.

We also tested how accurately each program was able to determine the correct hybrid regions assuming a correct sRNA/mRNA pair was detected. Since experimental validations of sRNA/target pairs do not always specify the duplex structure, we focused on the set of 55 pairs supported by compensatory mutations (see Methods). We excluded from this test all programs that predict a hybrid structure involving the full sRNA. **Table 2** shows the fraction of compensatory mutations that are indeed found within the predicted hybrid regions. Interestingly, this ratio is far from perfect (56% to 79%) even though we only considered correct RNA/target pairs.

Finally, we compared the speed of the different programs (**Table 3**). Program Texte vary by 2 orders of magnitude, the fastest programs are able to match a sRNA to the complete set of *E. coli* mRNAs in about one minute while some require one to 2 hours to complete the same task. The best performing predictors are generally the slowest, with the exception of RNAplex, which achieves both high prediction accuracy and fast running time.

Conclusion

RNA target prediction programs have improved continuously over the past 10 years owing to algorithmic innovation, as well as the exploitation of structural and evolutionary constraints. We confirm here that CopraRNA, a predictor that uses both RNA accessibility and evolutionary conservation of the interaction outperforms other programs in terms of ranking actual sRNA/mRNA pairs. However, conservation data is not often available and programs that predict sRNA targets based on sequence alone are still much needed. IntaRNA, RNAplex and RNAup are the best predictors in this class. The presence of RNAplex in this shortlist is surprising as this program is up to 100 times faster than competitors owing to a simplified structural model. Furthermore, our tests do not show significant accuracy gains from using RNA-seq derived UTRs instead of fixed-length 5′ regions, supporting the common practice for defining mRNA target regions.

In spite of software improvements, this study is also a reminder that the list of highest scoring mRNAs for a given *E. coli* sRNA remains populated by a majority of unconfirmed targets. One possible explanation for this could be that many true targets have not been validated yet, which is unlikely in this well-studied bacterium. Alternatively, *in vivo* sRNA-mRNA interaction can be positively affected by RNA-chaperon proteins (e.g. Hfq), which may significantly confuse computed targets ranks. Another reason that may explain the relatively poor ranking of many true targets is related to the way programs score an RNA-RNA pair. Predictors score targets based on binding free energy values (as well as conservation for CopraRNA and targetRNA2). However, a high binding energy does not necessarily imply an efficient processing of the mRNA target by the Hfq–sRNA complex. The situation in bacteria could be similar to that of miRNA-target pairs in eukaryotes, where, due to miRNA turnover[36] lower stability of the microRNA/target duplex may favor transient interactions and therefore an efficient down-regulation of multiple targets.[37] The large accuracy gain of CopraRNA obtained by introducing phylogenetic conservation in its scoring system illustrates again that there is much more than thermodynamics in understanding sRNA-based regulation.

Methods

Definition of UTR regions: «default» UTR regions were automatically extracted from MG1655 *E. coli* K12 genome, selecting the −200/+100 fragments around each start codon (4317 regions). «Real» UTR regions were extracted based on RNA-seq-derived transcriptional start site (TSS).[30] For each TSS, the region from the TSS to 100 nt past the start codon was extracted. Further details are provided in Suppl. methods.

Experimentally validated pairs (**Table S1**) were classified as described based on supporting experimental evidence[23]: (a) sRNA affecting the level of a protein or a translational fusion; (b) sRNA gene affecting the level of a targeted mRNA; (c) sRNA and mRNA forming a stable duplex *in vitro*; (d) sRNA mutation affecting a regulation; (e) sRNA mutations affecting a regulation rescued by compensatory mutations; (f) *in vitro* chemical/enzymatic probing of the sRNA/target duplex; (g) sRNA/target duplex cleaved by RNase III; (h) sRNA preventing or allowing ribosome binding demonstrated by toe-printing assay. We considered that the criteria c) to h) demonstrated direct or near-direct interactions for 82 sRNA/mRNA duplexes

Performance measures (**Fig. 1 and S2**) were obtained by running target predictors with each sRNA against the complete set of *E. coli* mRNAs (using "default" or "real" UTRs). Ranks of trusted targets (**Fig. 1A and S2A**) were computed based on the best ranking target of each sRNA. For the "random" distribution, we performed for each sRNA a set of N random draws between one and the total number of *E. coli* mRNAs, where N is the number of known targets for this sRNA, and kept the lowest ranking value. Receiver operating characteristic (ROC)-like curves (**Fig. 1B and S2B**) were obtained by counting the number of true targets found among the best X candidates (shown on X axis). Program versions, options and command lines are provided in Suppl. methods. Accessibility information was computed prior

to RNAplex runs using RNAplfold.[14] CopraRNA requires homologous sRNA sequences from other species and we used those provided by the CopraRNA server whenever possible. For IstR, RseX, RydC we provided our own sets (see Suppl. Files).

Recall of experimentally validated base pairs: 55 sRNA/target pairs have experimental support in the form of compensatory mutations. For each pair, we measured the ratio of experimentally validated bps that were included in the predicted hybrid region. This ratio was then averaged over all 55 sRNA/target pairs to produce percentages in **Table 2**. As TargetRNA2 could predict only 49 of the 55 experimentally supported pairs, the average value was corrected accordingly.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Funding

This work was funded by Agence Nationale pour la Recherche (grant 2010-BLAN-1602-01) 'Duplex-omics'.

Supplemental Material

Supplemental data for this article can be accessed on the publisher's website.

References

1. Lu X, Goodrich-Blair H, Tjaden B. Assessing computational tools for the discovery of small RNA genes in bacteria. RNA 2011; 17:1635-47; PMID:21768221; http://dx.doi.org/10.1261/rna.2689811
2. Gottesman S, Storz G. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. Cold Spring Harb Perspect Biol 2011; 3:pii: a003798; PMID:20980440; http://dx.doi.org/10.1101/cshperspect.a003798
3. Storz G, Vogel J, Wassarman KM. Regulation by small RNAs in bacteria: expanding frontiers. Mol Cell 2011; 43:880-91; PMID:21925377; http://dx.doi.org/10.1016/j.molcel.2011.08.022
4. Vogel J, Bartels V, Tang TH, Churakov G, Slagter-Jager JG, Huttenhofer A, Wagner EG. RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria. Nucleic Acids Res 2003; 31:6435-43; PMID:14602901; http://dx.doi.org/10.1093/nar/gkg867
5. Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, Binnewies TT, Hinton JC, Vogel J. Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. PLoS Genet 2008; 4:e1000163; PMID:18725932; http://dx.doi.org/10.1371/journal.pgen.1000163
6. Tafer H, Hofacker IL. RNAplex: a fast tool for RNA-RNA interaction search. Bioinformatics 2008; 24:2657-63; PMID:18434344; http://dx.doi.org/10.1093/bioinformatics/btn193
7. Busch A, Richter AS, Backofen R. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. Bioinformatics 2008; 24:2849-56; PMID:18940824; http://dx.doi.org/10.1093/bioinformatics/btn544
8. Will S, Joshi T, Hofacker IL, Stadler PF, Backofen R. LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. RNA 2012; 18:900-14; PMID:22450757; http://dx.doi.org/10.1261/rna.029041.111
9. Gerlach W, Giegerich R. GUUGle: a utility for fast exact matching under RNA complementary rules including G-U base pairing. Bioinformatics 2006; 22:762-4; PMID:16403789; http://dx.doi.org/10.1093/bioinformatics/btk041
10. Wenzel A, Akbasli E, Gorodkin J. RIsearch: fast RNA-RNA interaction search using a simplified nearest-neighbor energy model. Bioinformatics 2012; 28:2738-46; PMID:22923300; http://dx.doi.org/10.1093/bioinformatics/bts519
11. Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol 1999; 288:911-40; PMID:10329189; http://dx.doi.org/10.1006/jmbi.1999.2700
12. Andronescu M, Zhang ZC, Condon A. Secondary structure prediction of interacting RNA molecules. J Mol Biol 2005; 345:987-1001; PMID:15644199; http://dx.doi.org/10.1016/j.jmb.2004.10.082
13. Tulpan D, Andronescu M, Chang SB, Shortreed MR, Condon A, Hoos HH, Smith LM. Thermodynamically based DNA strand design. Nucleic Acids Res 2005; 33:4951-64; PMID:16145053; http://dx.doi.org/10.1093/nar/gki773
14. Bernhart SH, Tafer H, Muckstein U, Flamm C, Stadler PF, Hofacker IL. Partition function and base pairing probabilities of RNA heterodimers. Algorithms Mol Biol 2006; 1:3; PMID:16722605; http://dx.doi.org/10.1186/1748-7188-1-3
15. Tjaden B. TargetRNA: a tool for predicting targets of small RNA action in bacteria. Nucleic Acids Res 2008; 36:W109-13; PMID:18477632; http://dx.doi.org/10.1093/nar/gkn264
16. Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. Fast and effective prediction of microRNA/target duplexes. RNA 2004; 10:1507-17; PMID:15383676; http://dx.doi.org/10.1261/rna.5248604
17. Kruger J, Rehmsmeier M. RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res 2006; 34:W451-4; PMID:16845047; http://dx.doi.org/10.1093/nar/gkl243
18. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P. Fast folding and comparison of RNA secondary structures. Monatsh Chem 1994; 125:167-88; http://dx.doi.org/10.1007/BF00818163
19. Muckstein U, Tafer H, Hackermuller J, Bernhart SH, Stadler PF, Hofacker IL. Thermodynamics of RNA-RNA binding. Bioinformatics 2006; 22:1177-82; PMID:16446276; http://dx.doi.org/10.1093/bioinformatics/btl024
20. Wright PR, Richter AS, Papenfort K, Mann M, Vogel J, Hess WR, Backofen R, Georg J. Comparative genomics boosts target prediction for bacterial small RNAs. Proc Natl Acad Sci U S A 2013; 110:E3487-96; PMID:23980183; http://dx.doi.org/10.1073/pnas.1303248110
21. Wright PR, Georg J, Mann M, Sorescu DA, Richter AS, Lott S, Kleinkauf R, Hess WR, Backofen R. CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. Nucleic Acids Res 2014; 42 (Web Server issue):W119-23; PMID:24838564
22. Kery MB, Feldman M, Livny J, Tjaden B. TargetRNA2: identifying targets of small regulatory RNAs in bacteria. Nucleic Acids Res 2014; 42(Web Server issue):W124-9; PMID:24753424
23. Peer A, Margalit H. Accessibility and evolutionary conservation mark bacterial small-rna target-binding regions. J Bacteriol 2011; 193:1690-701; PMID:21278294; http://dx.doi.org/10.1128/JB.01419-10
24. Backofen R, Hess WR. Computational prediction of sRNAs and their targets in bacteria. RNA Biol 2010; 7:33-42; PMID:20061798; http://dx.doi.org/10.4161/rna.7.1.10655
25. Eggenhofer F, Tafer H, Stadler PF, Hofacker IL. RNApredator: fast accessibility-based prediction of sRNA targets. Nucleic Acids Res 2011; 39:W149-54; PMID:21672960; http://dx.doi.org/10.1093/nar/gkr467
26. Jorgensen MG, Nielsen JS, Boysen A, Franch T, Moller-Jensen J, Valentin-Hansen P. Small regulatory RNAs control the multi-cellular adhesive lifestyle of *Escherichia coli*. Mol Microbiol 2012; 84:36-50; PMID:22250746; http://dx.doi.org/10.1111/j.1365-2958.2012.07976.x
27. Beisel CL, Storz G. The base-pairing RNA spot 42 participates in a multioutput feedforward loop to help enact catabolite repression in *Escherichia coli*. Mol Cell 2011; 41:286-97; PMID:21292161; http://dx.doi.org/10.1016/j.molcel.2010.12.027
28. Lease RA, Belfort M. A trans-acting RNA as a control switch in *Escherichia coli*: DsrA modulates function by forming alternative structures. Proc Natl Acad Sci U S A 2000; 97:9919-24; PMID:10954740; http://dx.doi.org/10.1073/pnas.170281497
29. Pfeiffer V, Papenfort K, Lucchini S, Hinton JC, Vogel J. Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. Nat Struct Mol Biol 2009; 16:840-6; PMID:19620966; http://dx.doi.org/10.1038/nsmb.1631
30. Li S, Dong X, Su Z. Directional RNA-seq reveals highly complex condition-dependent transcriptomes in *E. coli* K12 through accurate full-length transcripts assembling. BMC Genomics 2013; 14:520; PMID:23899370; http://dx.doi.org/10.1186/1471-2164-14-520
31. Rice JB, Balasubramanian D, Vanderpool CK. Small RNA binding-site multiplicity involved in translational regulation of a polycistronic mRNA. Proc Natl Acad Sci U S A 2012; 109:E2691-8; PMID:22988087; http://dx.doi.org/10.1073/pnas.1207927109
32. Bohn C, Rigoulay C, Chabelskaya S, Sharma CM, Marchais A, Skorski P, Borezée-Durant E, Barbet R, Jacquet E, Jacq A, et al. Experimental discovery of small RNAs in *Staphylococcus aureus* reveals a riboregulator of central metabolism. Nucleic Acids Res 2010; 38:6620-36; PMID:20511587; http://dx.doi.org/10.1093/nar/gkq462
33. Richter AS, Backofen R. Accessibility and conservation: general features of bacterial small RNA-mRNA interactions? RNA Biol 2012; 9:954-65; PMID:22767260; http://dx.doi.org/10.4161/rna.20294
34. Vecerek B, Moll I, Blasi U. Control of Fur synthesis by the non-coding RNA RyhB and iron-responsive decoding. EMBO J 2007; 26:965-75; PMID:17268550; http://dx.doi.org/10.1038/sj.emboj.7601553
35. Argaman L, Altuvia S. fhlA repression by OxyS RNA: kissing complex formation at two sites results in a stable antisense-target RNA complex. J Mol Biol 2000; 300:1101-12; PMID:10903857; http://dx.doi.org/10.1006/jmbi.2000.3942
36. Hutvagner G, Zamore PD. A microRNA in a multiple-turnover RNAi enzyme complex. Science 2002; 297:2056-60; PMID:12154197; http://dx.doi.org/10.1126/science.1073827
37. Kozomara A, Hunt S, Ninova M, Griffiths-Jones S, Ronshaugen M. Target repression induced by endogenous microRNAs: large differences, small effects. PloS One 2014; 9:e104286; PMID:25141277; http://dx.doi.org/10.1371/journal.pone.0104286