Freiburg RNA Tools webserver

Patrick R. Wright[1,2], Martin Mann[1], Dragos A. Sorescu[1], Sita J. Lange[1], Omer S. Alkhnbashi[1], Dominic Rose[1], Steffen Heyne[1], Andreas S. Richter[1], Wolfgang R. Hess[2], Jens Georg[2], Anke Busch[1], Sebastian Will[1], Rolf Backofen[1]

[1]Chair of Bioinformatics, University of Freiburg, Germany; [2]Genetics and Experimental Bioinformatics, University of Freiburg, Germany

UNI FREIBURG

Department of Computer Science

RNA Tools

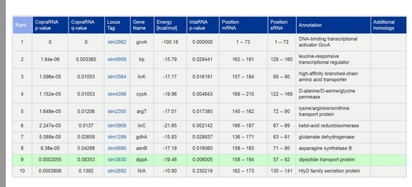
- **CopraRNA** — sRNA Targeting
- **LocARNA** — Alignment & Folding
- **CARNA** — Ensemble Alignment
- **CRISPRmap** — CRISPR Conservation
- **ExpaRNA** — Exact Matching
- **INFORNA** — Sequence Design
- **IntaRNA** — RNA-RNA Interaction
- **MARNA** — Structure Alignment

Overview

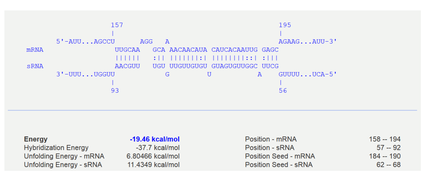
Due to novel experimental methods on the genomic scale, biologists are struggling with ever increasing magnitudes of data that can, in many cases, only be harnessed by previous bioinformatics analyses. Currently many tools are either only accessible on the command line and web servers tend to lack easy usability. The Freiburg RNA Tools webserver aims at supplying an easy to use and comprehensive web resource for RNA analysis, also for non-adept users. We designed a webserver framework that simplifies the access to our RNA analysis tools. The tools are accompanied by extensive help pages and direct help requests are rapidly answered. All tools incorporate individual post processing steps that aid result interpretation. The results can be viewed in the browser and/or downloaded for further local analysis or archiving. Individual job descriptions can be entered by the user, thus alleviating personal online archiving. Furthermore, results are stored for 30 days. The Freiburg RNA tools webserver currently integrates eight tools for RNA analysis. It includes CopraRNA [1] (sRNA target prediction), LocARNA (alignment and folding) [2], CARNA (ensemble alignment) [3], MARNA (structure alignment) [4], ExpaRNA (exact matching) [5], INFORNA (sequence design) [6], IntaRNA (RNA-RNA interaction) [7] and CRISPRmap (CRISPR conservation) [8]. The addition of several further tools is under construction.
The tools are available at: http://rna.informatik.uni-freiburg.de.

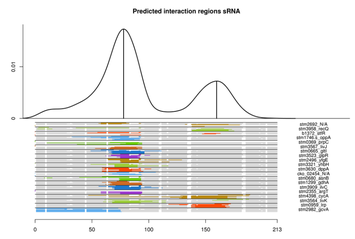
CopraRNA & IntaRNA

CopraRNA [1] is a tool for sRNA target prediction. It computes whole genome target predictions by combination of distinct whole genome IntaRNA [7] predictions derived from homologous sRNAs in different organisms. The images show an exemplary output for the well studied, amino acid related, GcvB sRNA. Its function is correctly predicted by CopraRNA.



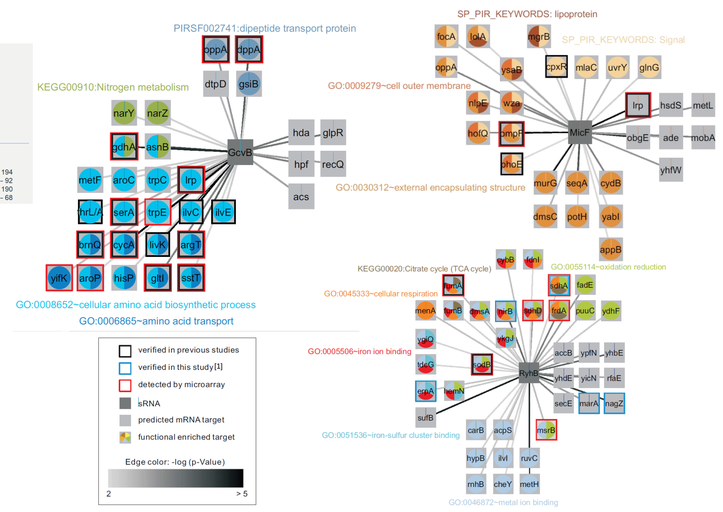
Rank	CopraRNA p-value	CopraRNA q-value	Locus Tag	Gene Name	Energy [kcal/mol]	IntaRNA p-value	Position mRNA	Position sRNA	Annotation	Additional homologs
1	0	0	stm2982	gcvA	-100.18	0.000000	1 – 73	1 – 73	DNA-binding transcriptional activator GcvA	
2	1.84e-06	0.003365	stm0959	lrp	-15.79	0.029441	162 – 191	129 – 160	leucine-responsive transcriptional regulator	
3	1.096e-05	0.01053	stm3564	livK	-17.17	0.016191	157 – 184	65 – 90	high-affinity branched-chain amino acid transporter	
4	1.152e-05	0.01053	stm4398	cycA	-19.96	0.004843	169 – 210	122 – 169	D-alanine/D-serine/glycine permease	
5	1.648e-05	0.01206	stm2355	argT	-17.01	0.017380	145 – 162	72 – 90	lysine/arginine/ornithine transport protein	
6	2.247e-05	0.0137	stm3909	ilvC	-21.85	0.002142	166 – 187	67 – 89	ketol-acid reductoisomerase	
7	5.068e-05	0.02859	stm1299	gdhA	-15.83	0.028937	138 – 171	63 – 91	glutamate dehydrogenase	
8	9.38e-05	0.04288	stm0680	asnB	-17.19	0.016060	158 – 183	71 – 90	asparagine synthetase B	
9	0.0002055	0.08353	stm3630	dppA	-19.46	0.006005	158 – 194	57 – 92	dipeptide transport protein	
10	0.0003806	0.1392	stm2692	N/A	-10.90	0.230219	162 – 173	130 – 141	HlyD family secretion protein	



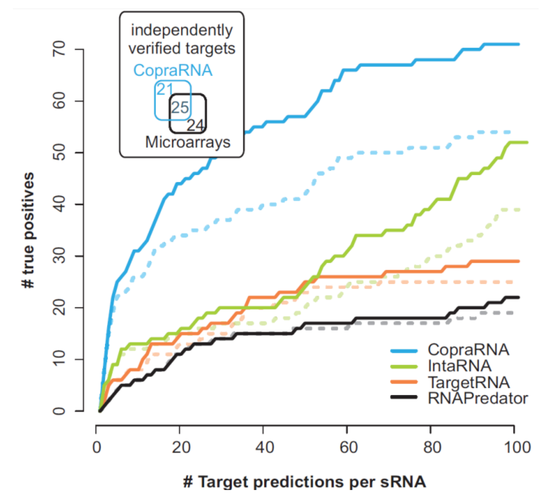


The density plots above, give a clear impression of the interacting regions within the mRNAs and sRNAs. Furthermore the degree of conservation throughout the organisms participating in the analysis becomes clearer.

Searching for functionally enriched terms within the top predictions, regularly yields insight into the correct *in vivo* function of an investigated sRNA. This alleviates construction of functional networks. See the examples below (GcvB, MicF, RyhB).



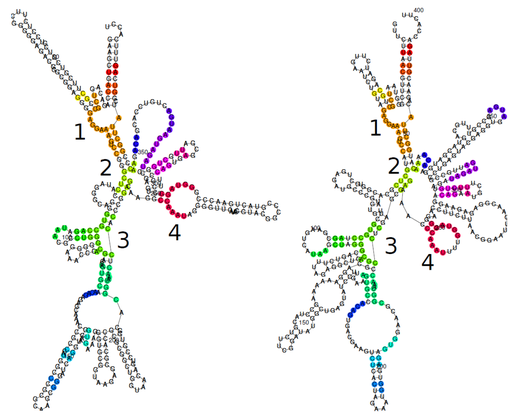
In some cases the interest is solely focused on single predictions for two or more interacting RNAs, or no homologous sRNAs are present for a whole genome target prediction. In these cases we suggest using IntaRNA [7].



Employing an extensive dataset of 18 enterobacterial sRNAs and 102 experimentally verified interactions we compared CopraRNA to 3 other state of the art sRNA target prediction approaches. CopraRNA yielded considerably superior results with respect to sensitivity. Furthermore CopraRNA appears to be at least as reliable as pulse expression microarray experiments.

ExpaRNA

ExpaRNA [5] is a tool for very fast comparison of RNAs by exact local matches. Instead of computing a full sequence-structure alignment, ExpaRNA efficiently computes the best arrangement of sequence-structure motifs common to two RNAs. Finding identical motifs is not directly addressed by sequence-structure alignment tools and they may remain hidden. In addition, the predicted set of motifs can be used as anchor constraints to speed-up and guide Sankoff-style alignment methods like LocARNA [2].

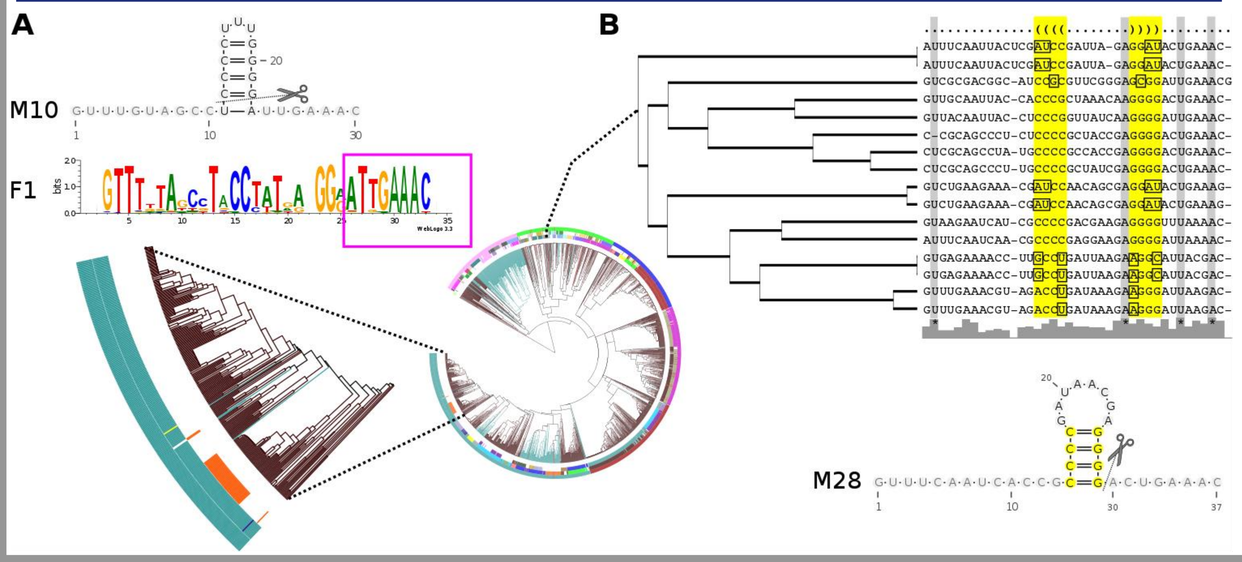


The figure shows annotated structures from the ExpaRNA output. Regions of exact pattern matches have the same color. Shown are two bacterial RNase~P RNAs from *Escherichia coli* and *Bacillus subtilis*.

CRISPRmap

The CRISPR-Cas system degrades foreign genetic elements and is wide-spread in bacteria and archaea. Central to CRISPR-Cas immune systems are repeated RNA sequences that serve as Cas-protein-binding templates. Their classification is mainly based on the architectural composition of associated Cas proteins; directly considering repeat evolution, however, is essential to complete the picture.

We compiled the largest dataset of CRISPRs to date; performed comprehensive, independent clustering analyses; and identified a novel set of 40 conserved sequence families and 33 potential structure motifs for Cas-endoribonucleases with some distinct conservation patterns.



CRISPRmap [8] is an easy-to-use web server that provides an automated assignment of newly sequenced CRISPRs to our classification systems and enables more informed choices on future hypotheses in CRISPR-Cas research.

As a visual map of both bacterial and archaeal CRISPR domains, we combined our categorisation into repeat families and motifs with a hierarchical tree based on sequence-and-structure similarities. This CRISPRmap tree details relationships between individual repeats and whole families and motifs.

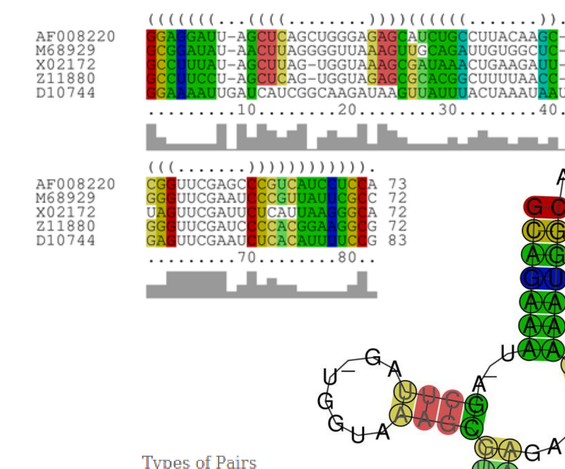
Left: Highlights the advantage of independent clustering approaches in the CRISPRmap tree.

(**A**) CRISPRs in the largest sequence family, F1, are mostly unstructured; however, for 50 CRISPRs also a conserved structure motif, M10, was identified.

(**B**) Structure motif M28 could not be identified by sequence conservation, but has been verified via mutational analyses and contains many compensatory base pair mutations (black squares).

LocARNA & MARNA

LocARNA [2] is a tool for multiple alignment of RNA molecules. LocARNA requires only RNA sequences as input and will simultaneously fold and align the input sequences. LocARNA outputs a multiple alignment together with a consensus structure. For the folding it makes use of a very realistic energy model for RNAs which is also employed by RNAfold of the Vienna RNA package (or mfold). For the alignment it features RIBOSUM-like similarity scoring and realistic gap cost. MARNA [4] is also offered, yet LocARNA supersedes it.

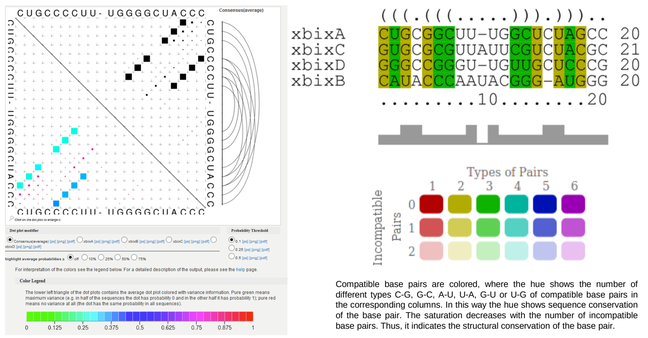


Compatible base pairs are colored, where the hue shows the number of different types C-G, G-C, A-U, U-A, G-U or U-G of compatible base pairs in the corresponding columns. In this way the hue shows sequence conservation of the base pair. The saturation decreases with the number of incompatible base pairs. Thus, it indicates the structural conservation of the base pair.

The images display an example alignment of multiple tRNA sequences and a visualization of the consensus structure.

CARNA

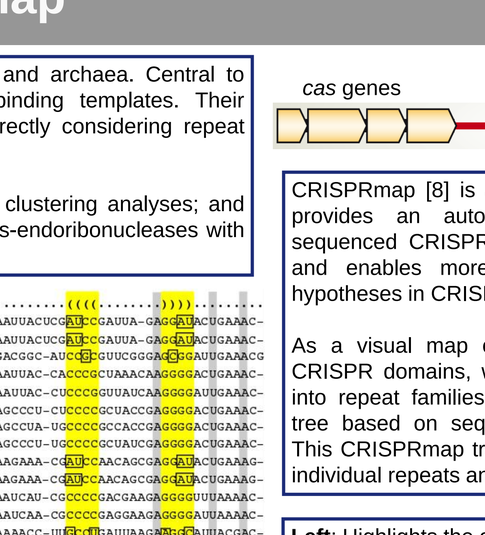
In contrast to LocARNA [2], CARNA [3] does not pick the most likely consensus structure, but computes the alignment that fits best to all likely structures simultaneously. Hence, CARNA is particularly useful when aligning RNAs like riboswitches, which have more than one stable structure. Also, CARNA is not limited to nested structures, but is able to align arbitrary pseudoknots.



Compatible base pairs are colored, where the hue shows the number of different types C-G, G-C, A-U, U-A, G-U or U-G of compatible base pairs in the corresponding columns. In this way the hue shows sequence conservation of the base pair. The saturation decreases with the number of incompatible base pairs. Thus, it indicates the structural conservation of the base pair.

CARNA optimizes all structural similarities in the input simultaneously, for example across an entire RNA structure ensemble. Even when compared to already costly Sankoff-style alignment, CARNA solves an intrinsically much harder problem by applying advanced, constraint-based, algorithmic techniques.

INFORNA



INFORNA [6] is a server for the design of RNA sequences that fold into a given pseudo-knot free RNA secondary structure.

References

1. Patrick R. Wright, Andreas S. Richter, Kai Papenfort, Martin Mann, Jörg Vogel, Wolfgang R. Hess, Rolf Backofen and Jens Georg,Comparative genomics boosts target prediction for bacterial small RNAs, Currently under review.
2. Sebastian Will, Kristin Reiche, Ivo L. Hofacker, Peter F. Stadler, and Rolf Backofen, Inferring non-coding RNA families and classes by means of genome-scale structure-based clustering, PLoS Computational Biology, 3 no. 4, pp. e65, 2007.
3. Dragos A. Sorescu, Mathias Moehl, Martin Mann, Rolf Backofen, and Sebastian Will, CARNA - alignment of RNA structure ensembles, Nucleic Acids Reseach, 40 no. W1 pp. W49-W53, 2012.
4. Sven Siebert and Rolf Backofen, MARNA: multiple alignment and consensus structure prediction of RNAs based on sequence structure comparisons, Bioinformatics 2005, Volume 21, Issue 16, 3352-3359.
5. Steffen Heyne, Sebastian Will, Michael Beckstette, and Rolf Backofen, Lightweight comparison of RNAs based on exact sequence-structure matches, Bioinformatics, 25 no. 16 pp. 2095-2102, 2009.
6. Anke Busch and Rolf Backofen, INFO-RNA - a fast approach to inverse RNA folding, Bioinformatics, 22 no. 15 pp. 1823-31, 2006.
7. Anke Busch, Andreas S. Richter, and Rolf Backofen, IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions, Bioinformatics, 24 no. 24 pp. 2849-56, 2008.
8. Sita J. Lange, Omer S. Alkhnbashi, Dominic Rose, Sebastian Will and Rolf Backofen, CRISPRmap: an automated classification of repeat conservation in prokaryotic adaptive immune systems, Currently under review.