Freiburg RNA tools – a central online resource for RNA-focused
research and teaching
- Supplementary material -

Martin Raden¹, Syed M. Ali¹, Omer S. Alkhnbashi¹, Anke Busch², Fabrizio Costa³,
Jason A. Davis⁴, Florian Eggenhofer¹, Rick Gelhausen¹, Jens Georg⁵, Steffen Heyne⁶,
Michael Hiller⁶, Kousik Kundu⁷,¹⁰, Robert Kleinkauf¹, Steffen C. Lott⁵,
Mostafa M. Mohamed¹, Alexander Mattheis¹, Milad Miladi¹, Andreas S. Richter¹¹,
Sebastian Will¹², Joachim Wolff¹, Patrick R. Wright¹³, Rolf Backofen¹,¹⁴

¹Bioinformatics, Computer Science, University of Freiburg, Georges-Koehler-Allee 106, 79110 Freiburg, Germany,
²Institute of Molecular Biology (IMB), Ackermannweg 4, 55128 Mainz, Germany, ³Department of Computer Science,
University of Exeter, Exeter EX4 4QF, UK, ⁴Coreva Scientific, Kaiser-Joseph-Str 198-200, 79098 Freiburg, Germany,
⁵Genetics and Experimental Bioinformatics, University of Freiburg, Schanzlestrae 1, 79104 Freiburg, Germany, ⁶Max
Planck Institute of Immunobiology and Epigenetics, St¨ubeweg 51, 79108 Freiburg, Germany, ⁷Max Planck Institute
of Molecular Cell Biology and Genetics, Dresden, Germany, ⁸Max Planck Institute for the Physics of Complex Systems,
Dresden, Germany, ⁹Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Long Road,
Cambridge CB2 0PT, UK, ¹⁰Department of Human Genetics, The Wellcome Trust Sanger Institute, Hinxton Cambridge
CB10 HH, UK, ¹¹Genedata AG, Margarethestrasse 38, 4053 Basel, Switzerland, ¹²Theoretical Biochemistry Group,
University of Vienna, Wahringerstrae 17, 1090, Vienna, Austria, ¹³Department of Clinical Research, Clinical Trial Unit,
University of Basel Hospital, Schanzenstrasse 55, 4031 Basel, Switzerland, and ¹⁴Centre for Biological Signalling Studies
(BIOSS), University of Freiburg, Schanzenstr. 18, 79104 Freiburg, Germany

Contents

1 CopraRNA - Application example ........................................... 2
  1.1 Input ........................................................................ 2
  1.2 Results and Interpretation ............................................ 2
    1.2.1 What does GcvB regulate? ........................................ 3
    1.2.2 Where does the sRNA bind the target mRNAs? ........... 3
    1.2.3 Which sRNA domains are used? ............................... 4

2 CMV - Visualization for RNA and Protein family models and their comparisons .................... 6
  2.1 Input ....................................................................... 6
  2.2 Results and Summary ................................................ 7

3 Teaching - Available Algorithms ........................................................................ 8
  3.1 RNA structure prediction ............................................. 8
  3.2 RNA-RNA interaction prediction ................................. 8
  3.3 Pairwise sequence alignment ....................................... 8
  3.4 Multiple sequence alignment ...................................... 8
  3.5 Evolutionary tree construction ................................... 8
1 CopraRNA - Application example

CopraRNA’s capabilities in characterizing bacterial small RNAs (sRNAs) via genome-wide target prediction will be demonstrated using GcvB [1]. The biological functions of the targets of an sRNA are the central indicator for the sRNA’s regulatory role. CopyraRNA provides analysis and visualization output in this regard and additionally supplies integrated functional enrichment analysis based on the DAVID resource [2].

1.1 Input

CopraRNA requires at least 3 homologous sRNA sequences from 3 distinct organisms in FASTA format as input. Furthermore, each organism’s genome has to be part of the NCBI Reference Sequence (RefSeq) database. Usually a user has a specific organism he is especially interested in. This organism of interest can be selected upon input. Figure 1 shows a correctly formatted input file for GcvB. Finally, the user needs to specify whether putative targeting shall be predicted around the start or stop codon.

>NC_003197
acuuccugacgcgaacgaaaguuuuacggaaucggccuuguaguguuugugucugcuugcuuuuuuuggcuuacgguaguguuguguu
ugugugugugugucuucgugaugggcuuguagugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugugu

Figure 1: Example of a CopraRNA input file. Eight homologous GcvB sRNA sequences from different species are displayed. The species are identified by the NCBI RefSeq identifier in the FASTA header.

1.2 Results and Interpretation

Depending on sequence length (target and sRNA), amount of input organisms and genome sizes, CopraRNA can take up to 24h or longer to compute (in most cases it is significantly faster).
1.2.1 What does GcvB regulate?

Many genes encoding mRNAs can be associated with gene ontology terms, which describe biological functions. These terms are grouped into categories (cellular compartment, biological process, molecular function) and represent different levels of detail. Aggregating terms from the predicted putative target mRNAs allows to analyze if specific functionalities are enriched. COPRARNAs’s heatmap for functional annotations summarizes genes and associated terms for interpretation. The result for GcvB shows multiple targets featuring amino acid metabolism related functionality, see Figure 2. This is in good agreement with the functions described in literature [1].

Figure 2: The heatmap summarizes the enriched terms for a specific COPRARNAs prediction list. This instance of the functional annotation chart shows the result for GcvB with a focus on Salmonella (NC_003197). Each line corresponds to a putative target gene and can be identified by the labels on the left. Each column corresponds to an ontology term. These terms are specified in the legend at the bottom. The colors indicate larger groups of related functional terms. Color opacity indicates the CopraRNA p-value for a predicted interaction. The barchart on the bottom right represents the DAVID enrichment scores for the individual functional groups.

1.2.2 Where does the sRNA bind the target mRNAs?

The binding site of an sRNA to its mRNA target gives an indication for the mode of regulation. sRNAs that down-regulate their targets often interact with the ribosome-binding site (RBS) of the mRNA, and thereby block it for translation initiation. Alternatively sRNAs can bind elsewhere and open a secondary structure that otherwise renders the RBS inaccessible. COPRARNAs predicts binding sites on the target mRNAs, which allows to evaluate the effect on single RNAs, as well as the general behavior. The mRNA regions plot for GcvB gives an overview of the binding sites for highly ranked targets, see Figure 3.
Predicted interaction regions mRNAs

Figure 3: COPRarna predicted interaction regions on putative mRNA targets, highlighting a preferred binding region around the start codon. The table below the plot lists the targets and their individual binding regions.

1.2.3 Which sRNA domains are used?

The regions plot clearly outlines that GcvB employs its previously reported regions R1 and R2 for target interactions, see Figure 4.

Summary and Performance

COPRarna allows to find and analyze target RNAs, thereby enabling informed planning of experimental strategies. Experimental testing restricted to high confidence targets has the potential to save time and resources when determining the biological function of sRNAs. The predictions of COPRarna often agree with experimental studies as is also shown here for GcvB. COPRarna’s comparative approach outperforms traditional approaches like IntaRNA, with the requirement that the necessary homologous sequences are available (see Table 1).

<table>
<thead>
<tr>
<th>Target Name</th>
<th>Rank IntaRNA</th>
<th>Rank COPRarna</th>
</tr>
</thead>
<tbody>
<tr>
<td>lrp</td>
<td>97</td>
<td>2 ↑</td>
</tr>
<tr>
<td>livK</td>
<td>47</td>
<td>3 ↑</td>
</tr>
<tr>
<td>cycA</td>
<td>12</td>
<td>4 ↑</td>
</tr>
<tr>
<td>ilvC</td>
<td>3</td>
<td>5 ↓</td>
</tr>
<tr>
<td>livK</td>
<td>94</td>
<td>6 ↑</td>
</tr>
<tr>
<td>cycA</td>
<td>16</td>
<td>8 ↑</td>
</tr>
</tbody>
</table>

Table 1:

Multiple additional examples of experimentally well-characterized RNAs can be found on the input page of the COPRarna webservice.
Figure 4: COPRarna predicted interaction regions of the sRNA, highlighting two preferred binding domains. The table below the plot lists the targets and their individual binding regions on the sRNA.
2 CMV - Visualization for RNA and Protein family models and their comparisons

Homology search is a standard method for the identification of novel RNAs or proteins. Homologous proteins and RNAs that serve the same biological function can be defined as a family. This families can be encoded as covariance models (CMs) for RNAs or Hidden Markov Models (HMMs) for Proteins and are powerful tools for homology search. RNA families that originate from a common ancestor gene, potentially share conserve sequence or structure features and can be grouped into so called RNA family clans. However, their complexity makes it tedious to investigate them in their tabulated default form. The Covariance model visualization tools (CMV) visualize CMs or HMMs and their comparisons. In this application example we will demonstrate how to visualize a RNA family model constructed from an structural alignment done with LocARNA, also a Freiburg-RNA tool.

2.1 Input

We will start from a Stockholm Alignment file generated with LocARNA, which contains the sequence information of several homologous members of the tRNA-family as well as a consensus secondary structure (see Table ) The first step is to copy and paste the following five tRNA FASTA-format sequences from Rfam[3], in the LocARNA webservice and press the submit button.

```
>AB003409.1/96-167
GGGCCCAUAUGCUGUAGUGAGUGCCUCCUUGCAAGGAGGAUGG--UAGAGUGCCUCCUUUGCAAGGAGGAUG---------CCCUGGG-UUCGAAUCCCAGUGGGUCCA

>AB009835.1/1-71
CAUUUAUGACUGGAAAGCAAGUACUGGUCUCUUAAACCAUUUAAUAGUAAAUUAGCACUUACUUCUAAUGA

>AB013372.1/8-81
GCGCCCGUAGUUGAUGAGCGUUUGACUACGGAUCAAAAGGUGUAGGGGUUGCAUCUUCCUCGGCGCGG

>AB013373.1/3764-3825
GCGGAAGUAGUGGAGAAGAACCCUUGCAAGGUGGGGUUGCAUCUUCCUCGGCGCGG

>AB017063.1/58819-58900
GUGGACGUGCUGCAGUUGUUUUCUGCAUGUAGAAACUAUGUGCUGCCUUGCAAGGUGCCGCUUGGAGAAUCCCCUCUGUCACG

# STOCKHOLM 1.0
#=GF CC Generated by LocARNA 1.9.2
#=GF SQ 5

AB003409.1/96-167 GGGCCCAUAUGCUGUAGUGAGUGCCUCCUUGCAAGGAGGAUG---------CCCUGGG-UUCGAAUCCCAGUGGGUCCA
AB013373.1/3764-3825 GCGGAAGUAGUGGAGAAGAACCCUUGCAAGGUGGGGUUGCAUCUUCCUCGGCGCGG
AB017063.1/58819-58900 GUGGACGUGCUGCAGUUGUUUUCUGCAUGUAGAAACUAUGUGCUGCCUUGCAAGGUGCCGCUUGGAGAAUCCCCUCUGUCACG

#GC SS_cons
((((((.(((((((......)))))))))))))))))))))))))
```

On the LocARNA output page download the Stockholm-format alignment. You should obtain a file with following content.

# STOCKHOLM 1.0
#=GF CC Generated by LocARNA 1.9.2
#=GF SQ 5

```
AB003409.1/96-167 GGGCCCAUAUGCUGUAGUGAGUGCCUCCUUGCAAGGAGGAUG---------CCCUGGG-UUCGAAUCCCAGUGGGUCCA
AB013373.1/3764-3825 GCGGAAGUAGUGGAGAAGAACCCUUGCAAGGUGGGGUUGCAUCUUCCUCGGCGCGG
AB017063.1/58819-58900 GUGGACGUGCUGCAGUUGUUUUCUGCAUGUAGAAACUAUGUGCUGCCUUGCAAGGUGCCGCUUGGAGAAUCCCCUCUGUCACG

#GC SS_cons
((((((.(((((((......)))))))))))))))))))))))))
```

The alignment provides all the information to construct an RNA family model with the INFERNAL tool suite [4], specifically the cmbuild tool. You can use to following command to construct a covariance model named tRNA.cm from the Stockholm alignment downloaded:

```
cmbuild tRNA.cm tRNA.stockholm
```

Now navigate to the CMV webservice and select the CMV tool at the navigation bar at the top right. The following Table 2 gives an overview over the available options. Press the button on the topmost upload field and select the tRNA.cm covariance model. Then select the tRNA Stockholm alignment obtained from the LocARNA webservice for the upload field below. Uploading the alignment is optional, but will provide you with a additional visualization for the alignment. Three levels of increasing detail for model visualization can be set via the dropdown menu below. The emission layout option controls how emission probabilities are displayed, while the model layout allows to arrange to model visualization layout according to the guide tree. Transition probabilities are only displayed if they are higher than the set cutoff, excluding low probabilities. For submission press the 'Send it' button.
2.2 Results and Summary

Minimal detail setting displays per node, which corresponds to a single nucleotide or two paired nucleotides, one box labeled with the node index (see Figure 5 A).

Setting the detail level to simple additionally displays the node type for each state, see Figure 5 B. The detailed setting also visualizes states and transition probabilities, see Figure 6.

The alignment visualization with column indices can be seen in Figure 5 C, moreover the consensus secondary structure is used to create secondary structure visualization via R2R [5] (see Figure 5 D) and forna [6] (not shown).

CMV can also be used to visualize Hidden Markov Models, and comparisons between Hidden Markov Models and Covariance Models, for more Information and examples please refer to the Help page of the webservice or the supplementary material of the CMV publication [7].
3 Teaching - Available Algorithms

In the following, we list the algorithms currently covered by the Teaching section of the Freiburg RNA tools webserver. All implementations are also freely available on github at

https://github.com/BackofenLab/RNA-Playground

for extension, download and offline-usage.

3.1 RNA structure prediction
- counting of all foldable nested structures [8, 9]
- optimal structure prediction (base pair maximization) [10, 11] including suboptimal-structure enumeration [12]
- computation of partition functions, base pair and unpaired probabilities [13, 14]
- maximum expected accuracy (MEA) structure prediction [15]

3.2 RNA-RNA interaction prediction
- hybrid-only interaction prediction [16]
- co-folding / concatenation-based interaction prediction [17, 18]
- accessibility-based interaction prediction [19, 20, 21]

3.3 Pairwise sequence alignment
- global [22] and local [23] alignment with linear gap cost
- global alignment in linear space [24] with linear gap cost
- global (and local) alignment with affine gap cost [25]
- global alignment with arbitrary gap cost [26]
- length-normalized local alignment [27] with linear gap cost

3.4 Multiple sequence alignment
- progressive alignment [28, 29]
- iterative refinement of progressive alignment
- t-coffee approach [30]

3.5 Evolutionary tree construction
- WPGMA - Weighted Pair Group Method with Arithmetic Mean (Simple Average) [31]
- UPGMA - Unweighted PGMA (Group Average or Average Linkage)
- Single Linkage (Nearest Neighbour on elements between two different clusters)
- Neighbour-Joining (Nearest Neighbour on clusters)
- Complete Linkage (Furthest Neighbour)
References


Figure 5: A) Minimal tRNA family visualization, using the example input, showing model nodes with their index arranged according to guide tree. B) Simple tRNA family visualization, showing model nodes, labeled with node type and index, arranged according to guide tree. C) tRNA family alignment visualization, using the example above, the number of alignment entries can be set via option. The displayed column indices allow association with the corresponding nodes in the model. The consensus secondary structure is shown in the last line. D) Consensus secondary structure visualization for tRNA with R2R.
Figure 6: Detailed tRNA family visualization, showing nodes with states, emission and transition probabilities, arranged according to guide tree. Transitions below threshold that can be set via option are not shown and line strength corresponds to probability.