

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

Martin Raden<sup>1</sup>, Syed M. Ali<sup>1</sup>, Omer S. Alkhnbashi<sup>1</sup>, Anke Busch<sup>2</sup>, Fabrizio Costa<sup>3</sup>,  
Jason A. Davis<sup>4</sup>, Florian Eggenhofer<sup>1</sup>, Rick Gelhausen<sup>1</sup>, Jens Georg<sup>5</sup>, Steffen Heyne<sup>6</sup>,  
Michael Hiller<sup>7,8</sup>, Kousik Kundu<sup>9,10</sup>, Robert Kleinkauf<sup>1</sup>, Steffen C. Lott<sup>5</sup>,  
Mostafa M. Mohamed<sup>1</sup>, Alexander Mattheis<sup>1</sup>, Milad Miladi<sup>1</sup>, Andreas S. Richter<sup>11</sup>,  
Sebastian Will<sup>12</sup>, Joachim Wolff<sup>1</sup>, Patrick R. Wright<sup>13</sup>, Rolf Backofen<sup>1,14</sup>

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

## Contents

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>CopraRNA - Application example</b>  | <b>2</b> |
| 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        |
| <b>2</b> | <b>CMV - Visualization for RNA and Protein family models and their comparisons</b> | <b>6</b> |
| 2.1      | Input . . . . .  | 6        |
| 2.2      | Results and Summary . . . . .  | 7        |
| <b>3</b> | <b>Teaching - Available Algorithms</b>   | <b>8</b> |
| 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. COPRARNA 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
acuuccugagccggaacgaaaaguuuauucggaugcguguucugaugggcuuuuggcuuacgguugugauguuguguu
guuguguuugcaauuggucugcgauucagaccacgguagcgagacuacccuuuuucacuuccuguacauuuaccuguc
uguccauagugauuaauguagcaccgccaauugcgguuguuu
>NC_000913
acuuccugagccggaacgaaaaguuuauucggaugcguguucuggugaacuuuuggcuuacgguugugauguuguguu
guuguguuugcaauuggucugcgauucagaccaugguagcaaagcuacccuuuuucacuuccuguacauuuaccuguc
uguccauagugauuaauguagcaccgccaauugcgguuguuu
>NC_009792
acuuccugagccggaacgaaaaguuuauucggaugcguguucugaugggcuuuuggcuuacgguugugauguuguguu
guuguguuugcaauuggucugcgauucagaccaugguagcgaagcuacccuuuuucacuuccuguacauuuaccuguc
uguccauagugauuaauguagcaccgccaauugcgguuguuu
>NC_013716
acuuccugagccggaacgaaaaguuuauucggaugcguguucuggggcuuuuggcuuacgguugugauguuguguu
guuguguuugcaauuggucugcgauucagaccacgguagcgaagcuacccuuuuucacuuccuguacauuuaccuguc
uguccauagugauuaauguagcaccgccaauugcgguuguuu
>NC_011740
acuuccugagccggaacgaaaaguuuauucggaugcguguucuggugaacuuuuggcuuacgguugugauguuguguu
guuguguuugcaauuggucugcgauucagaccacgguagcgaagcuacccuuuuucacuuccuguacauuuaccuguc
uguccauagugauuaauguagcaccgccaauugcgguuguuu
>NC_012917
acuuccugggcccgaacgaaaagugcggaugggugaccugaggugcuuuuggcuugugguugugauguuguguuugcua
uuuguuugucugccuuuugcagaugguagcgagucuacccuauucacuuccuguacauuuaccugucuguccaaag
ugauuuuuugugagcaccgcaauugcgguuguuu
>NC_009832
acuuccggggccggaacgaaaaggagugggugucgaggacgccgaugaacuuuuggcuugugguugugauguuguguu
ugcaauuugucuggcaauccagacgugguagcuaagcuacuguuuuuucacuuccuguacauuuaccugucuguccaa
gugauuuuugcagcaccgcaauucguugcgguuguuu
>NC_003143
acuucccuagccggaacgaaaaguagguuugguauccagguacugaaaugcuuuuggcuuugugguggugauguugu
guuugcaauugucugguauugcagacauaguagcuaagcuacugcuuuuuuucacuuccuguacauuuaccugucug
uccauagugauuuuugcagcaccguuuuuugcgguuguuu
```

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. COPRARNA'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].

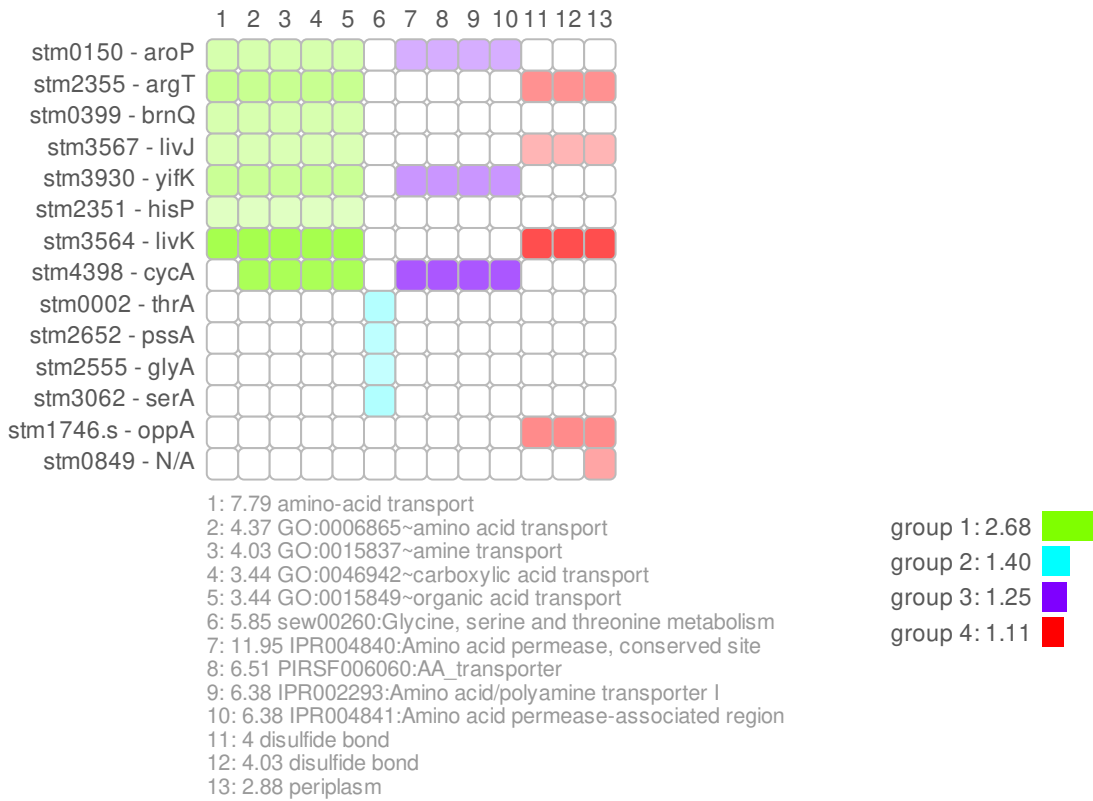


Figure 2: The heatmap summarizes the enriched terms for a specific COPRARNA prediction list. This instance of the functional annotation chart shows the result for GcvB with a focus on *Salmonella* (NC\_003197). Each line corresponds to a putative target gene and can be identified by the labels on the 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. COPRARNA 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.

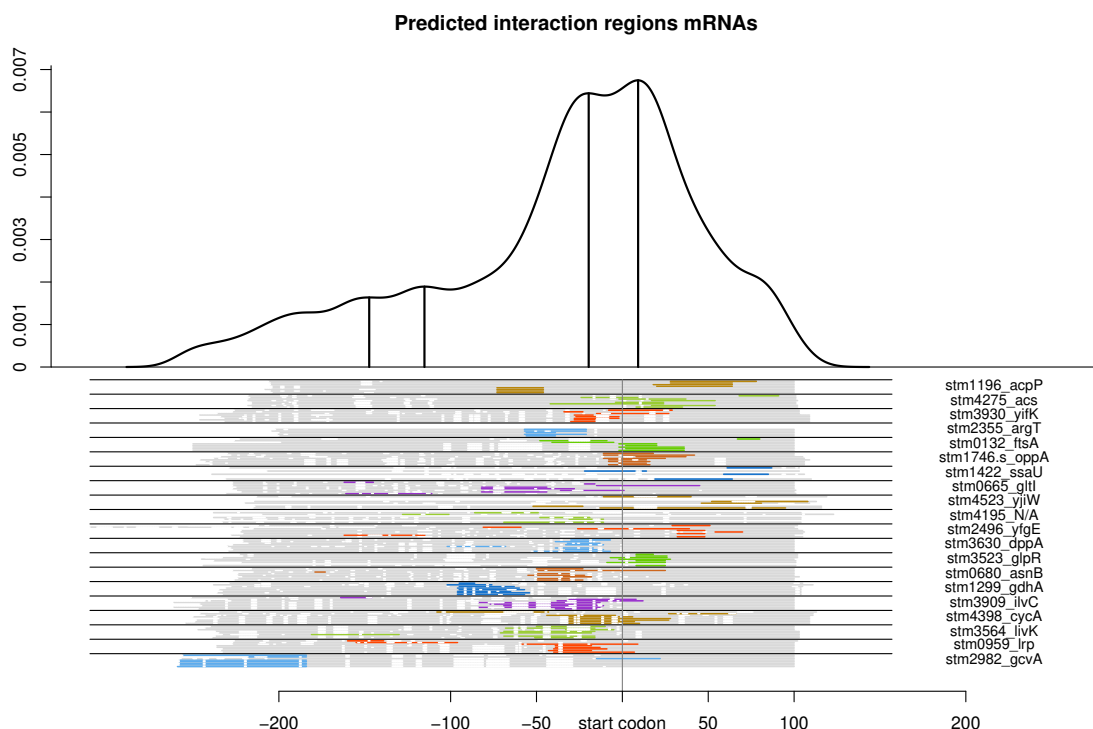


Figure 3: COPRARNA predicted interaction regions on putative mRNA targets, highlighting a preferred binding region around the start codon. The table below the plot lists the targets and their individual binding regions.

### 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).

| Target Name | Rank IntaRNA | Rank CopraRNA |
|-------------|--------------|---------------|
| lrp         | 97           | 2 ↑           |
| livK        | 47           | 3 ↑           |
| cycA        | 12           | 4 ↑           |
| ilvC        | 3            | 5 ↓           |
| livK        | 94           | 6 ↑           |
| cycA        | 16           | 8 ↑           |

Table 1:

Multiple additional examples of experimentally well-characterized RNAs can be found on the input page of the COPRARNA webservice.

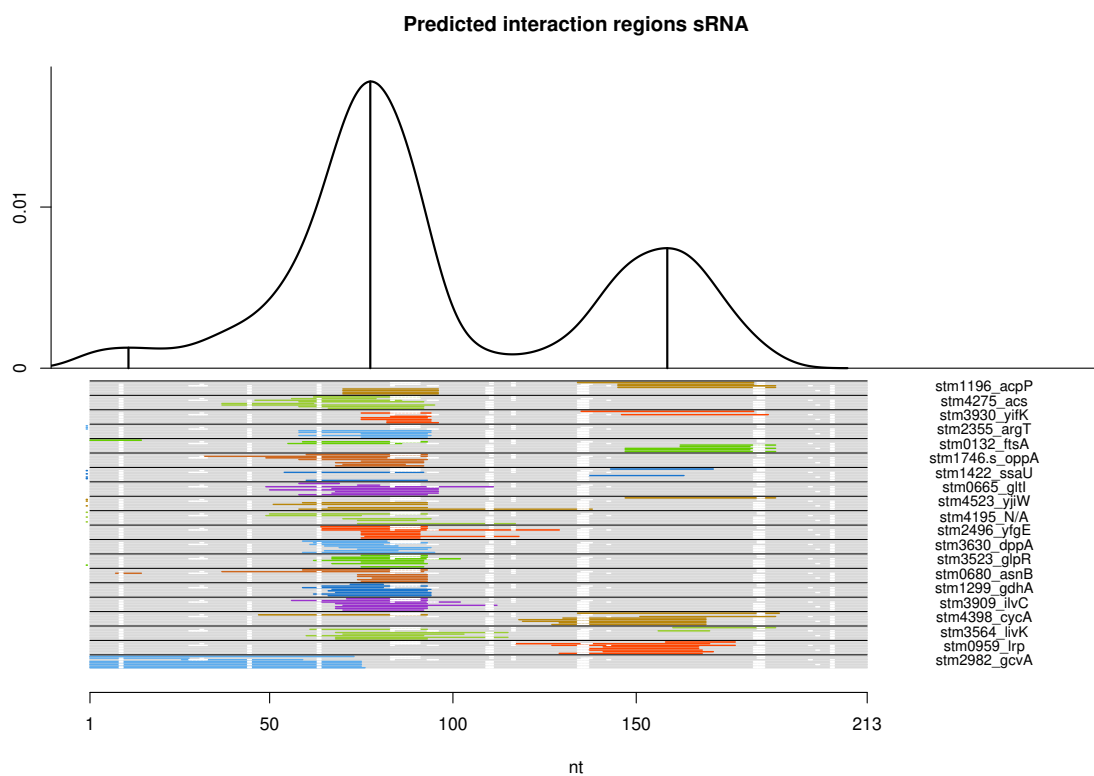


Figure 4: COPRARNNA predicted interaction regions of the sRNA, highlighting two preferred binding domains. The table below the plot lists the targets and their individual binding regions on the sRNA.

## 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 conserved 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
GGGCCCAUAGCUCAGUGGUAAGAGUGCCUCCUUUGCAAGGAGGAUGCCCUGGGUUCGAAUCCAGUGGGUCCA
>AB009835.1/1-71
CAUUAAGAUGACUGAAAGCAAGUACUGGUCUCUUAACCAUUAUUAAGUAAAUAAGCACUUAUCUUAUAUGA
>AB013372.1/8-81
GCGCCCGUAGCUCAAUUGGAUAGAGCGUUUGACUACGGAUCAAAGGUUAGGGGUUCGACUCCUCUGGGCGCG
>AB013373.1/3754-3825
GCGGAAGUAGUUCAGUGGUAGAACACCACCUUGCCAAGGUGGGGUGCGGGUUCGAAUCCCGUCUCCGCU
>AB017063.1/58819-58900
GUGGACGUGCCGAGUGGUUAUCGGGCAUGACUAGAAAUAUGUGGGCUUUGCCCGCGCAGGUUCGAAUCCUGCCGUUCACG
```

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      GGGCCCAUAGCUCAGUGG--UAGAGUGCCUCCUUUGCAAGGAGGAUG-----CCCUGGG-UUCGAAUCCAGUGGGUCCA
AB013373.1/3754-3825  GCGGAAGUAGUUCAGUGG--UAGAACACCACCUUGCCAAGGUGGGGG-----UCGCGGG-UUCGAAUCCCGUCUCCGCU
AB013372.1/8-81       GCGCCCGUAGCUCAAUUGGAUAGAGCGUUUGACUACGGAUCAAAGG-----UUAGGGG-UUCGACUCCUCUGGGCGCG
AB017063.1/58819-58900 GUGGACGUGCCGAGUGGU-UUACGGGCAUGACUAGAAAUAUGUGGGCUUUGCCCGCGCAGG-UUCGAAUCCUGCCGUUCACG
AB009835.1/1-71       CAUUAAGAUGACUGAAAGC----AAGUACUGGUCUCUUAACCAUUA-----AUAGUAAAUAAGCACUUAUCUUAUGA
#=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.

| Name                             | Description   |
|----------------------------------|---|
| Detail level                     | minimal - showing the node number, simple - + node type and emission probabilities, detailed - + states, transition probabilities |
| Emission layout                  | box - fill state of a box, score - bit score, as a floating pint number, probability - as a floating point number                 |
| Output format                    | Available output formats are pdf, png, svg and ps. The webservice always generates svg by default                                 |
| Max. number of alignment entries | This controls how many entries are displayed for optionally uploaded alignments.  |
| Image size scaling factor        | Scales the result image by set factor. Please consider that resulting .svg output can be easily rescaled.                         |
| Transition probability cutoff    | Cutoff for a transition probability to be displayed.  |

Table 2:

## 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

- [1] Cynthia M. Sharma, Kai Papenfort, Sandy R. Pernitzsch, Hans-Joachim Mollenkopf, Jay C. D. Hinton, and Jorg Vogel. Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol Microbiol*, 81(5):1144–65, 2011. [PubMed:21696468] [doi:10.1111/j.1365-2958.2011.07751.x].
- [2] Xiaoli Jiao, Brad T. Sherman, Da Wei Huang, Robert Stephens, Michael W. Baseler, H. Clifford Lane, and Richard A. Lempicki. David-ws: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics*, 28(13):1805–1806, 2012. [PubMed:22543366] [PubMed Central:PMC3381967] [doi:10.1093/bioinformatics/bts251].
- [3] Ioanna Kalvari, Joanna Argasinska, Natalia Quinones-Olvera, Eric P. Nawrocki, Elena Rivas, Sean R. Eddy, Alex Bateman, Robert D. Finn, and Anton I. Petrov. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Research*, 46(D1):D335–D342, January 2018. [PubMed:29112718] [PubMed Central:PMC5753348] [doi:10.1093/nar/gkx1038].
- [4] Eric P. Nawrocki and Sean R. Eddy. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, 29(22):2933–2935, November 2013. [PubMed:24008419] [PubMed Central:PMC3810854] [doi:10.1093/bioinformatics/btt509].
- [5] Zasha Weinberg and Ronald R. Breaker. R2R - software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics*, 12(1):3, 2011. [PubMed:21205310] [PubMed Central:PMC3023696] [doi:10.1186/1471-2105-12-3].
- [6] Peter Kerpedjiev, Stefan Hammer, and Ivo L. Hofacker. Forna (force-directed RNA): simple and effective online RNA secondary structure diagrams. *Bioinformatics*, page btv372, 2015. [PubMed:26099263] [PubMed Central:PMC4595900] [doi:10.1093/bioinformatics/btv372].
- [7] Rolf Backofen Florian Eggenhofer, Ivo L. Hofacker and Christian Höner zu Siederdisen. Cmv - visualization for rna and protein family models and their comparisons. *Bioinformatics*, 2018. in revision final stage.
- [8] M.S. Waterman and T.F. Smith. RNA secondary structure: a complete mathematical analysis. *Mathematical Biosciences*, 42(3):257 – 266, 1978.
- [9] Michael Waterman. Secondary structure of single-stranded nucleic acids. In *Studies on foundations and combinatorics, Advances in mathematics supplementary studies*, Academic Press N.Y., 1:167 – 212, pages 167–212, 1978. [PubMed:21626368] [doi:10.1140/epje/i2011-11055-2].
- [10] Ruth Nussinov, George Pieczenik, Jerrold R. Griggs, and Daniel J. Kleitman. Algorithms for loop matchings. *SIAM J Appl Math*, 35(1):68–82, 1978. [doi:10.1137/0135006].
- [11] Richard Durbin, Sean R. Eddy, Anders Krogh, and Graeme Mitchison. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, 1998.
- [12] S. Wuchty, W. Fontana, I. L. Hofacker, and P. Schuster. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, 49(2):145–65, 1999. [PubMed:10070264] [doi:10.1002/(SICI)1097-0282(199902)49:2<145::AID-BIP4>3.0.CO;2-G].
- [13] J. S. McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29(6-7):1105–19, 1990. [PubMed:1695107] [doi:10.1002/bip.360290621].
- [14] Stephan H. Bernhart, Ulrike Mückstein, and Ivo L. Hofacker. Rna accessibility in cubic time. *Algorithms for Molecular Biology*, 6(1):3, 2011. [PubMed:21388531] [PubMed Central:PMC3063221] [doi:10.1186/1748-7188-6-3].
- [15] Chuong B. Do, Daniel A. Woods, and Serafim Batzoglou. CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics*, 22(14):e90, 2006. [PubMed:16873527] [doi:10.1093/bioinformatics/btl246].

- [16] Brian Tjaden, Sarah S. Goodwin, Jason A. Opdyke, Maude Guillier, Daniel X. Fu, Susan Gottesman, and Gisela Storz. Target prediction for small, noncoding RNAs in bacteria. *Nucleic Acids Research*, 34(9):2791, 2006. [PubMed:[16717284](#)] [PubMed Central:[PMC1464411](#)] [doi:[10.1093/nar/gkl356](#)].
- [17] Ivo L. Hofacker, Walter Fontana, Peter F. Stadler, Sebastian Bonhoeffer, Manfred Tacker, and Peter Schuster. Fast folding and comparison of RNA secondary structures. *Monatshefte Chemie*, 125:167–188, 1994. [doi:[10.1007/BF00818163](#)].
- [18] David H. Mathews, Mark E. Burkard, Susan M. Freier, Jacqueline R. Wyatt, and Douglas H. Turner. Predicting oligonucleotide affinity to nucleic acid targets. *RNA*, 5(11):14581469, 1999. [PubMed:[10580474](#)] [PubMed Central:[PMC1369867](#)].
- [19] Ulrike Mückstein, Hakim Tafer, Jörg Hackermüller, Stephan H. Bernhart, Peter F. Stadler, and Ivo L. Hofacker. Thermodynamics of RNARNA binding. *Bioinformatics*, 22(10):1177, 2006. [doi:[10.1093/bioinformatics/btl024](#)].
- [20] Martin Mann, Patrick R. Wright, and Rolf Backofen. IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. *Nucl Acid Res*, 45(W1):W435–W439, 2017. [PubMed:[28472523](#)] [PubMed Central:[PMC5570192](#)] [doi:[10.1093/nar/gkx279](#)].
- [21] Anke Busch, Andreas S. Richter, and Rolf Backofen. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*, 24(24):2849–56, 2008. [PubMed:[18940824](#)] [PubMed Central:[PMC2639303](#)] [doi:[10.1093/bioinformatics/btn544](#)].
- [22] Saul B. Needleman and Christian D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3):443 – 453, 1970. [PubMed:[5420325](#)].
- [23] T.F. Smith and M.S. Waterman. Identification of common molecular subsequences. *Journal of Molecular Biology*, 147(1):195 – 197, 1981. [PubMed:[7265238](#)].
- [24] D. S. Hirschberg. A linear space algorithm for computing maximal common subsequences. *Commun. ACM*, 18(6):341–343, June 1975. [doi:[10.1145/360825.360861](#)].
- [25] Osamu Gotoh. An improved algorithm for matching biological sequences. *Journal of Molecular Biology*, 162(3):705 – 708, 1982. [PubMed:[7166760](#)].
- [26] M.S Waterman, T.F Smith, and W.A Beyer. Some biological sequence metrics. *Advances in Mathematics*, 20(3):367 – 387, 1976. [doi:[10.1016/0001-8708\(76\)90202-4](#)].
- [27] Abdullah N. Arslan, Ömer Eeciolu, and Pavel A. Pevzner. A new approach to sequence comparison: normalized sequence alignment. *Bioinformatics*, 17(4):327–337, 2001. [PubMed:[11301301](#)].
- [28] Da-Fei Feng and Russell F. Doolittle. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *Journal of Molecular Evolution*, 25(4):351–360, Aug 1987. [doi:[10.1007/BF02603120](#)].
- [29] Da-Fei Feng and Russell F. Doolittle. [21] progressive alignment of amino acid sequences and construction of phylogenetic trees from them. In *Computer Methods for Macromolecular Sequence Analysis*, volume 266 of *Methods in Enzymology*, pages 368 – 382. Academic Press, 1996. [doi:[10.1016/S0076-6879\(96\)66023-6](#)].
- [30] Cdric Notredame, Desmond G Higgins, and Jaap Heringa. T-coffee: a novel method for fast and accurate multiple sequence alignment edited by j. thornton. *Journal of Molecular Biology*, 302(1):205 – 217, 2000. [doi:[10.1006/jmbi.2000.4042](#)].
- [31] Louis L. McQuitty. Single and multiple hierarchical classification by reciprocal pairs and rank order types. *Educational and Psychological Measurement*, 26(2):253–265, 1966. [doi:[10.1177/001316446602600201](#)].



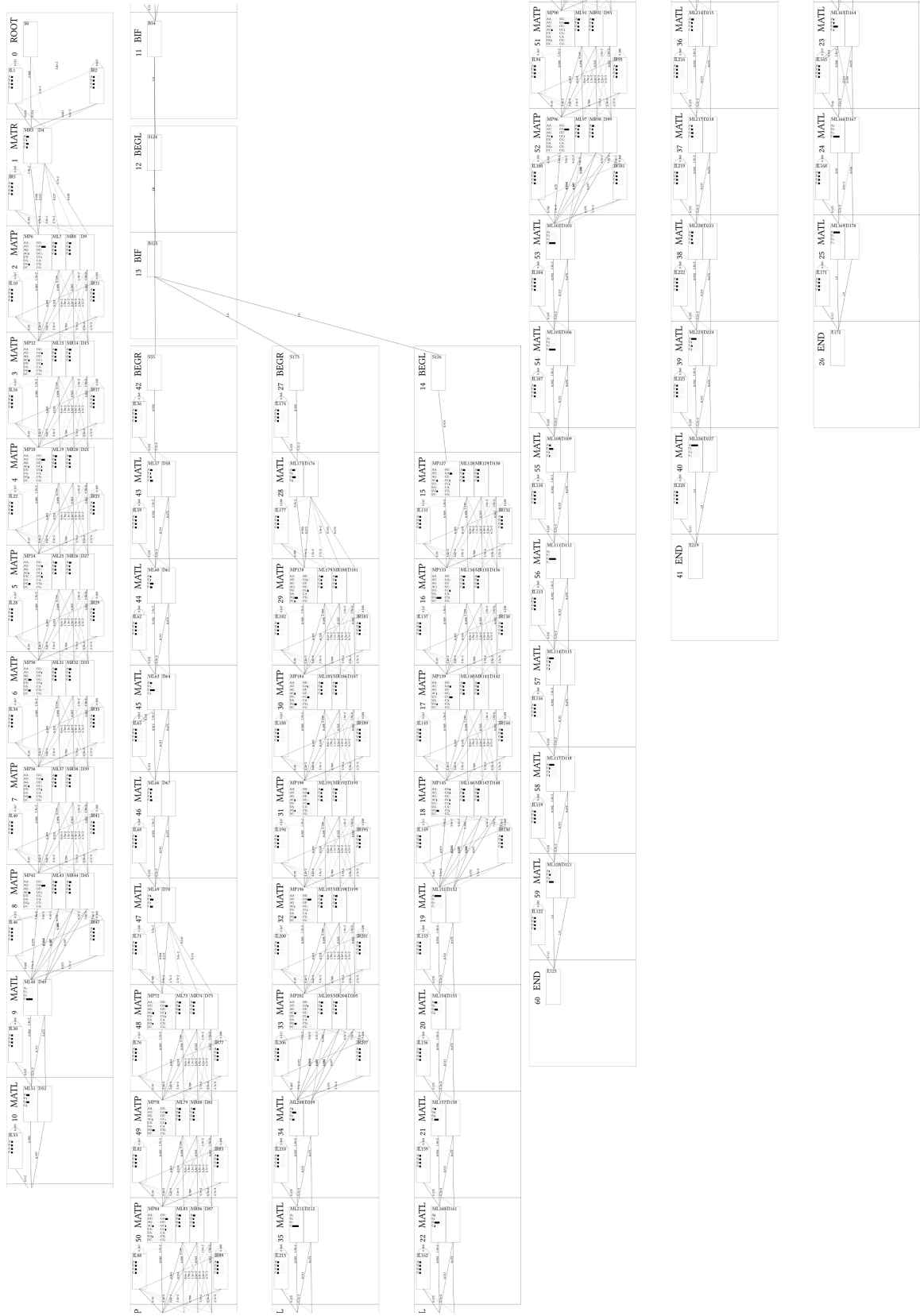


Figure 6: Detailed tRNA family visualization, showing nodes with states, emission and transition probabilities, arranged according to guide tree. Transitions below threshold that can be set via option are not shown and line strength corresponds to probability