

Help:

cmv offers tools for the visualisation of RNA family models, also known as covariance models (CM) and Hidden Markov Models (HMM). Moreover, comparsions between models, the multiple sequence alignments they were constructed from and, in the case of RNA families, the consensus secondary structure can be visualised. The first part of the guide explains the application of the tools for hidden markov models (*hmmv*, *hmmcvc*) on the command line, as well as on the web service. The second part addresses the usage of the tools for covariance models (*cmv*, *cmcvc*). Each of the four included webservices features a sample button at the bottom of the page, complete with example input files. If you look for source code or a installation guide for the tool please refer to the Tool subpage. This manual is included with the tool as manual.pdf. (<https://github.com/eggzilla/CMV/raw/master/manual.pdf>)

Table of contents:

- HMM-Tools
 - Input
 - Commandline usage
 - Output
 - Pfam-Examples
- CM-Tools
 - Input
 - Commandline usage
 - Output
 - Rfam-Examples

HMM-Tools:

Hidden Markov models are used to represent the sequence information of biopolymers. Nodes of the model represent columns of a multiple sequence alignment. The guide describes the required input and parameters for both tools als well as the output. A visualisation for the EGF-protein family and a comparison visualisation for the hammerhead-RNA clan with corresponding command line calls are used as examples.

Input

HMM (-m)

Input models are supported in HMMER3 (see HMMER User-guide (<http://hmmer.org/documentation.html>)) format, as used by Pfam (<http://pfam.xfam.org/>) and as part of Rfam (<http://rfam.xfam.org/>) INFERNAL (<http://eddylab.org/infernal/>) models. Here is the EGF (<http://192.52.2.124/data/examples/EGF/PF00008.hmm>) protein family as example. Multiple input models can be provided by concatenating them in one file, see hammerhead-RNA (<http://192.52.2.124/data/examples/hammerheadClan/hammerheadClan.hmm>) family clan. The webservice accepts a file upload, the commandline tool a absolute filepath.

Alignment (-s)

Optionally the multiple sequence alignment used to construct the input model can be provided in Stockholm format, e.g. EGF (<http://192.52.2.124/data/examples/EGF/PF00008.hmm>) alignment. For multiple input models the same number of alignments must be provided in the same sequence as the models, also concatenated into one file, e.g. hammerhead-RNA (<http://192.52.2.124/data/examples/hammerheadClan/aln/hammerheadClan.aln>). The webservice accepts a file upload, the commandline tool a absolute filepath.

Comparison (-r)

HMMCVC requires a comparison file, detailing the relationship between the input models. The webservice accepts a file upload, the commandline tool a absolute filepath. This format is derived from the CMCompare output format and contains on each line following white space separated fields:

```
model1Name model2Name linkscore1 linkscore2 linksequence model1matchednodes model2matchednodes
Here is a example line from a hammerhead clan comparison, the whole file can be found here: hammerheadClan-comparison
(http://192.52.2.124/data/examples/hammerheadClan/hammerheadClan.hmmcompare).
Hammerhead_1 Hammerhead_3 6.168 5.244 GUCCAGUAUAGGAC [17,18,19,20,21,22,23,23,24,25,26,27,28]
```

[36,37,38,39,40,41,42,43,44,45,46,47,48]

Detail level (-d)

Three detail level for each node are available:

- minimal - showing the node number.
- simple - showing emission probabilities
- detailed - showing emission and transition probabilities

Emission layout (-e)

Controls display of emission probabilities for detail levels simple and detailed. The selected variant is shown next to the emitted symbol.

- box - fill state of a box
- score - bit score, as a floating point number
- probability - as a floating point number

Output format (-f)

Available output formats are pdf, png, svg and ps. The webservice always generates svg by default for rendering of the preview.

Max. number of alignment entries (-n)

This controls how many entries are displayed for optionally uploaded alignments.

Image size scaling factor (-c)

Scales the result image by set factor. Please consider that resulting .svg output can be easily rescaled.

Transition probability cutoff (-t)

Minimum necessary cutoff for a transition probability to be displayed.

Output directory path (only cmdline, -o)

Absolute path to output directory

Help (only cmdline, --help)

Prints help with all default options and commandline parameters

Commandline usage

hmmv-visualisation for the Piwi protein family, as used for the webservice, can be obtained with the following commandline call:

hmmv -d detailed -m Piwi.hmm -s /home/user/PF02171_seed.txt

hmmc comparison visualisation for the hammerheadRNA clan can be computed like this:

hmmc -d detailed -m /home/user/hammerhead.hmm -s /home/user/hammerhead.stockholm.txt -r /home/user/hammerhead.hmmc -f pdf

Output

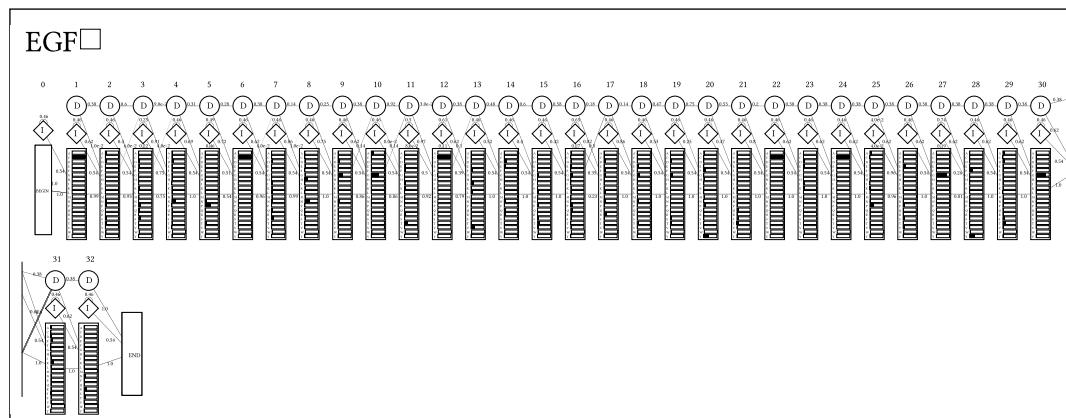
HMMV-Output

For each input model a outputfile in the requested format is generated. Als filename the modelname encoded in the file is used and as file extension the request output format (.png,.svg,.ps,.pdf). If stockholm alignments have been provided, then for each model a alignment visualisation with index columns is created. The alignment file name is the model name, followed by ".aln" and then the requested file format extension. The webservice gzips all results and provides a download link. Following is a example result table for the EGF-protein family from Pfam (PF00008 (<http://pfam.xfam.org/family/PF00008>)).

Model name	Model-pdf (http://192.52.2.124/data/examples/EGF/EGF.pdf)	Model-svg (http://192.52.2.124/data/examples/EGF/EGF.svg)	Alignment-pdf (http://192.52.2.124/data/examples/EGF/EGF.aln.pdf)
EGF	pdf	svg	pdf

The webservice also creates a zoom- and panable preview, that can be expanded by clicking.

EGF



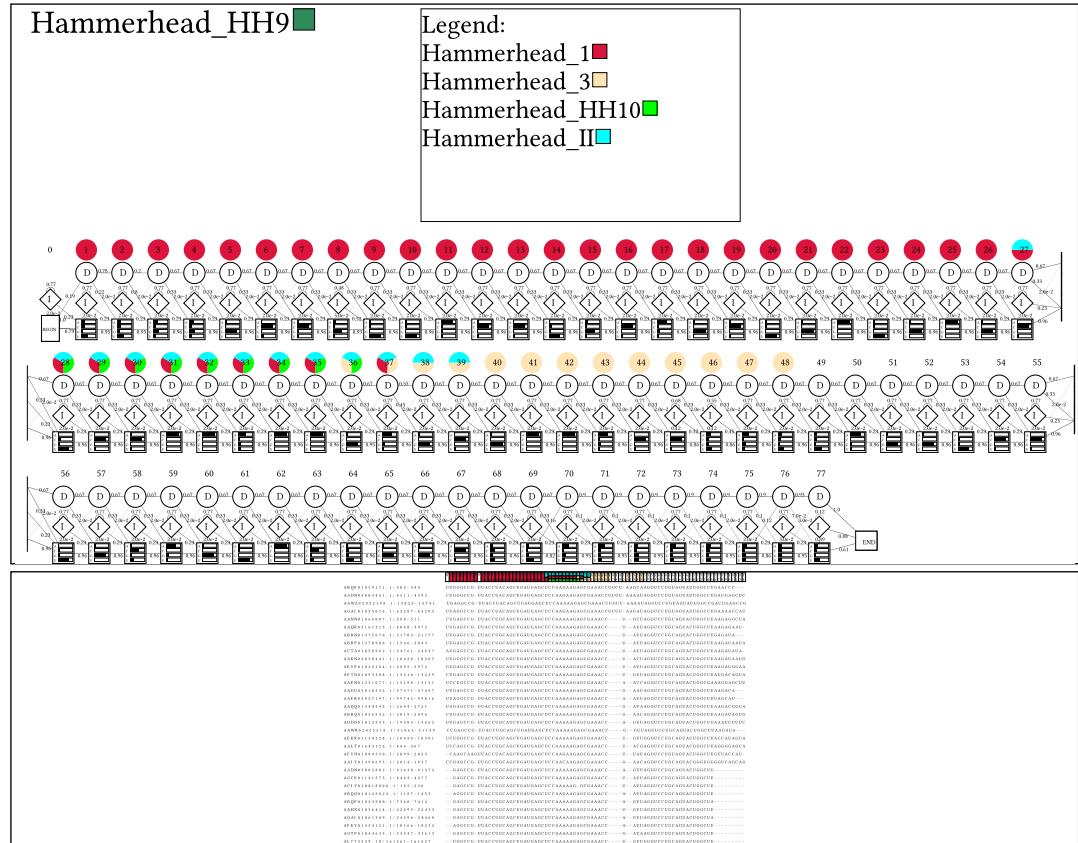
hmmcv-output

hmmcv also provides visualisation output for each model and corresponding alignment. The naming scheme is similar to *hmmv*. However it annotates regions found to be linked by model comparison via color tags. Each provided model is associated with a color that is depicted in the legend table. The nodes of each model visualisation are then tagged with the color of the linked model. The coloring is also applied to the runing index of the alignment visualisation. This allows to study in detail which parts of models are connected, or shared in e.g. clans and which not. The following example shows only the result for the Hammerhead_HH9 RNA.

Model name	Model-pdf	Model-svg
Hammerhead_HH9	pdf (http://192.52.2.124/data/examples/hammerheadClan/hmmcv/Hammerhead_HH9.pdf)	svg (http://192.52.2.124/data/examples/hammerheadClan/h)

The webservice also creates a zoom- and panable preview, that can be expanded by clicking

Hammerhead_HH9_HMM



Examples

Precomputed example table (<http://192.52.2.124/data/HMMs>) for minimal.simple and detailed visualisation for 1500 Pfam models.

CM-Tools:

Covariance, also called RNA family models are used to represent the sequence and structure information of RNA molecules. Nodes of the model represent

columns of a multiple sequence alignment. Parameters that only work for the command line tool and not on the webserver are annotated with "cmdline-only". A visualisation for the tRNA family and a comparison visualisation for the hammerhead-RNA clan with corresponding command line calls are used as examples.

Input

CM (-m)

Input models are supported in INFERNAL (see Infernal User-guide (<http://eddylab.org/infernal/>)) format, as used by Rfam (<http://rfam.xfam.org/>) or constructed by RNAlien (<http://rna.tbi.univie.ac.at/rnalien/>). Here is the tRNA (<http://192.52.2.124/data/examples/tRNAClan/RF00005.cm>) family as example. Multiple input models can be provided by concatenating them in one file, see hammerhead-RNA (<http://192.52.2.124/data/examples/hammerheadClan/hammerheadClan.cm>) family clan. The webservice accepts a file upload, the commandline tool a absolute filepath.

Alignment (-s)

Optionally the multiple sequence alignment used to construct the input model can be provided in Stockholm format, e.g. tRNA (<http://192.52.2.124/data/examples/tRNAClan/RF00005.stockholm.txt>) alignment. For multiple input models the same number of alignments must be provided in the same sequence as the models, also concatenated into one file. The webservice accepts a file upload, the commandline tool a absolute filepath.

Comparison (-r)

cmcv requires a comparison file, detailing the relationship between the input models. The webservice accepts a file upload, the commandline tool a absolute filepath. The used CMCompare output format contains on each line following white space separated fields:

```
model1Name model2Name linkscore1 linkscore2 linksequence model1matchednodes model2matchednodes
Here is a example line from a hammerhead clan comparison, the whole file can be found here: hammerheadClan-comparsion
```

(<http://192.52.2.124/data/examples/hammerheadClan/hammerheadClan.hmmcompare>).
Hammerhead_1 Hammerhead_3 6.168 5.244 GUCCAGUAAUAGGAC (((.....))) (((.....))) [17,18,19,20,21,22,23,23,24,25,26,27,28]
[36,37,38,39,40,41,42,43,44,45,46,47,48]

It is possible to use output files from the CMCompare Webserver (<http://rna.tbi.univie.ac.at/cgi-bin/cmcws/cmcws.cgi>), only the filepaths in the first two fields must be replaced with model names. CMV offers a commandlinetool (cmcwshtocmcv) to perform this task automatically, taking the concatenated model file and the corresponding comparsion file as input.

```
cmcwshtocmcv -m /home/user/models.cm -r /home/user/model.cmcwsresult
```

Detail level (-d)

Three detail level for each node are available:

- minimal - showing the node number.
- simple - showing node number and node type
- detailed - showing node number, type, emission and transition probabilities

Emission layout (-e)

Controls display of emission probabilities for detail levels simple and detailed. The selected variant is shown next to the emitted symbol.

- box - fill state of a box
- score - bit score, as a floating point number
- probability - as a floating point number

Model layout (-l)

- tree - model is branched at each bifurcation according to guide tree
- flat - nodes are all arranged on a line

Secondary structure visualisation tool (only cmdline, -x)

cmv and cmcv can create input files for the secondary structure visualisation tools forna and R2R. This input file contains the consensus secondary structure of the input alignment. Moreover cmcv labels nucleotides that are annotated to be similar with compared models in the input files and in case of R2R color encodes them according to the color scheme used for the model comparison visualisation. In case of the webservice both visualisations are triggered on providing a input alignment.

- forna - for each model a input structure file and a color scheme is generated for upload to forna
- r2r - for each model a input file for r2r is generated

Output format (-f)

Available output formats are pdf, png, svg and ps. The webservice always generates svg by default for rendering of the preview.

Max. number of alignment entries (-n)

This controls how many entries are displayed for optionally uploaded alignments.

Image size scaling factor (-c)

Scales the result image by set factor. Please consider that resulting .svg output can be easily rescaled.

Transition probability cutoff (-t)

Minimum necessary cutoff for a transition probability to be displayed.

Output directory path (only cmdline, -o)

Absolute path to output directory

Help (only cmdline, --help)

Prints help with all default options and commandline parameters

Commandline usage

cmv-visualisation for the tRNA family, as used as sample for the webservice, can be obtained with the following commandline call:

```
cmv -m tRNA.cm -s RF00005.stockholm.txt -f pdf
```

cmcv comparsion visualisation for the hammerheadRNA clan can be computed like this:

```
cmcv -d simple -m /home/user/hammerheadClan.cm -s /home/user/hammerheadClan.sto -r /home/user/hammerheadClan.cmcompare -f pdf -l tree
```

Output

cmv-output

For each input model a outputfile in the requested format is generated. Als filename the modelname encoded in the file is used and as file extension the request output format (.png,.svg,.ps.,.pdf). If stockholm alignments have been provided, then for each model a alignment visualisation with index colums is created. The alignment file name is the model name, followed by ".aln" and then the requested file format extension. The webservice gzips all results and provides a download link. Following is a example result table for the tRNA family from Rfam (RF00005 (<http://rfam.xfam.org/family/RF00005>)).

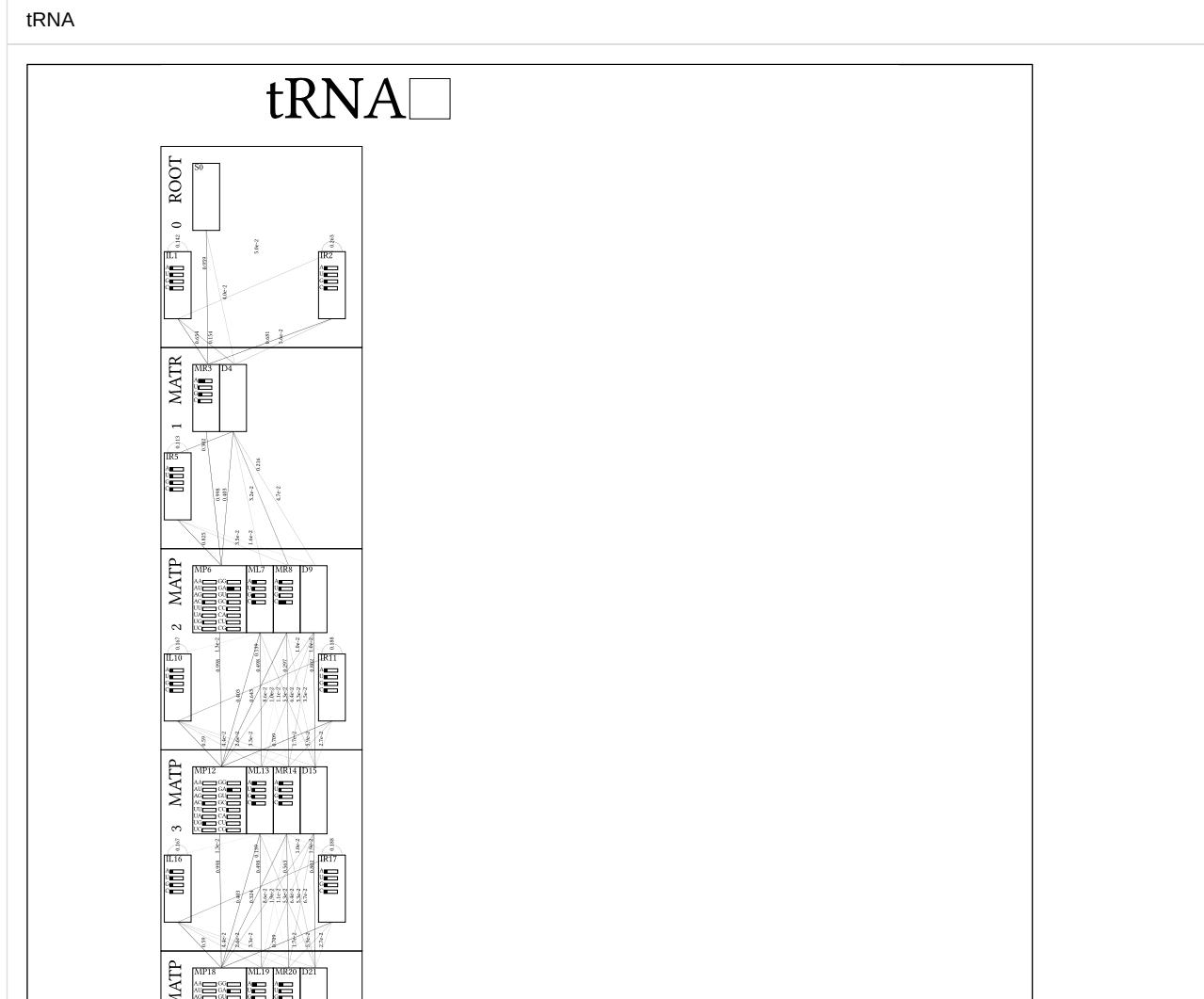
Model name	Model-pdf	Model-svg	Alignment-pdf
tRNA	pdf (http://192.52.2.124/data/examples/tRNAClan/tRNA.pdf)	svg (http://192.52.2.124/data/examples/tRNAClan/tRNA.svg)	pdf (http://192.52.2.124/data/examples/tRNAClan/tRNA.aln.pdf)

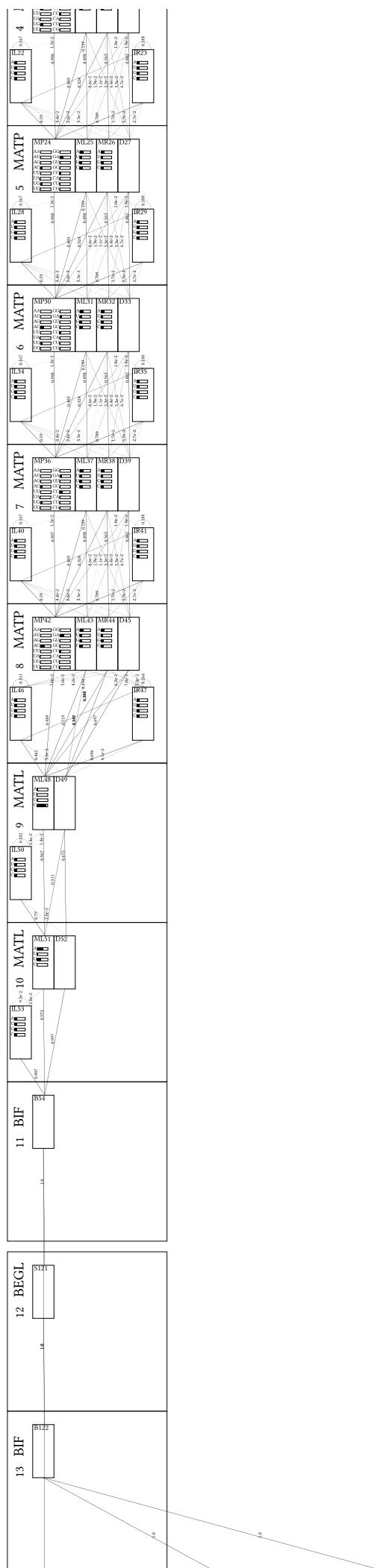
Secondary structure visualisation

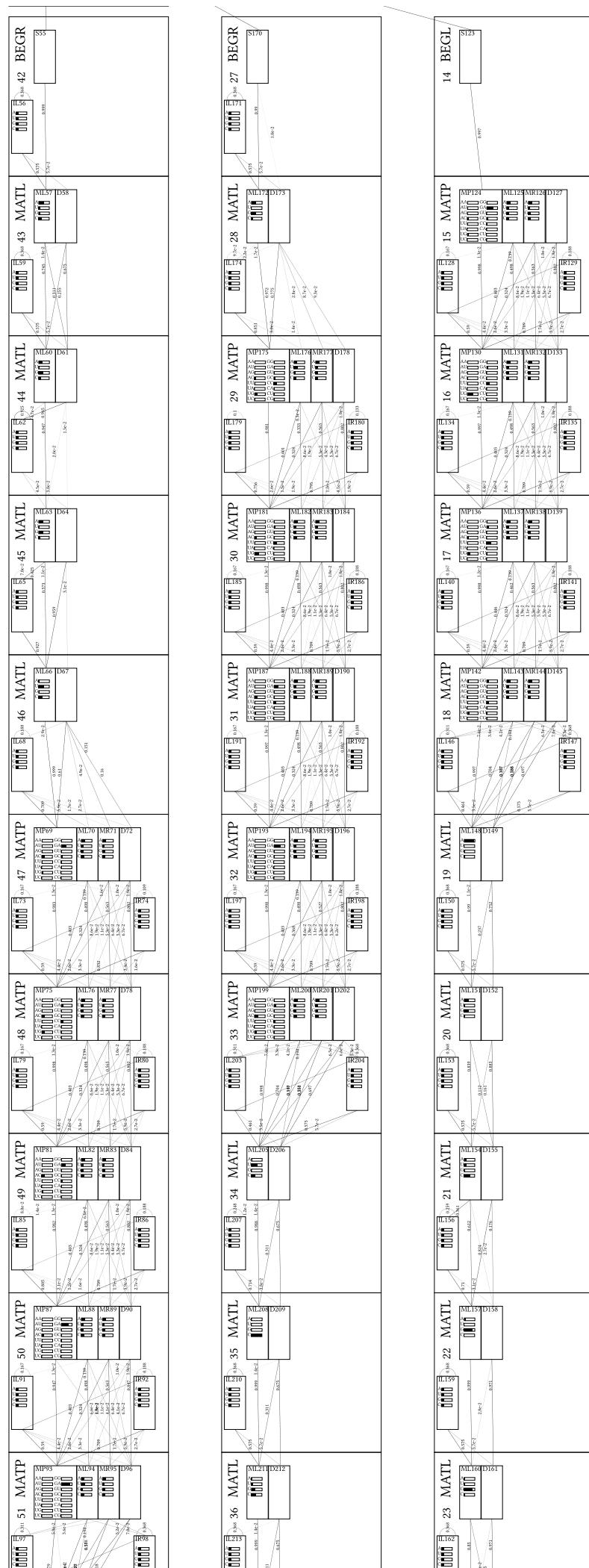
Table provides the consensus secondary structure visualisation of the structural alignment for each model by R2R (<http://breaker.research.yale.edu/R2R/>) as well as forna (<http://rna.tbi.univie.ac.at/forna/>)

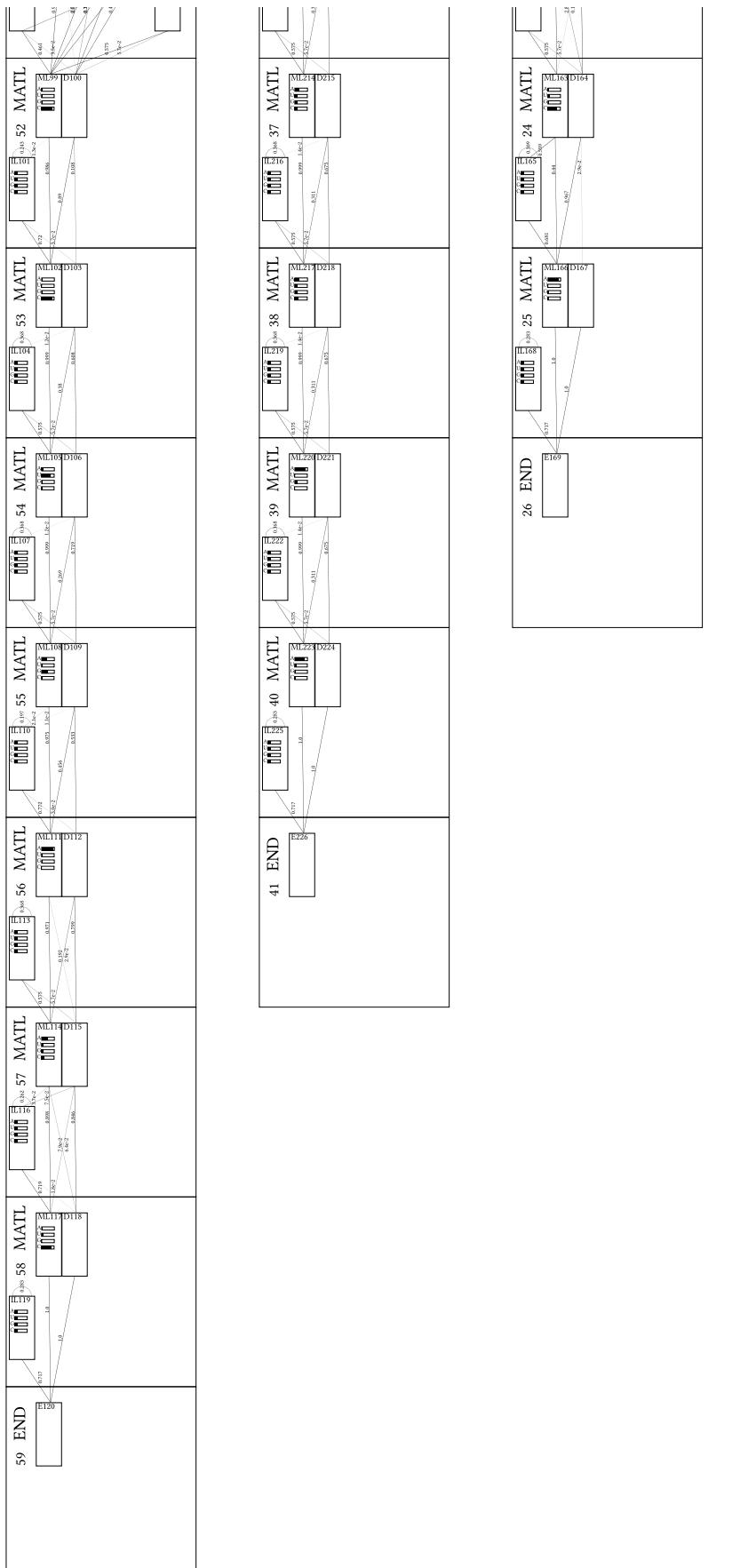
tRNA
r2r (http://192.52.2.124/data/examples/tRNAClan/tRNA.r2r.svg) / forna (http://rna.tbi.univie.ac.at/forna/forna.html) id=fasta&file=%3Eheader\nGGAGAUUAUGUCAGUGGUAGAGCGUCGGACUUAAAUCGAAGGGCGGGUUCGAAUCGCCUAUCUCCA\n((((.....))))((((.....))))....((((.....)))))))..)

The webservice also creates a zoom- and panable preview, that can be expanded by clicking.









cmcvt-output

cmcv also provides visualisation output for each model and corresponding alignment. The naming scheme is similar to *cmv*. However it annotates regions found to be linked my model comparison via color tags. Each provided model is associated with a color that is depicted in the legend table. The nodes of each model visualisation are then tagged with the color of the linked model. The coloring is also applied to the runing index of the alignment visualisation. This allows to study in detail which parts of models are connected, or shared in e.g. clans and which not. The following example shows only the result for the Hammerhead HH9 RNA.

Model name	Model-pdf	Model-svg
Hammerhead_HH9	pdf (http://192.52.2.124/data/examples/hammerheadClan/cmcv/Hammerhead_HH9.pdf)	svg (http://192.52.2.124/data/examples/hammerheadClan/cmcv/Hammerhead_HH9.svg)

Secondary structure visualisation

Moreover linked nodes are also annotated in the consensus secondary structure of the provided models, if a stockholm alignment with consensus structure has been provided. This allows to identify specific secondary structure elements that are linked between families. This is done for pairs of models with R2R (<http://breaker.research.yale.edu/R2R/>) as well as forna (<http://rna.tbi.univie.ac.at/forna/>).

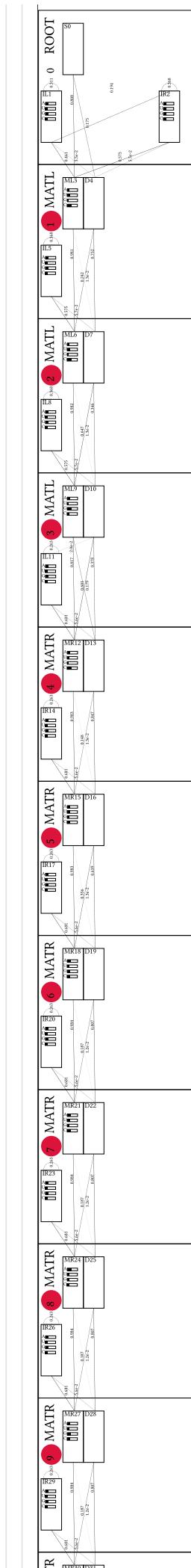
The webservice also creates a zoom- and panable preview, that can be expanded by clicking.

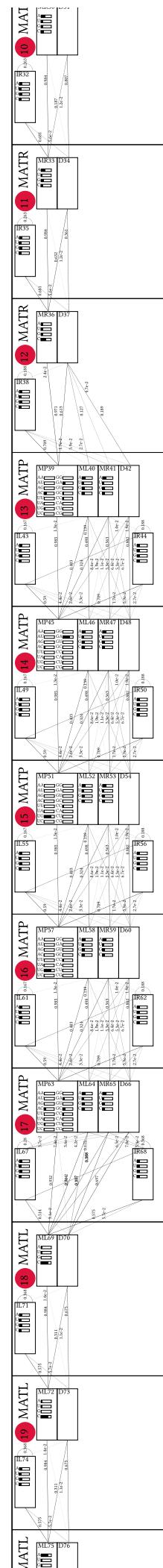
Hammerhead_HH9_CM

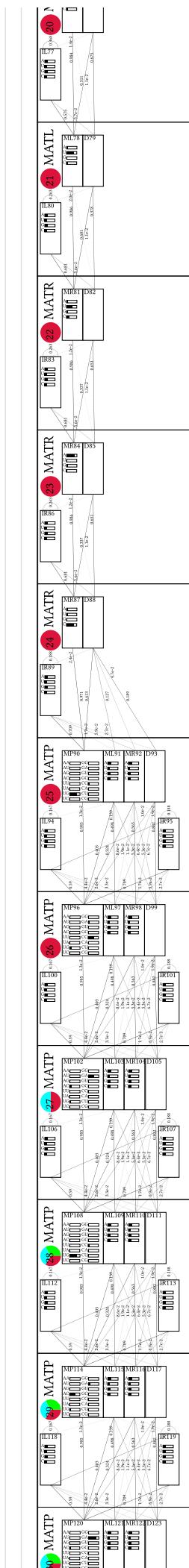
Hammerhead_HH9 [■]

Legend:

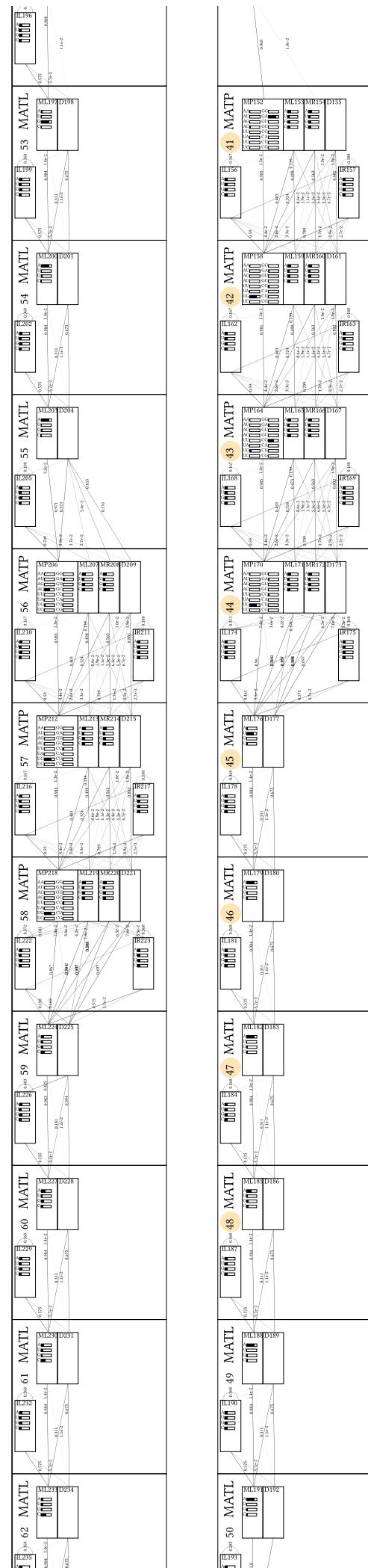
- Hammerhead_1 [■]
- Hammerhead_3 [□]
- Hammerhead_HH10 [■]
- Hammerhead_II [□]

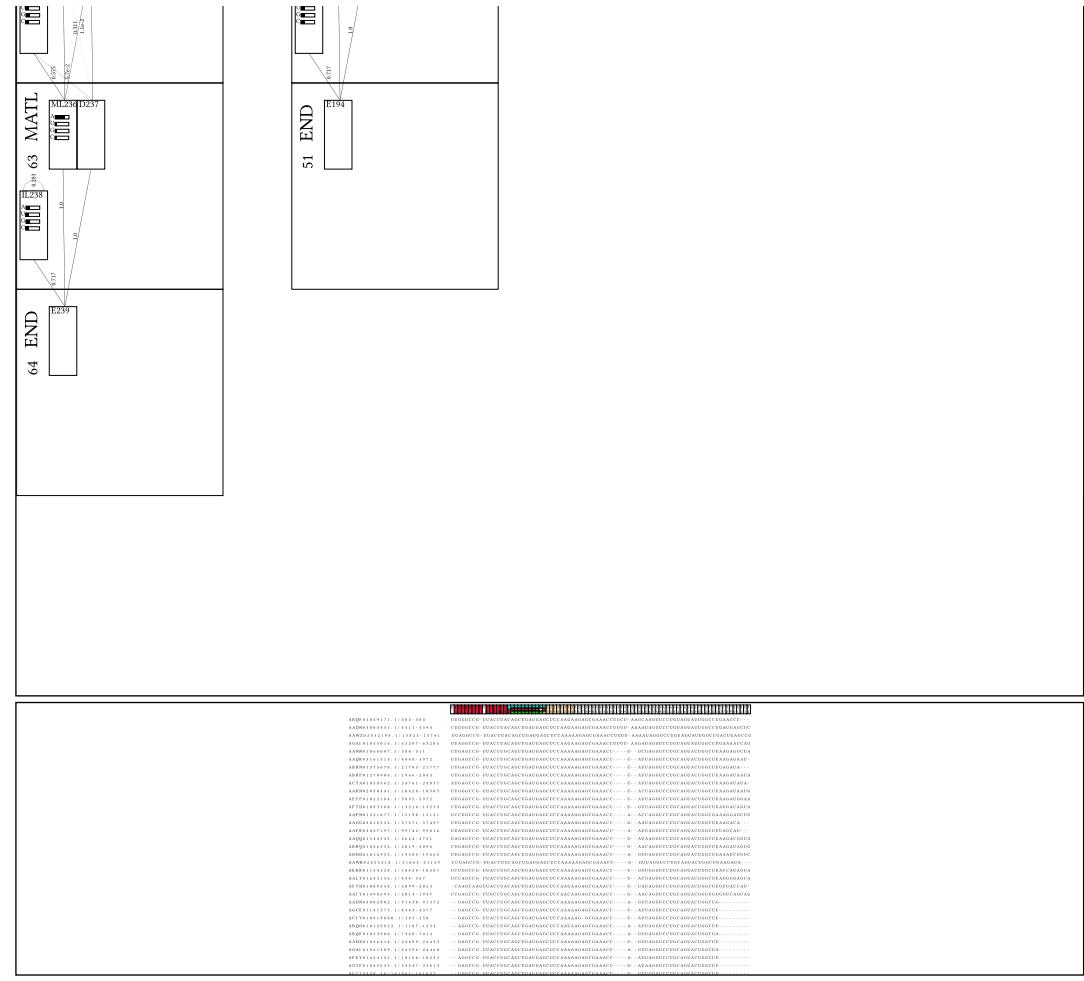












Examples

Example table (<http://192.52.2.124/data/CMS>) for minimal, simple and detailed visualisation of all Rfam models.

FE-2017