LRIscan

Long-range RNA-RNA interactions (LRIs) play an important role in viral replication. Only a few of these interactions are known in a limited number of viral species. Up to now, it has been impossible to screen a full viral genome for LRIs experimentally or in silico.

We present LRIscan, a tool for the prediction of long-range interactions in full viral genomes based on a multiple genome alignment. LRIscan is able to find interactions spanning thousands of nucleotides.

LRIscan is available at `http://www.rna.uni-jena.de/en/supplements/lriscan/`

LRIscan is based on the C-library of the ViennaRNA Package 2.0

Contact: markus.fricke2@uni-jena.de

REQUIREMENTS

Ruby 2.x.x (https://www.ruby-lang.org/ or http://rubyinstaller.org/)

Firefox/Chrome to display the results.

INSTALL (Linux)

pre compiled version

extract `LRIscan.tar.gz`
navigate to the destination folder of LRIscan `(HOME_LRISCAN)`
type: `./LRIscan.rb`

self compile

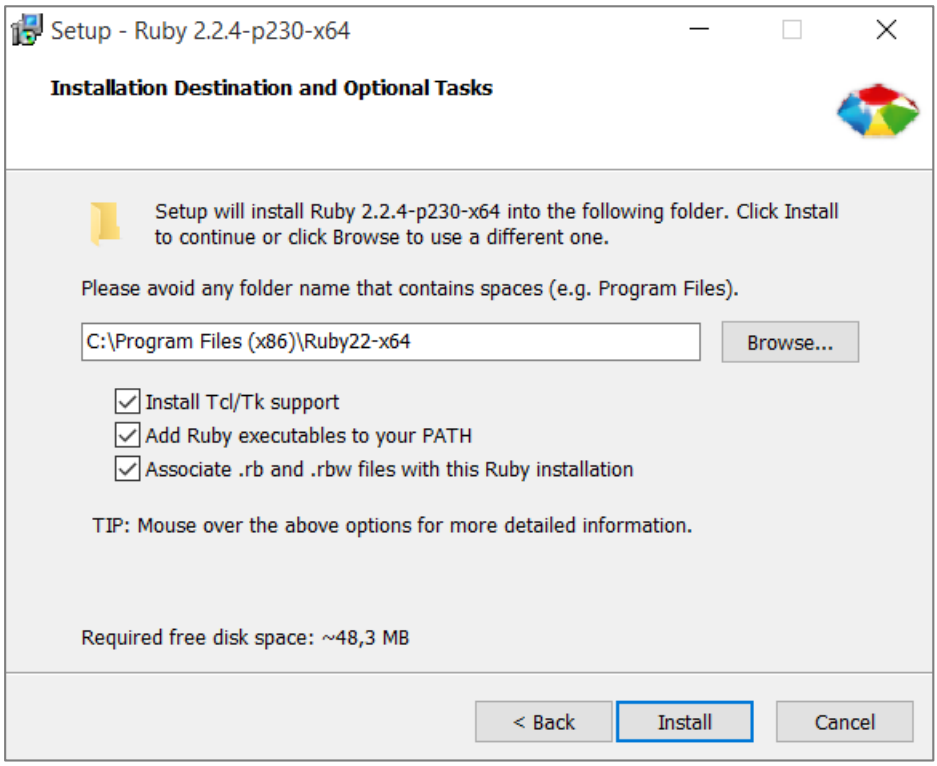
extract `LRIscan.tar.gz`
navigate to the destination folder of LRIscan `(HOME_LRISCAN)`

```
cd ./src/RNAalifold_interface
./configure
make

cd  ../../
./LRIscan.rb
```

INSTALL (Windows)

Install Ruby (http://rubyinstaller.org/) with **all** options



A Ruby installation without TCL/Tk supports no user interface.

Install `LRIscan_install.exe`

Run `LRIscan.exe` for LRIscan with user interface

Or

Run `ruby LRIscan.rb` for command line version

GENERAL OPTIONS

```
Usage: LRIscan [options]

-f    PATH        Input alignment file, [Clustal W format]
-o    [PATH]      Output file, [default='lri_output.tsv']
-d    [INT]       Minimum distance between interaction sides, [default=100nt]
-s    [INT]       Minimum size of seed interaction, [default=5bp]
-m    [DOUBLE]    MFE threshold, [default=-10 kcal/mol]
-i    [DOUBLE]    Minimum percentage of involved sequences in LRIs, [default=0.5]
-t    [DOUBLE]    Minimum conservation per base-pair, [default=0.95]
-c    [INT]       Number CPUs, [default=1]
-z    [INT]       z-score on [1], z-score off [0], z-score extension off [2], [default=2]
-a    [PATH]      List of SHAPE input files
```

INPUT

ALIGNMENT

The alignment input format should be in Clustal W format:

```
CLUSTAL W

SEQ_01       GCCAGCCCGGTGACGGACCCACCCCCGGAAAGGGACACCCTCGCGCGTACAGGACGCGAA
SEQ_02       GCCAGCCCGGAGACGGACCCACCCCGGGAAAGGGACGCCCACGCGAGAACAGGACGCGGA
SEQ_03       GCCAACCCGGGGGCGGACCCACCCCGGGAAAGAAGCACCCGCGCGGGGACAGGACACGAA
SEQ_04       GCCAACCCGGGGGAGGACCCACCCCGGGAAAGGGACGCCCCCGCGGGCACAGGACTCGCA
```

No whitespace character are allowed in the sequence name!

SHAPE

The shape input should be a tab separated list, including the sequence name and the absolute path of the shape reactivity files:

```
SEQ_01  /PATH/TO/SHAPE/REACTIVITY/FILE/FOR/SEQ_01
SEQ_02  /PATH/TO/SHAPE/REACTIVITY/FILE/FOR/SEQ_02
…
```

No whitespace character are allowed in the sequence name! The sequence name has to be identical with the sequence name of the input alignment file.

Shape reactivity file format:

```
1       A       -999
2       C       -999
3       C       -999
4       T       -6.3144
5       G       2.7304
6       C       -999
7       C       -0.061907
8       C       0.07257
9       C       -5.3412
10      T       1.1829
11      A       6.7803
12      A       1.717
13      T       0.82481
14      A       2.1647
15      G       -0.16705
16      G       0.42786
17      G       -0.033121
18      G       1.4219
19      C       -1.5601
20      G       0.30733
```

HOW TO

Here you can find a short guide of the most important options and scores. For a detailed explanation read our paper.

The detection of long-range RNA-RNA interaction is a complex task. Unfortunately, it is not possible to use a single scores to find a reliable interaction out of all possible interaction. But a combination of the different scores can reduce the big amount of possible interactions.

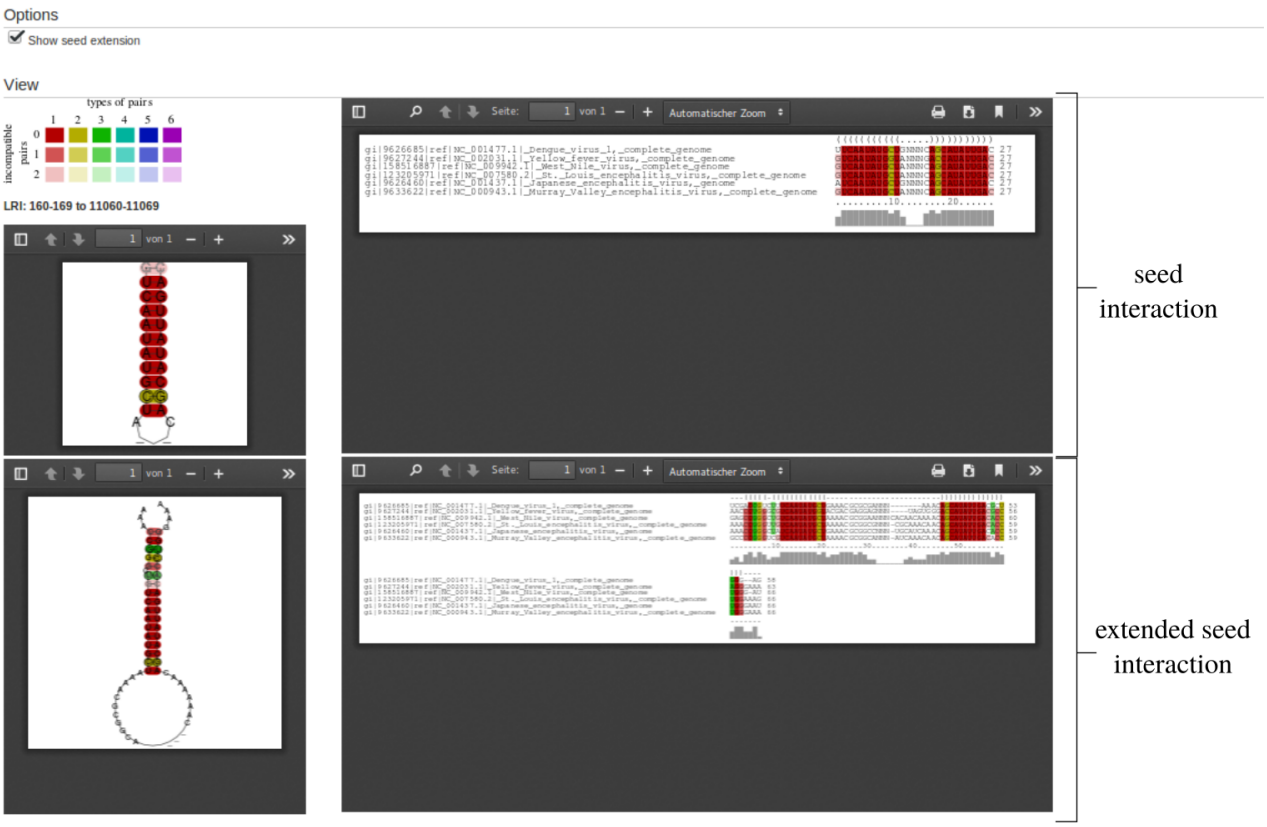
(i) It is crucial to decide **how many sequences** should be involved in your interaction. At default you will get all interactions which occur in at least 50% of the sequences. If you want to find only highly conserved interactions it is recommended to increase the `(-i)` option. This value depends on your alignment. For example, if you have regions of interest which are only covered by 30% of your sequences, set `-i` to `0.3`.

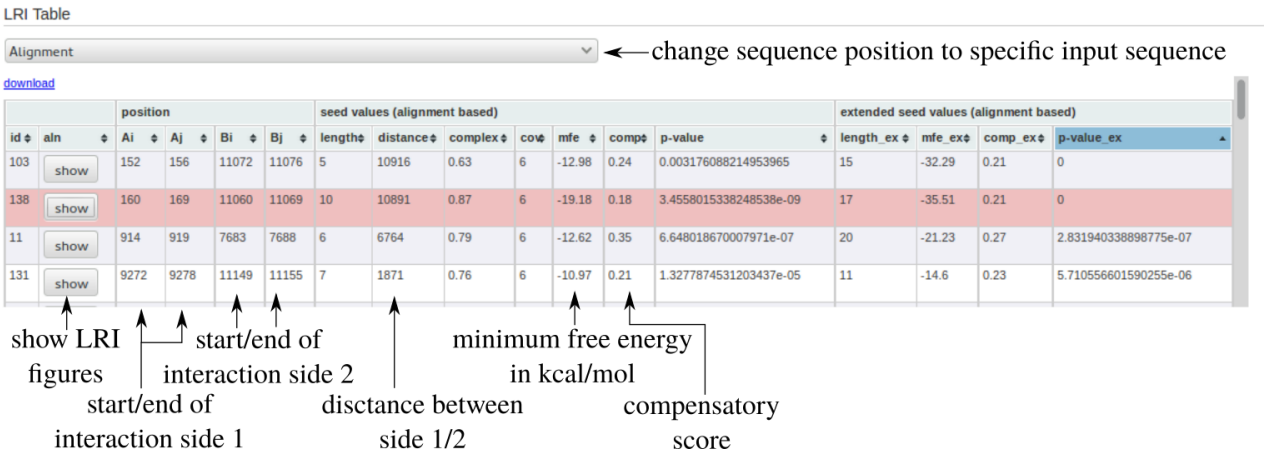
(ii) The **LRI score** (`-m`) is one of the most important values. At default you will get LRIs with a maximum score of -10 (the more negative the better). This score is the energy to build the LRI (kcal/mol). It is recommended to decrease the score if you want to have only the most stable interactions.

(iii) The **compensatory score** can be used to sort your results for interactions which have a good relation between compensatory base-pair mutations and incompatible base-pairs (the higher the better). Interactions with a high compensatory score are possible LRIs under evolutionary pressure and therefore good candidates for further analysis.

(iv) The **Z-score** and **p-value** can be used to reject LRIs which occur by chance (because of the nucleotide composition). LRIs with a p-value higher 0.05 should be rejected.

OUTPUT





You can find a html output in LRI.html :

The alignment table is also available as tab separated file (default `lri_output.tsv`)
The specific sequence position files can be found in the `tables` folder.
All figures can be found as PS and PDF file in the `ps` folder.