Theoretical and practical metagenomic approaches to viral discovery

Practical Session: ViennaRNA for RNA-RNA Interactions

Kevin Lamkiewicz, Manja Marz

24.10.2019

European Virus Bioinformatics Center

WHY RNA-RNA INTERACTIONS?

WHY RNA-RNA INTERACTIONS?

Table 1 | **Functional intragenomic interactions in positive-strand RNA viruses**

Virus (genus)	RNA–RNA interaction	Viral process regulated	Refs
Many plant viruses in the *Tombusviridae* family and the Umbravirus and Luteovirus genera	3′ CITE–5′ UTR or 3′ CITE–5′ coding region	Translation initiation[5,6]	5,6
BYDV (Luteovirus)	3′ CITE–5′ UTR	Translation initiation	21,25,30
	Frameshift site–3′ UTR	Ribosomal frameshifting	47
CIRV (Tombusvirus)	3′ CITE–5′ UTR	Translation initiation	22,28
	PRTE–DRTE	Stop codon readthrough	50
TBSV (Tombusvirus)	3′ CITE– 5′ UTR	Translation initiation	26,27
	UL–DL	Genome replication	52
	AS1–RS1; AS2–RS2; DE–CE	sgmRNA transcription	81–84
FMDV (Apthovirus)	IRES– 3′ UTR	Translation initiation	32,33
	S-region– 3′ UTR	Possibly genome replication	33
CSFV (Pestivirus)	IRES– 3′ terminus	Translation	34
HCV (Hepacivirus)	IRES–5BSL3.2	Translation initiation	35–37
	5BSL3.2– 3′ UTR	Genome replication	38–42
DENV and WNV (Flavivirus)	5′ UAR–3′ UAR; 5′ DAR–3′ DAR; 5′ CS–3′ CS	Genome replication	53–61,65,66
TGEV (Coronavirus)	DE–PE; cBM–BM	sgmRNA-N transcription	77–79

Beth L. Nicholson and K. Andrew White (2014) ”Functional long-range RNA-RNA interactions in positive-strand RNA viruses.”,

Nature Reviews Microbiology volume 12, pages 493–504

Interactions for single sequences

RNAcofold

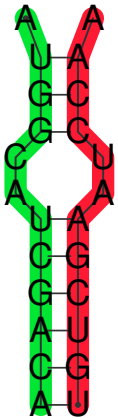
```
# RNAcofold works like RNAfold, but allows to specify two RNA sequences.
# These sequences are then allowed to form a dimer structure. In order
# to calculate the hybrid structure, it is necessary to concatenate the
# two RNA sequence, using & as a separator.

$> RNAcofold [OPTIONS] < sequences.fasta > sequences.cofold

# >seq1
# AUGGCAUCGACA
# >seq2
# UGUCGAAUCCAA

# RNAcofold Input:
# AUGGCAUCGACA&UGUCGAAUCCAA
```

RNAcofold



- First sequence is colored green
- Second sequence is colored red

RNACOFOLD WITH PARTITION FUNCTION

```
# We calculated the MFE structure of the interacting molecules (RNA dimer).
# RNAcofold also has the -p parameter implemented.

$> RNAcofold -p < sequences.fasta > sequences.cofold
```

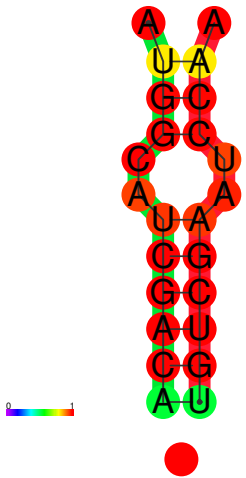
RNAcofold with partition function

```
# We calculated the MFE structure of the interacting molecules (RNA dimer).
# RNAcofold also has the -p parameter implemented.

$> RNAcofold -p < sequences.fasta > sequences.cofold
```

We can use `relplot.pl` on the PostScript files as well, but it looks a bit... weird.

RNACOFOLD WITH PARTITION FUNCTION



AA, AB, BB

```
# Until now, we just looked at the heterodimer of the two sequences.
# But how do the molecules behave individually?

$> RNAcofold -a < sequences.fasta > sequences.cofold
```

AA, AB, BB

```
# Until now, we just looked at the heterodimer of the two sequences.
# But how do the molecules behave individually?

$> RNAcofold -a < sequences.fasta > sequences.cofold
```

The AA and BB dimer describe the MFE structure of two RNA molecules of sequence one and sequence two, respectively.

RNAduplex

```
# RNAduplex is very similar to RNAcofold. Actually,
# it is a special case of RNAcofold, where only inter-molecular
# base pairs are allowed.

$> RNAduplex [OPTIONS] < sequences.fasta > sequences.duplex

```

RNAduplex

```
# RNAduplex is very similar to RNAcofold. Actually,
# it is a special case of RNAcofold, where only inter-molecular
# base pairs are allowed.

$> RNAduplex [OPTIONS] < sequences.fasta > sequences.duplex

# Alternative:
# RNAcofold -C < sequences_constrained.fasta
# sequences_constrained.fasta
# UAGCUAGCAUGCAUCGACGAU&CGAUGCAUGCAUGCAUGCAUC
# <<<<<<<<<<<<<<<<<<<<&>>>>>>>>>>>>>>>>>>>>
```

Co-Folding with MSAs

RNAALIDUPLEX

RNAALIDUPLEX

Not much implemented...

Unfortunately, ViennaRNA does not provide many possibilities for alignment-based co-folding analyses. Indeed, only the alignment version of `RNAduplex` is implemented in `RNAaliduplex`.

RNAALIDUPLEX

Not much implemented...

Unfortunately, ViennaRNA does not provide many possibilities for alignment-based co-folding analyses. Indeed, only the alignment version of `RNAduplex` is implemented in `RNAaliduplex`.

```
RNAaliduplex [OPTIONS] <file1.aln> <file2.aln>

# RNAaliduplex expects two input files (both CLUSTAL alignments)
# and predicts optimal and suboptimal binding sites.
# However, only inter-molecular base pairs are taken into account.
```

ALIGNMENT-BASED INTRA-MOLECULAR BASE PAIRS?

ALIGNMENT-BASED INTRA-MOLECULAR BASE PAIRS?

What would you do?

Discuss, play around, try to make some examples - I will go around and answer questions, discuss your ideas and help you as good as I can.

ALIGNMENT-BASED INTRA-MOLECULAR BASE PAIRS!

ALIGNMENT-BASED INTRA-MOLECULAR BASE PAIRS!

Different ways to do it

Most commonly, you'd want to do the following:

1. Extract your sequences and align them individually
2. Merge the alignments, use `'NNNNN'` as a separator
3. Apply `RNAalifold` on the alignment

Long-Range Interactions

We remember...

We remember...

Table 1 | **Functional intragenomic interactions in positive-strand RNA viruses**

Virus (genus)	RNA–RNA interaction	Viral process regulated	Refs
Many plant viruses in the *Tombusviridae* family and the Umbravirus and Luteovirus genera	3′ CITE–5′ UTR or 3′ CITE–5′ coding region	Translation initiation[5,6]	5,6
BYDV (Luteovirus)	3′ CITE–5′ UTR	Translation initiation	21,25,30
	Frameshift site–3′ UTR	Ribosomal frameshifting	47
CIRV (Tombusvirus)	3′ CITE–5′ UTR	Translation initiation	22,28
	PRTE–DRTE	Stop codon readthrough	50
TBSV (Tombusvirus)	3′ CITE– 5′ UTR	Translation initiation	26,27
	UL–DL	Genome replication	52
	AS1–RS1; AS2–RS2; DE–CE	sgmRNA transcription	81–84
FMDV (Apthovirus)	IRES– 3′ UTR	Translation initiation	32,33
	S-region– 3′ UTR	Possibly genome replication	33
CSFV (Pestivirus)	IRES– 3′ terminus	Translation	34
HCV (Hepacivirus)	IRES–5BSL3.2	Translation initiation	35–37
	5BSL3.2– 3′ UTR	Genome replication	38–42
DENV and WNV (Flavivirus)	5′ UAR–3′ UAR; 5′ DAR–3′ DAR; 5′ CS–3′ CS	Genome replication	53–61,65,66
TGEV (Coronavirus)	DE–PE; cBM–BM	sgmRNA-N transcription	77–79

Beth L. Nicholson and K. Andrew White (2014) "Functional long-range RNA-RNA interactions in positive-strand RNA viruses.",

Nature Reviews Microbiology volume 12, pages 493–504

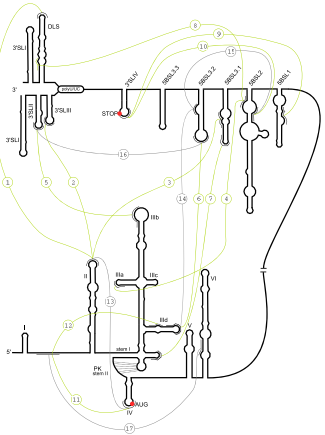
WHY LRIS?

- Interaction spans distances between a few hundred and several thousands of nucleotides
- few are described in positive stranded RNA viruses
- often located in loop regions (bulges, hairpins, ...)

WHY LRIS?

- Interaction spans distances between a few hundred and several thousands of nucleotides
- few are described in positive stranded RNA viruses
- often located in loop regions (bulges, hairpins, ...)
  $\Rightarrow$ pseudo-knots!

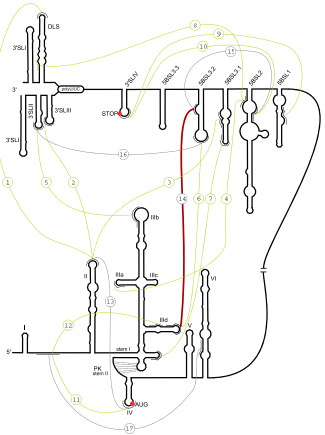
RNA-RNA INTERACTIONS ARE CRUCIAL FOR RNA VIRUSES



Fricke, M. *et al.* (2015). Conserved RNA secondary structures and long-range interactions in hepatitis C viruses. RNA, http://doi.org/10.1261/rna.049338.114

- Hepatitis C virus: (+)ssRNA
- around 9 kb in size

RNA-RNA INTERACTIONS ARE CRUCIAL FOR RNA VIRUSES

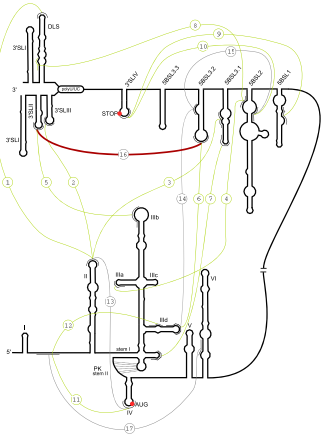


Fricke, M. *et al.* (2015). Conserved RNA secondary structures and long-range interactions in hepatitis C viruses. RNA, http://doi.org/10.1261/rna.049338.114

- Hepatitis C virus: (+)ssRNA
- around 9 kb in size

- Initiation of Translation

RNA-RNA INTERACTIONS ARE CRUCIAL FOR RNA VIRUSES



- Hepatitis C virus: (+)ssRNA
- around 9 kb in size

- Initiation of Translation
- Initiation of Replication

Fricke, M. *et al.* (2015). Conserved RNA secondary structures and long-range interactions in hepatitis C viruses. RNA, http://doi.org/10.1261/rna.049338.114

Exercise:

Take any LRI from HCV, described in the following paper, and try to reconstruct / predict it with the ViennaRNA package.

Fricke, M. *et al.* (2015). Conserved RNA secondary structures and long-range interactions in hepatitis C viruses. RNA, http://doi.org/10.1261/rna.049338.114