

# Characterization of conserved Flavivirus 5'UTR elements

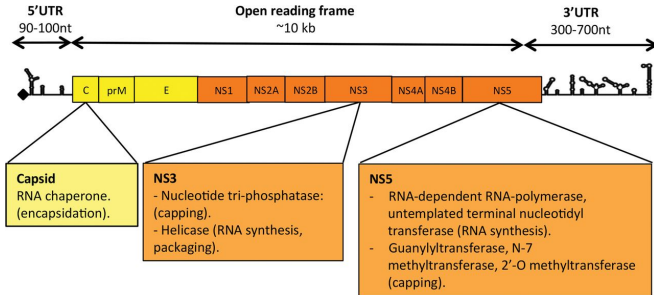
Michael T. Wolfinger

Department of Theoretical Chemistry  
University of Vienna

Bled, Slovenia, February 16, 2017

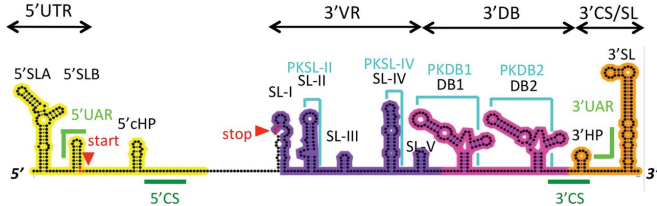
## Flavivirus genome characteristics

- Single-stranded, positive-strand RNA viruses
- Enters cell through receptor-mediated endocytosis
- Capped, non-polyadenylated genome (gRNA) of 10-12kb length
- Encodes a single ORF, flanked by structured 5'-UTR and 3'-UTR
- Translation of FV ORF yields a single polyprotein



## Flavivirus genome characteristics

- Single-stranded, positive-strand RNA viruses
- Enters cell through receptor-mediated endocytosis
- Capped, non-polyadenylated genome (gRNA) of 10-12kb length
- Encodes a single ORF, flanked by structured 5'-UTR and 3'-UTR
- Translation of FV ORF yields a single polyprotein

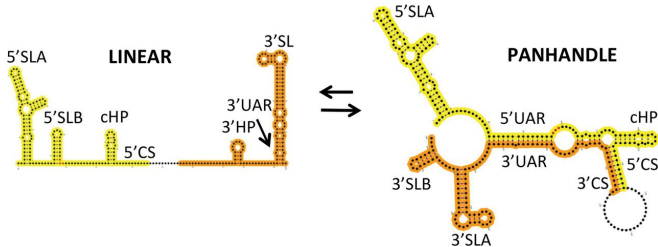


## Flavivirus replication

- FV RNA replication happens via asymmetric replication cycle
- Viral RNA-dependent RNA polymerase (RdRp) synthesizes a full-length minus-strand (-)gRNA
- (-)gRNA is a template for several rounds of (+)gRNA synthesis
- Ratio (-)gRNA / (+)gRNA between 1:10 and 1:100
- Genome cyclisation is required for (-)gRNA synthesis

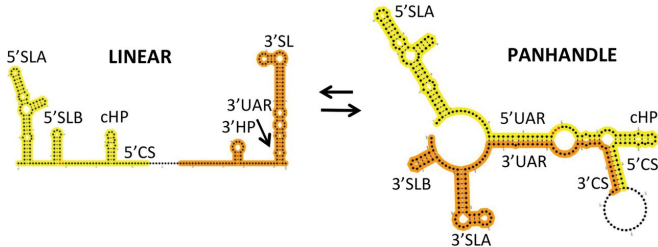
## Flavivirus replication

- FV RNA replication happens via asymmetric replication cycle
- Viral RNA-dependent RNA polymerase (RdRp) synthesizes a full-length minus-strand (-)gRNA
- (-)gRNA is a template for several rounds of (+)gRNA synthesis
- Ratio (-)gRNA / (+)gRNA between 1:10 and 1:100
- Genome cyclisation is required for (-)gRNA synthesis



## Flavivirus replication

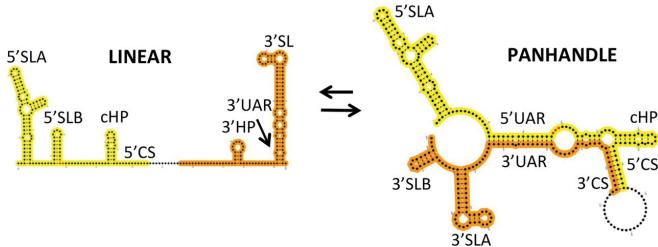
- FV RNA replication happens via asymmetric replication cycle
- Viral RNA-dependent RNA polymerase (RdRp) synthesizes a full-length minus-strand (-)gRNA
- (-)gRNA is a template for several rounds of (+)gRNA synthesis
- Ratio (-)gRNA / (+)gRNA between 1:10 and 1:100
- Genome cyclisation is required for (-)gRNA synthesis



- SLA acts as promoter / enhancer for viral RdRp

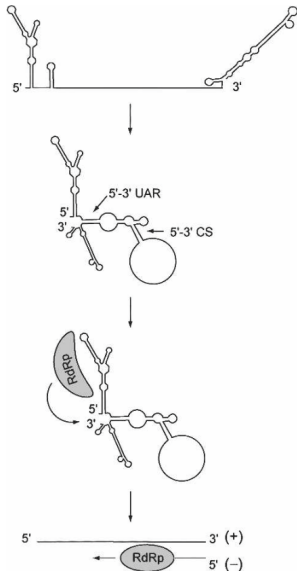
## Flavivirus replication

- FV RNA replication happens via asymmetric replication cycle
- Viral RNA-dependent RNA polymerase (RdRp) synthesizes a full-length minus-strand (-)gRNA
- (-)gRNA is a template for several rounds of (+)gRNA synthesis
- Ratio (-)gRNA / (+)gRNA between 1:10 and 1:100
- Genome cyclisation is required for (-)gRNA synthesis



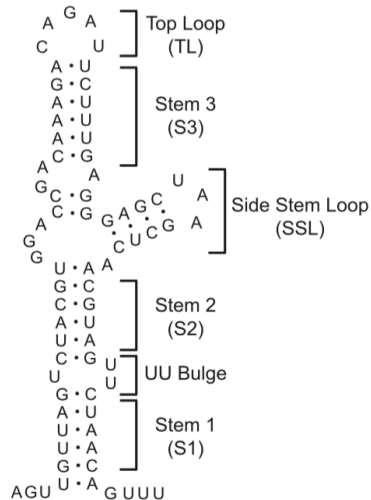
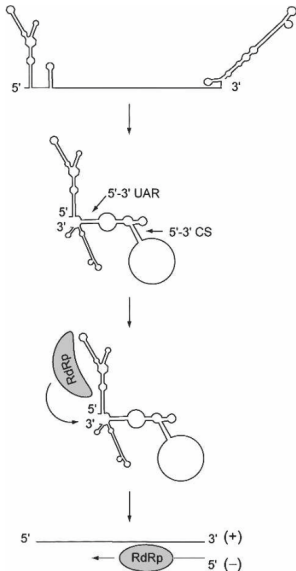
- SLA acts as promoter / enhancer for viral RdRp
- Viral RdRps are **error-prone**: Misincorporations every 40-70kb

# Flavivirus replication





# Flavivirus replication



Consensus structures computed from mlocarna alignments of all non-redundant representatives for each group obtained from NCBI

5'-R G0 A R

R

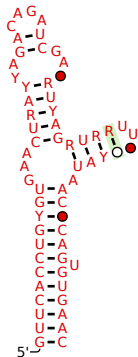
SGEV\_2.1.14.1/1  
TBV\_2.1.11.1/1  
OHV\_2.1.7.1/1  
LTV\_2.1.1.1/1  
LTV\_2.1.12.1/1  
ALKV\_2.1.2.1/1  
POW\_2.1.10.1/1  
TYUV\_2.2.1.1/1  
KAM\_2.2.3.1/1  
#=GC\_SS\_cons

[illegible]

## SLA in different FV groups

Consensus structures computed from mlocarna alignments of all non-redundant representatives for each group obtained from NCBI

### Kokobera virus group (KOKVG)



KOKV.1.4.1.1.1.1

NMV.1.4.2.1.1.1

#=GC SS\_cons

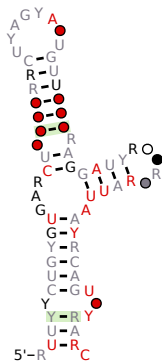
GUUACCCUGUGUAACUAACCAGACAGAUCGAGUUAGGUGAUUA-CAUAACACAGUGUGAAC

GUUACCCUGCGUGAACUGAUUAGACAGAUCGACAUCAGAUAGUUUCUAUAAUCACAGUGUGAAC

(((((((((((((((.....((((((.....)).....))))))((((.....))))).)))))).))))))

## SLA in different FV groups

Japanese encephalitis virus group (JEVG)

[illegible]

## SLA in different FV groups

## Spondweni virus group (SPOVG)

ZIKV\_1.6.2.21.1.1 UGUUGAUCUGUGAGUCCAGUCGACAGUUCAGUUUGAAGCGAAAGCUAGCAACAGUAUCAACA  
ZIKV\_1.6.2.161.1.1 UGUUGAUCUGUGUGAAUCCAGUCGCCAGAGUUCAGUUUGAAGCGAAAGCUAGCAACAGUAUCAACA  
ZIKV\_1.6.2.158.1.1 UGUUGAUCUGUGUGAGUCCAGUCGCCAGAGUUCAGUUUGAAGCGAAGCAGUAAACAGUAUCAACA  
ZIKV\_1.6.2.169.1.1 UGUUGAUCUGUGUGAAUCCAGUCGCCAGAGUUCAGUUUGAAGCGAGAGCUAAACAGUAUCAACA  
ZIKV\_1.6.2.84.1.1 AGUUGAUCUGUGUGAAUCCAGUCGCCAGUUCAGUUUGAAGCGAAAGCUAGCAACAGUAUCAACA  
ZIKV\_1.6.2.150.1.1 UGUU--UCUGUGUGAAUCCAGUCGCCAGUUCAGUUUGAAGCGAAAGCUAGCAACAGUAUCAACA  
ZIKV\_1.6.2.141.1.1 UGUUACUUGUUGCCAGUCAGUCGCCAGAGUUCAGUUUGAAGCGAAAGCUAGCAACAGUAUCAACA

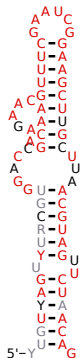
#=GC SS\_cons ((((((((((((C-)))))))))(((C-))))))(((C-))))))(((C-))))))(((C-))))))



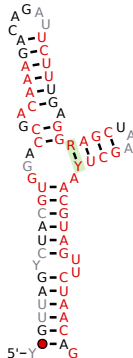
## Dengue virus SLA

There are four DENV serotypes, probably five ...

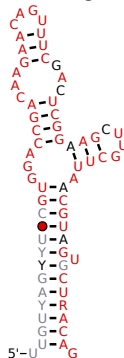
DENV1



DENV2



DENV3



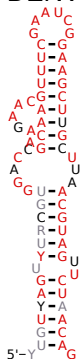
DENV4



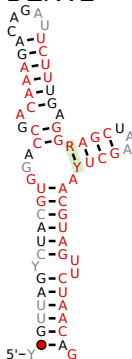
## Dengue virus SLA

There are four DENV serotypes, probably five ...

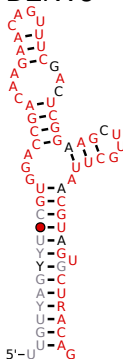
DENV1



DENV2



DENV3



DENV4



### Alternative facts:

- DENV1 folds into another consensus structure
- We need to deconvolute that !



## RNAaliSplit objectives

```
DENV1.1.2.1.156.1.1 CUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGWUCUAACAG
DENV1.1.2.1.21.1.1 CUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGKUCUAACAG
DENV1.1.2.1.1731.1.1 UUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUCGC---UUAACGUAGUUCUAACAG
DENV1.1.2.1.229.1.1 UUGUUAGU-CUGUCUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAACAG
DENV1.1.2.1.339.1.1 UUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUACAG
DENV1.1.2.1.157.1.1 UUGUUAGU-CUACGUGGACU-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAACAG
DENV1.1.2.1.168.1.1 UUGUUAGUACUACGUGGACC-GACAAGAAUCAGUUUCGAAUCGGAAGCUUGCCGGUUAACGUAGUUCUAACAG
DENV1.1.2.1.402.1.1 GACUCAGA-UACACGCGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAACAG
#=GC SS_cons .((((((.(((((((.(.((...((....)))...)))...)).....)))))..)))))..
```

## RNAaliSplit objectives

```
DENV1.1.2.1.156.1.1 CUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGWUCUAAACAG
DENV1.1.2.1.21.1.1 CUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGKUCUAAACAG
DENV1.1.2.1.1731.1.1 UUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUCGC---UUAACGUAGUUCUAAACAG
DENV1.1.2.1.229.1.1 UUGUUAGU-CUGUCUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAAACAG
DENV1.1.2.1.339.1.1 UUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAAACAG
DENV1.1.2.1.157.1.1 UUGUUAGU-CUACGUGGACU-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAAACAG
DENV1.1.2.1.168.1.1 UUGUUAGUACUACGUGGACC-GACAAGAAUCAGUUUCGAAUCGGAAGCUUGC CGGUUACGUAGUUCUAAACAG
DENV1.1.2.1.402.1.1 GACUCAGA-UACCACGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAAACAG
#=GC SS_cons .((((((..(((((((..((..((...)))...)))...)).....)))))..)))))..
```

- Find subalignments that have alternative consensus structures
- Identify sequences that do not fit the MSA

## RNAaliSplit objectives

```
DENV1.1.2.1.156.1.1 CUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGWUCUAAACAG
DENV1.1.2.1.21.1.1 CUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGKUCUAAACAG
DENV1.1.2.1.1731.1.1 UUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUCGC---UUAACGUAGUUCUAAACAG
DENV1.1.2.1.229.1.1 UUGUUAGU-CUGUCUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAAACAG
DENV1.1.2.1.339.1.1 UUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAAACAG
DENV1.1.2.1.157.1.1 UUGUUAGU-CUACGUGGACU-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAAACAG
DENV1.1.2.1.168.1.1 UUGUUAGUACUACGUGGACC-GACAAGAAUCAGUUUCGAAUCGGAAGCUUGC CGGUUAAACGUAGUUCUAAACAG
DENV1.1.2.1.402.1.1 GACUCAGA-UACCACGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAAACAG
#=GC SS_cons .(((((((.(((((((.(.((...((....)))...)))...)).....)))))..)))))..
```

- Find subalignments that have alternative consensus structures
- Identify sequences that do not fit the MSA

Structure conservation

Covariation

## RNAaliSplit objectives

```
DENV1.1.2.1.156.1.1 CUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGWUCUAACAG
DENV1.1.2.1.21.1.1 CUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGKUCUAACAG
DENV1.1.2.1.1731.1.1 UUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUCGC---UUAACGUAGUUCUAACAG
DENV1.1.2.1.229.1.1 UUGUUAGU-CUGUCUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAACAG
DENV1.1.2.1.339.1.1 UUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUACAG
DENV1.1.2.1.157.1.1 UUGUUAGU-CUACGUGGACU-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAACAG
DENV1.1.2.1.168.1.1 UUGUUAGUACUACGUGGACC-GACAAGAAUCAGUUUCGAAUCGGAAGCUUGC CGGUUACGUAGUUCUAACAG
DENV1.1.2.1.402.1.1 GACUCAGA-UACCACGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAACAG
#=GC SS_cons .(((((((.(((((((.((.((...(((.....)))...)))...)).....)))))..)))))..
```

- Find subalignments that have alternative consensus structures
- Identify sequences that do not fit the MSA

Structure conservation

Covariation

How can we quantify them?

## RNAaliSplit objectives

```
DENV1.1.2.1.156.1.1 CUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGWUCUAAACAG
DENV1.1.2.1.21.1.1 CUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGKUCUAAACAG
DENV1.1.2.1.1731.1.1 UUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAAACAG
DENV1.1.2.1.229.1.1 UUGUUAGU-CUGUCUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAAACAG
DENV1.1.2.1.339.1.1 UUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAAACAG
DENV1.1.2.1.157.1.1 UUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAAACAG
DENV1.1.2.1.168.1.1 UUGUUAGUACUACGUGGACC-GACAAGAAUCAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAAACAG
DENV1.1.2.1.402.1.1 GACUCAGA-UACACGAGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAAACAG
#=GC SS_cons .(((((((.(((((((((((((...(((.....)))...)))...)).....)))))).))))).
```

- Find subalignments that have alternative consensus structures
- Identify sequences that do not fit the MSA

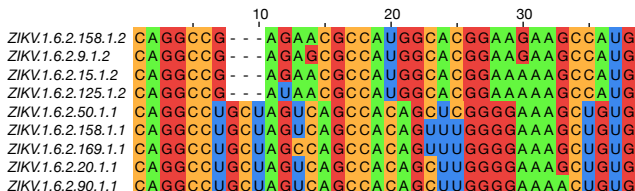
### Structure conservation

- SCI
  - z-score
  - mean pairwise identity
  - GC content
- RNAz SVM RNA-class probability

### Covariation

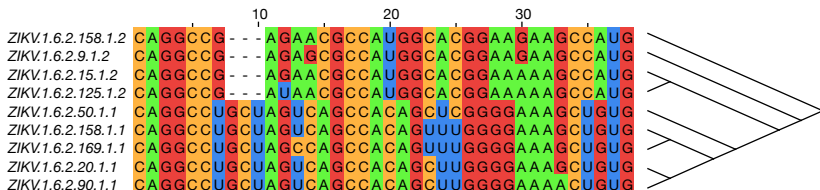
- compensatory mutations
- RNAalifold
- statistical significance
- R-scape

# RNAaliSplit workflow



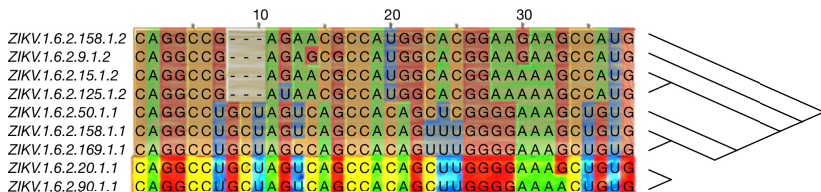
SVM prob	z-score	SCI	seq consensus	structure	SP	alignment
0.856846	-1.06	0.8613	9	((((((((.....))).(((((((.....)))))))).))))))	3	result.aln
0.939812	-1.17	0.8497	7	(((((.....(((.((((.....)))))))).))))))	3	split7.set1.aln
0.036973	-0.65	0.9671	2	(((((.....(((.....)))(((((((.....)))))))).))))))	0	split7.set2.aln
0.051212	-0.51	1.0091	2	(((((.....(((.....(((.....)))(((((((.....)))))))).))))))	0	split8.set1.aln
0.887830	-1.18	0.8409	7	(((((.....(((.....)))))))).))))))	2	split8.set2.aln
0.899553	-1.43	0.9442	5	(((((.....(((.....)))))))).))))))	2	split9.set1.aln
0.063624	-0.56	0.9789	4	(((((.....(((.....(((.....)))(((((((.....)))))))).))))))	0	split9.set2.aln
...						

# RNAaliSplit workflow



SVM prob	z-score	SCI	seq consensus structure	SP	alignment
0.856846	-1.06	0.8613	9 (((((((((.....))).((((((((.....)))))))).)))))) 3	result.aln	
0.939812	-1.17	0.8497	7 (((((((.....(((.((((((((.....)))))))).)))))) 3	split7.set1.aln	
0.036973	-0.65	0.9671	2 (((((((.....(((.((((((((.....)))))))).)))))) 0	split7.set2.aln	
0.051212	-0.51	1.0091	2 (((((((.....(((.((((((((.....)))))))).)))))) 0	split8.set1.aln	
0.887830	-1.18	0.8409	7 (((((((.....(((.((((((((.....)))))))).)))))) 2	split8.set2.aln	
0.899553	-1.43	0.9442	5 (((((((.....(((.((((((((.....)))))))).)))))) 2	split9.set1.aln	
0.063624	-0.56	0.9789	4 (((((((.....(((.((((((((.....)))))))).)))))) 0	split9.set2.aln	
...					

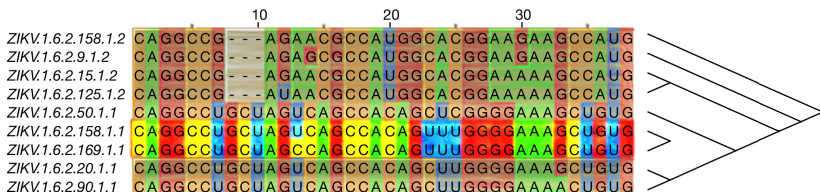
# RNAaliSplit workflow



SVM prob	z-score	SCI	seq consensus	structure	SP alignment
0.856846	-1.06	0.8613	9	((((((((.....))).(((((((.....)))))))).))))))	3 result.aln
0.939812	-1.17	0.8497	7	(((((.....(((.....)))))))).))))))	3 split7.set1.aln
0.036973	-0.65	0.9671	2	(((((.....(((.....)))))))).))))))	0 split7.set2.aln
0.051212	-0.51	1.0091	2	(((((.....(((.....)))))))).))))))	0 split8.set1.aln
0.887830	-1.18	0.8409	7	(((((.....(((.....)))))))).))))))	2 split8.set2.aln
0.899553	-1.43	0.9442	5	(((((.....(((.....)))))))).))))))	2 split9.set1.aln
0.063624	-0.56	0.9789	4	(((((.....(((.....)))))))).))))))	0 split9.set2.aln
...					

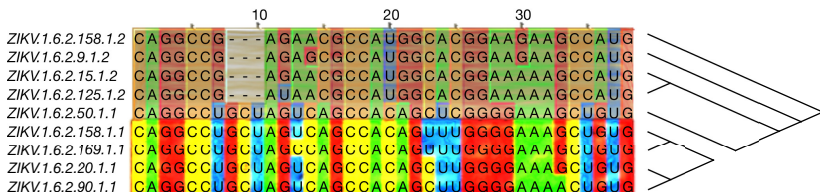


# RNAaliSplit workflow



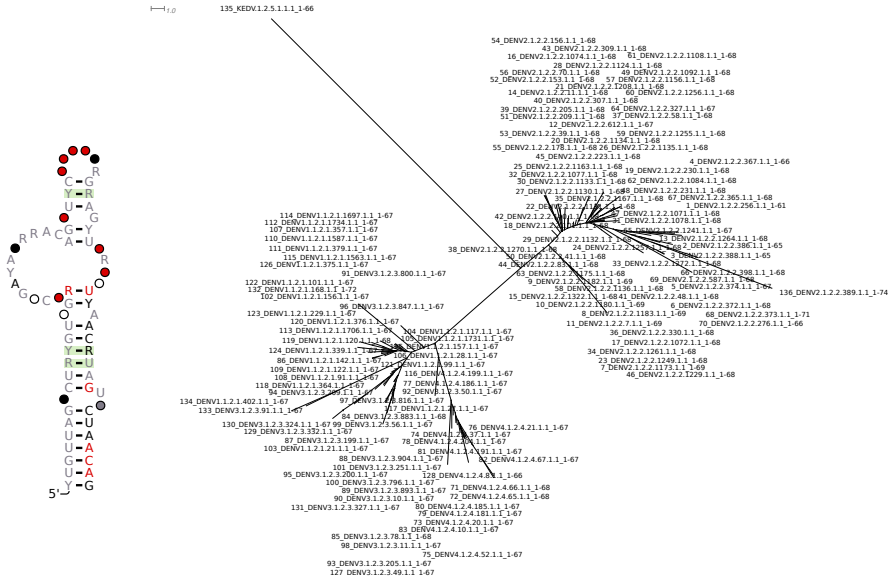
SVM prob	z-score	SCI	seq consensus	structure	SP alignment
0.856846	-1.06	0.8613	9	((((((((.....))).(((((((.....)))))))).))))))	3 result.aln
0.939812	-1.17	0.8497	7	(((((.....(((.....)))))))).))))))	3 split7.set1.aln
0.036973	-0.65	0.9671	2	(((((.....(((.....)))))))).))))))	0 split7.set2.aln
0.051212	-0.51	1.0091	2	(((((.....(((.....)))))))).))))))	0 split8.set1.aln
0.887830	-1.18	0.8409	7	(((((.....(((.....)))))))).))))))	2 split8.set2.aln
0.899553	-1.43	0.9442	5	(((((.....(((.....)))))))).))))))	2 split9.set1.aln
0.063624	-0.56	0.9789	4	(((((.....(((.....)))))))).))))))	0 split9.set2.aln
...					

# RNAaliSplit workflow



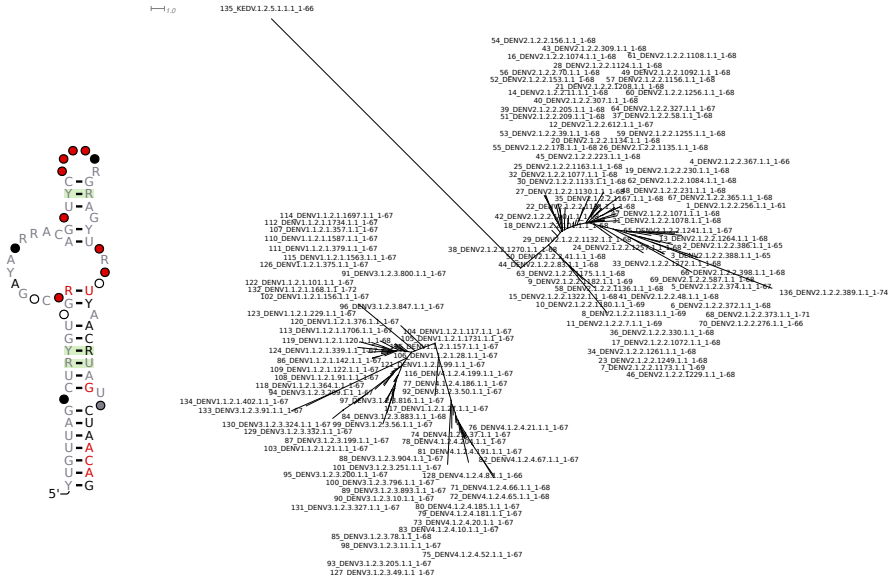
SVM prob	z-score	SCI	seq consensus	structure	SP alignment
0.856846	-1.06	0.8613	9	((((((((.....))).(((((((.....)))))))).))))))	3 result.aln
0.939812	-1.17	0.8497	7	(((((.....(((.....)))))))).))))))	3 split7.set1.aln
0.036973	-0.65	0.9671	2	(((((.....(((.....)))))))).))))))	0 split7.set2.aln
0.051212	-0.51	1.0091	2	(((((.....(((.....)))))))).))))))	0 split8.set1.aln
0.887830	-1.18	0.8409	7	(((((.....(((.....)))))))).))))))	2 split8.set2.aln
0.899553	-1.43	0.9442	5	(((((.....(((.....)))))))).))))))	2 split9.set1.aln
0.063624	-0.56	0.9789	4	(((((.....(((.....)))))))).))))))	0 split9.set2.aln
...					

## Dengue virus SLA



NJ tree; 146 pairwise different DENVG SLA sequences

## Dengue virus SLA



NJ tree; 146 pairwise different DENVG SLA sequences



NJ tree, 64 pairwise different DENV 1,3,4 SLA sequences

A horizontal number line with arrows at both ends. A tick mark is labeled 10.0 at the right end.



NJ tree, 64 pairwise different DENV 1,3,4 SLA sequences

[illegible]

5'-U

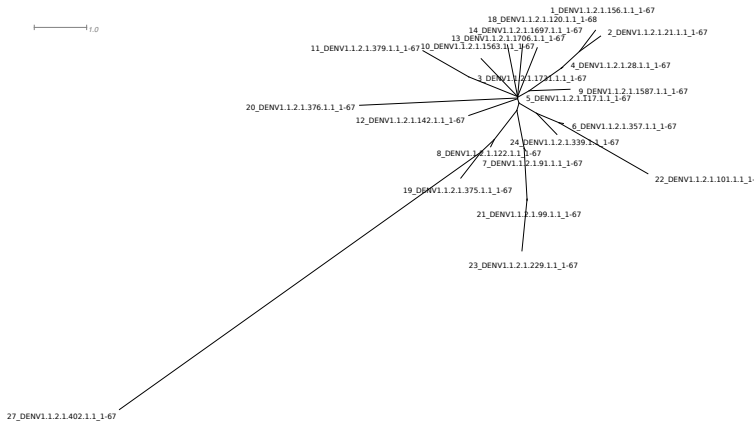
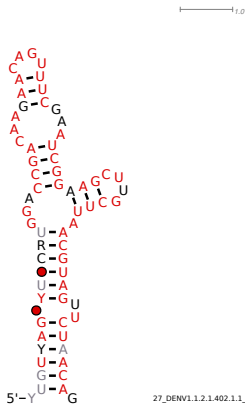
5'-Y





[illegible]

# Dengue virus SLA



NJ tree; 22 pairwise different DENV1 SLA sequences

## Summary

Flavivirus SLA elements are

- required for FV replication, mediating (-)gRNA synthesis by RdRp
- conserved among Mosquito-bourne and Tick-bourne FV

## Summary

### Flavivirus SLA elements are

- required for FV replication, mediating (-)gRNA synthesis by RdRp
- conserved among Mosquito-bourne and Tick-bourne FV

### RNAaliSplit is

- a heuristic approach for splitting MSA into sets that have alternative consensus structures
- a means for cleaning MSA before constructing covariance models
- universally applicable to moderate-sized RNA MSA
- implemented in a modular, object-oriented way based on the Perl Moose framework
- available soon as a stand-alone Perl Module

## Acknowledgments

### Collaborators

Andrea Tanzer

Roman Ochsenreiter

Ivo L. Hofacker



DK RNA Biology (FWF-W 1207)



Der Wissenschaftsfonds.

SFB RNA regulation of the  
transcriptome (FWF-F43)



universität  
wien



MEDIZINISCHE  
UNIVERSITÄT  
WIEN

# Thank you!