Characterization of conserved Flavivirus 5'UTR elements

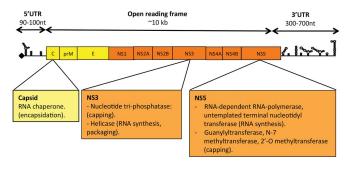
Michael T. Wolfinger

Department of Theoretical Chemistry University of Vienna

Bled, Slovenia, February 16, 2017

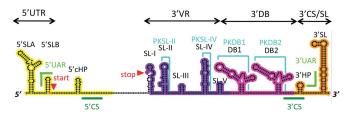
Flavivirus genome characteristics

- 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



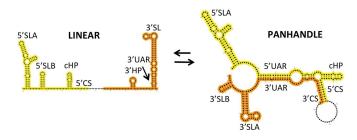
Flavivirus genome characteristics

- 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

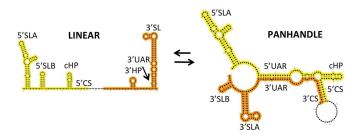


- FV RNA replication happens via asymmetric replication cycle
- Viral RNA-dependent RNA polymerase (RdRp) synthesizes a full-length minus-strand (-)gRNA
- (-)gRNA is a template for several rounds of (+)gRNA synthesis
- Ratio (-)gRNA / (+)gRNA between 1:10 and 1:100
- Genome cyclisation is required for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

- 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

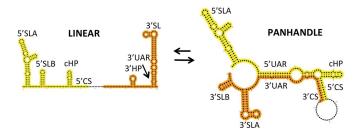


- FV RNA replication happens via asymmetric replication cycle
- Viral RNA-dependent RNA polymerase (RdRp) synthesizes a full-length minus-strand (-)gRNA
- (-)gRNA is a template for several rounds of (+)gRNA synthesis
- Ratio (-)gRNA / (+)gRNA between 1:10 and 1:100
- Genome cyclisation is required for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

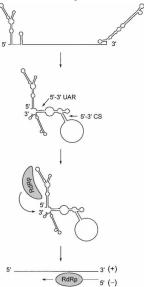


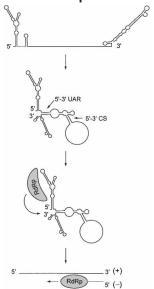
SLA acts as promoter / enhancer for viral RdRp

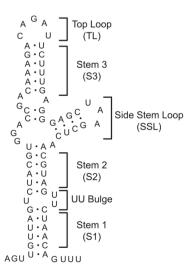
- FV RNA replication happens via asymmetric replication cycle
- Viral RNA-dependent RNA polymerase (RdRp) synthesizes a full-length minus-strand (-)gRNA
- (-)gRNA is a template for several rounds of (+)gRNA synthesis
- Ratio (-)gRNA / (+)gRNA between 1:10 and 1:100
- · Genome cyclisation is required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



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



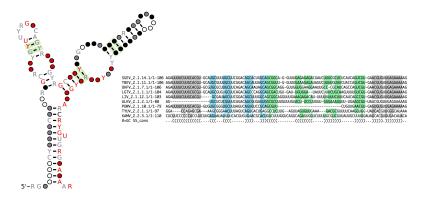




Lodeiro et al., J. Virol. (2009)

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

Tick-bourne Flavivirus group (TBFVG)



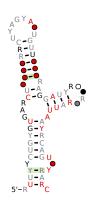
Consensus structures computed from mlocarna alignments of all non-redundant representatives for each group obtained from NCBI

Kokobera virus group (KOKVG)



KOKV.1.4.1.1.1.1 NMV.1.4.2.1.1.1 #=GC SS_cons Consensus structures computed from mlocarna alignments of all non-redundant representatives for each group obtained from NCBI

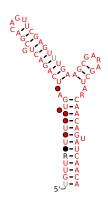
Japanese encephalitis virus group (JEVG)



JEV.1.3.1.1.1.1	G-UUUAUCUGUGUGAACUUCUUGGCUUAGUAUCGU-UGAGAAGAAUCC-AGA-GAUUAGU-GCAGUUUAAAC
JEV.1.3.1.163.1.1	G-UUAAUCUGUGUGAACUUCUUGGCUUAGUAUCGU-UGAGAAGAAUCG-AGA-GAUUAGU-GCAGUUUAAAC
JEV.1.3.1.130.1.1	G-UUUAUCUGUGUGAACUUCUUGGCUUUGUAUCGU-UGAGAAGAAUCG-AGA-GAUUAGU-GCAGUUUAAAC
ALFV.1.3.5.1.1.1	G-UUCAUCUGUGUGAACUUUUUGACUCAGGAUUGU-UGGAAGGGAUUG-AAA-GAUUAAU-ACAGUUUGAAC
JEV.1.3.1.165.1.1	G-UUUAUCUGUGUGAACUUCUUGGUUUAGUAUCGU-UGAGAAGAAUCG-AAA-GAUUAGU-GCAGUUUAAAC
JEV.1.3.1.75.1.1	G-UUUUAACGUGUGAACUUCUUGGCUUAGUAUCGU-CGAGAGAGAUCG-AGA-GAUUAGU-GCAGUUUAAAC
JEV.1.3.1.82.1.1	G-UUUU-UUUUGUGUGACUUCUGGGCUUAGUAUCGU-UGAGAAGAAUCG-AGA-GAUUAGU-GCAGUUUAAAC
JEV.1.3.1.84.1.1	G-UUUU-AACUGUGUGACUUCUUGGCUUAGUAUCGU-UGAGAAGAAUCG-AGA-GAUUAGU-GCAGUUUAAAC
USUV.1.3.7.89.1.1	G-UUGGCCUGUGUGAGCUCUACUACUUAGUAUUGU-UUUGGAGGAUCG-UUA-GAUUAAU-ACAGUGUUGAC
MVEV.1.3.6.2.1.1	G-UUCAUCUGCGUGAGCUUCCGAUCUCAGUAUUGU-UUGGAAGGAUCAUU-GAUUAAC-GCGGUUUGAAC
MVEV.1.3.6.1.1.1	G-UUCAUCUGCGUGAGCUUCCGAUCUCAGUAUUGU-UUGGAAGGAUCAUU-GAUUAAC-GUGGUUUGAAC
MVEV.1.3.6.4.1.1	G-UUCAUCUGCGUGAGCUUCCGAUCUCAGUAUUGU-UUGGGAGGAUCA-UUU-GAUUAAC-GCGGUUUGAAC
MVEV.1.3.6.9.1.1	G-UUCAUCUGCGUGAGCUUUCGAUCUCAGUAUUGU-UUGAAAGGAUCAUUUU-GAUUAAC-GCGGUUUGAAC
	GUUUCGCCUGUGUGAGCUGACAAACUUAGUAGUGU-UUGUGAGGAUUAACAACAAUUAAC-ACAGUGCGAGC
	GUAUUGACCCUGUGUGAGCUGACAAACUUAGUAGUGU-UUGUGAGGAUUAACAACAAUUAAC-ACAGUGCGAGC
	GUCGCCUGUGUGAGCUGACAAACUUAGUAGUGU-UUGUGAGGAUUAACAACAAUUAAC-ACAGUGCGAGC
WNV.1.3.2.973.1.1	G-UUCGCCUGUGUGAACUGACAAACUUAGUAGUGU-UUGUGAGGAUUAAUAACAAUUAAC-ACAGUGCGAGC
KUNV.1.3.4.43.1.1	G-UUCGCCUGUGUGAGCUGACAAACUUAGUAGUGU-UUGUGAGGAUUUUGAACAAUUAAC-ACAGUGCGAGC
	ACAACCCUGUGUGAGCUGACAAACUUAGUAGUGU-UUGUGAGGAUUAACAACAAUUAAC-ACAGUGCGAGC
	UUCGCCUGUGUGAGCUGACAAACUUAGUAGUGU-UUGUGAGGAUUACGAGGAUUAAU-ACAGUACGAGC
	GCCUGUGUGAGCUGACAAACUUAGUAGUGU-UUGUGAGGAUUAACAACAAUUAAC-ACAGUGCGAGC
USUV.1.3.7.10.1.1	GUUCGUCUGGGUGAGCUCUACUUAGUAUUGUUUUUGGAGGAUCGUGAGAUUAAC-ACAGUGCCGGC
USUV.1.3.7.95.1.1	GUUCGUCUGCGUGAGCUCUACUUAGUAUUGUUUUUGGAGGAUCGUGAGAUUAAC-ACAGUGCCGGC
USUV.1.3.7.24.1.1	GUCCGUCUGCGUGAGCUCUACUUAGUAUUGUUUUUGGAGGAUCGUGAGAUUAAC-ACAGUGCCGGC
USUV.1.3.7.46.1.1	GUUCGCCUGCGUGAGCUCUACUUAGUAUUGUUUUUGGAGGAUCGUGAGAUUAAC-ACAGUGCCGGC
USUV.1.3.7.9.1.1	GUUCGUCGGCGUGAGCUCUACUACUUAGUAUUGUUUUUGGAGGAUCGUGAGAUUAAC-ACAGUGCCGGC
SLEV.1.3.3.28.1.1	UGUUCGCGUCGGUGAGCGAAGAGGAAACAGAUUUC-UUUUUUGGAGGAUAACAACUUAACUU
SLEV.1.3.3.30.1.1	UGUUCGCGUUGGUGAGCGAGGAAGAAACAGAUUUC-UUUUUUGGAGGAUUACAACUUAACUU
	UCUCCUCAGAAUCCUGAAAAGAUUCUCAGUGU-UUGUGAGGAUUAACAACAAUUAAC-ACAGUGCGAGC
#=GC SS_cons	(((((((((()).)))).))((()))).)))).)))).

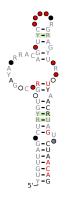
Consensus structures computed from mlocarna alignments of all non-redundant representatives for each group obtained from NCBI

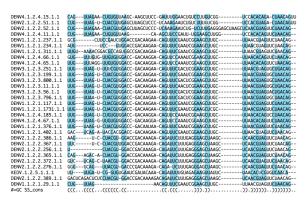
Spondweni virus group (SPOVG)



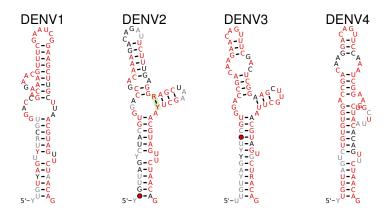
Consensus structures computed from mlocarna alignments of all non-redundant representatives for each group obtained from NCBI

Dengue virus group (DENVG)

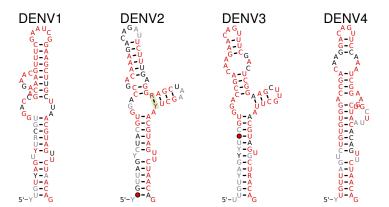




There are four DENV serotypes, probably five ...



There are four DENV serotypes, probably five ...



Alternative facts:

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

```
DENV1.1.2.1.156.1.1
                     CUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGWUCUAACAG
DENV1.1.2.1.21.1.1
                     CUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGKUCUAACAG
DENVI. 1. 2. 1. 1731. 1. 1. UUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUCGC---UUAACGUAGUUCUAACAG
DENV1.1.2.1.229.1.1
                     UUGUUAGU-CUGUCUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAACAG
                     UUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUUACAG
DENV1.1.2.1.339.1.1
DENV1.1.2.1.157.1.1
                     UIGUIJAGU-CUACGUGGACU-GACAAGAA-CAGUUJCGAAUCGGAAGCUJGC---JUJAACGJJAGUJCUAACAG
DENV1.1.2.1.168.1.1
                     LIJIGIJIJAGIJACIJACGUGGACC-GACAAGAAUCAGUUUCGAAUCGGAAGCUUGCCGGUUAACGUAGUUCUAACAG
                     GACUCAGA-UACCACGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAACAG
DENV1.1.2.1.402.1.1
#=GC_SS_cons
                     .((((((,((((((,((,(,(,,,,)),..)))...))...))))...))))...
```

```
CUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGWUCUAACAG
DENV1 1 2 1 156 1 1
DENV1.1.2.1.21.1.1
                    CUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGKUCUAACAG
DENV1.1.2.1.1731.1.1
                    UUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUCGC---UUAACGUAGUUCUAACAG
                    UUGUUAGU-CUGUCUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAACAG
DENV1.1.2.1.229.1.1
DENV1.1.2.1.339.1.1
                    UUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUUACAG
                    UUGUUAGU-CUACGUGGACU-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAACAG
DENV1.1.2.1.157.1.1
DENV1.1.2.1.168.1.1
                    UUGUUAGUACUACGUGGACC-GACAAGAAUCAGUUUCGAAUCGGAAGCUUGCCGGUUAACGUAGUUCUAACAG
                    GACUCAGA-UACCACGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAACAG
DENV1.1.2.1.402.1.1
#=GC SS_cons
                    .((((((,((((((,((,(,(,,,,)),..)))...))...))))...))))...
```

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

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

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

Structure conservation

Covariation

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

Structure conservation

Covariation

How can we quantify them?

```
DENV1.1.2.1.156.1.1
                    CUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGWUCUAACAG
DENV1.1.2.1.21.1.1
                    CUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGKUCUAACAG
                    UUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUCGC---UUAACGUAGUUCUAACAG
                    UUGUUAGU-CUGUCUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAACAG
DENV1.1.2.1.229.1.1
DENV1.1.2.1.339.1.1
                    UUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUUACAG
                    UUGUUAGU-CUACGUGGACU-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAACAG
DENV1 1 2 1 157 1 1
DENV1.1.2.1.168.1.1
                    UUGUUAGUACUACGUGGACC-GACAAGAAUCAGUUUCGAAUCGGAAGCUUGCCGGUUAACGUAGUUCUAACAG
                    GACUCAGA-UACCACGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAACAG
DENV1.1.2.1.402.1.1
#=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

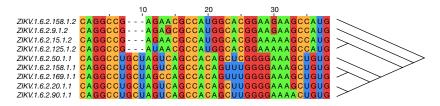
RNAz SVM RNA-class probability

Covariation

- compensatory mutations
 RNAalifold
- statistical significance
 R-scape

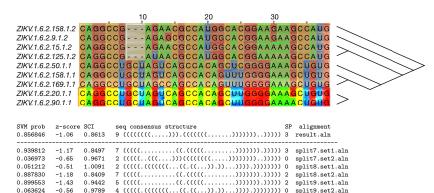
```
10
                        - AGAACGCCAUGGCACGGAAGAAGCCAUG
            CAGGCCG-
                         AGAGCGCCAUGGCACGGAAGAAGCCAUG
ZIKV.1.6.2.9.1.2
            CAGGCCG - - - AGAACGCCAUGGCACGGAAAAGCCAUG
ZIKV.1.6.2.15.1.2
            CAGGCCG - - - AUAACGCCAUGGCACGGAAAAAGCCAUG
ZIKV.1.6.2.125.1.2
            CAGGCCUGCUAGUCAGCCACAGCUCGGGGAAAGCUGUG
ZIKV.1.6.2.50.1.1
ZIKV.1.6.2.158.1.1
            CAGGCCUGCUAGUCAGCCACAGUUUGGGGAAAGCUGUG
            CAGGCCUGCUAGCCAGCCACAGUUUGGGGAAAGCUGUG
ZIKV.1.6.2.169.1.1
ZIKV.1.6.2.20.1.1
            CAGGCCUGCUAGUCAGCCACAGCUUGGGGAAAGCUGUG
            CAGGCCUGCUAGUCAGCCACAGCUUGGGGAAAACUGUG
ZIKV.1.6.2.90.1.1
SVM prob
                                                    SP alignment
       z-score SCI
                   seq consensus structure
0.856846
       -1.06
             0.8613
                                                      result.aln
0.939812
       -1.17
             0.8497
                                                      split7.set1.aln
0.036973
       -0.65
             0.9671
                                                      split7.set2.aln
0.051212
       -0.51
             1.0091
                   split8.set1.aln
0.887830
       -1.18
             0.8409
                                                      split8.set2.aln
                                     0.899553
       -1.43
             0.9442
                                                      split9.set1.aln
                   split9.set2.aln
0.063624
       -0.56
             0.9789
```

. . .



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

. . .

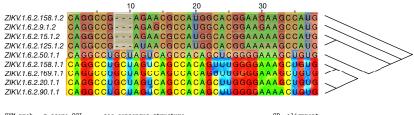


• • •

		10	20	30	
ZIKV.1.6.2.158.1.2	CAGGCCG	AGAACO	CCAUGGCAC	GGAAGAAGCCAU	G 🔍
ZIKV.1.6.2.9.1.2	CAGGCCG	AGAGC	CCAUGGCAC	GGAAGAAGCCAU	G
ZIKV.1.6.2.15.1.2	CAGGCCG	AGAACO	CCAUGGCAC	GGAAAAAGCCAU	G \
				GGAAAAAGCCAU	
				GGGGAAAGCUGU	
				J <mark>GGGG</mark> AAA <mark>GCU</mark> GU	
				J <mark>GGGG</mark> AAA <mark>GCU</mark> GU	
				J G G G G A A G C U G U	
ZIKV.1.6.2.90.1.1	CAGGCCU	GCUAGUCA	CCACAGCUL	JGGGGAAAACUGU	G /

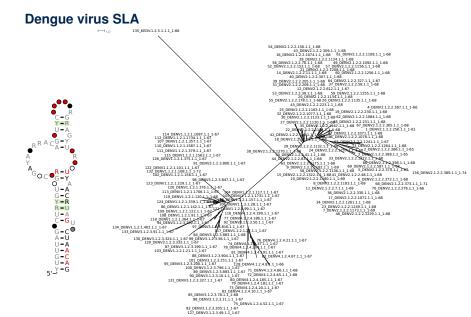
SVM prob z-score SCI 0.856846 -1.06 0.8		ensus structure (())).((((((())))))))		alignment result.aln
0.939812 -1.17 0.8 0.036973 -0.65 0.9 0.051212 -0.51 1.0 0.887830 -1.18 0.8 0.899553 -1.43 0.9 0.063624 -0.56 0.9	71 2 (((((. 91 2 ((((. 09 7 ((((. 42 5 ((((.	.((()))((((((.(((.(((.((()))))))))))))))))))))))))))))))))))) 0)))) 0)))) 2)))) 2	split7.set1.aln split7.set2.aln split8.set1.aln split8.set2.aln split9.set1.aln split9.set2.aln

• • •

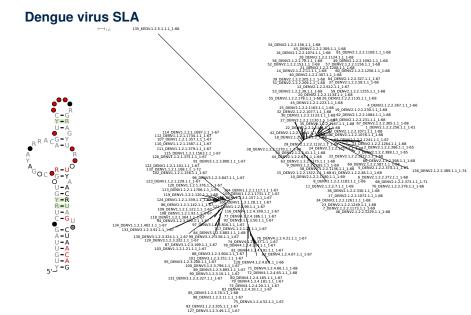


```
SVM prob
         z-score SCI
                       seq consensus structure
                                                                  alignment
0.856846
         -1.06
                0.8613
0.939812
                0.8497
                                                                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

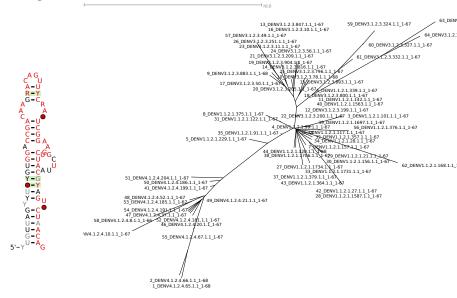
• •



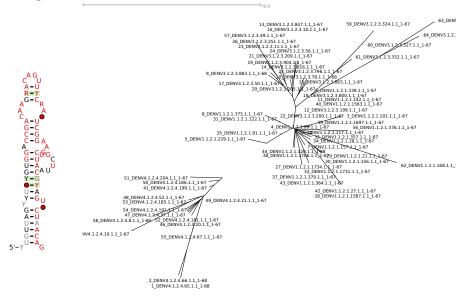
NJ tree; 146 pairwise different DENVG SLA sequences



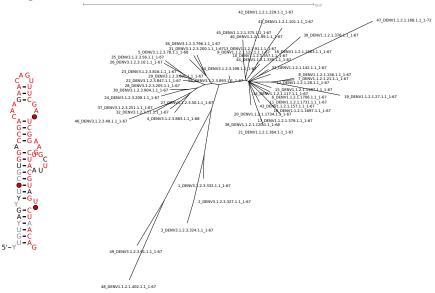
NJ tree; 146 pairwise different DENVG SLA sequences



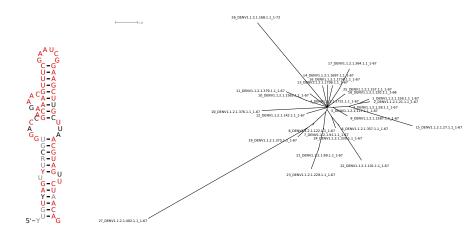
NJ tree; 64 pairwise different DENV1,3,4 SLA sequences



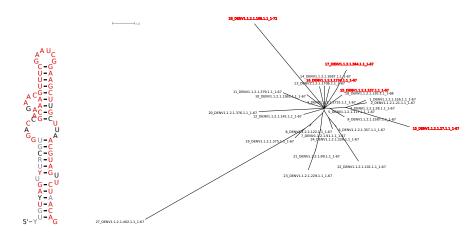
NJ tree; 64 pairwise different DENV1,3,4 SLA sequences



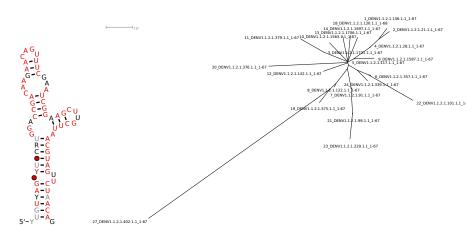
NJ tree; 49 pairwise different DENV1,3 SLA sequences



NJ tree; 27 pairwise different DENV1 SLA sequences



NJ tree; 27 pairwise different DENV1 SLA sequences



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 consructing covariance models
- unversally 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)



Thank you!