

Characterization of conserved Flavivirus 5'UTR elements

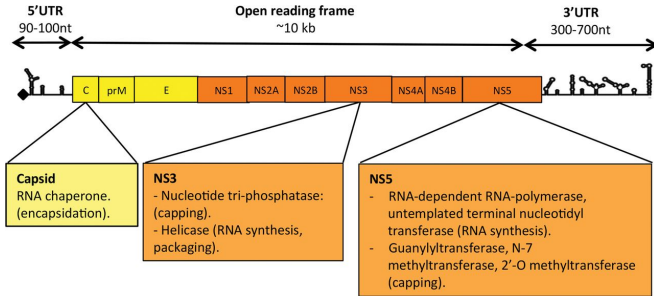
Michael T. Wolfinger
Andrea Tanzer
Roman Ochsenreiter

Department of Theoretical Chemistry
University of Vienna

Bled, Slovenia, February 16, 2017

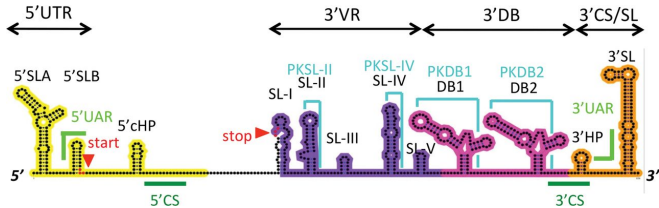
Flavivirus genome characteristics

- 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

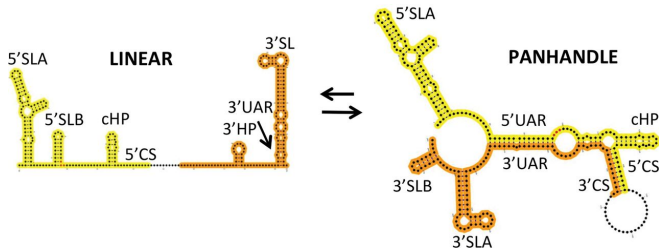


Flavivirus replication

- 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

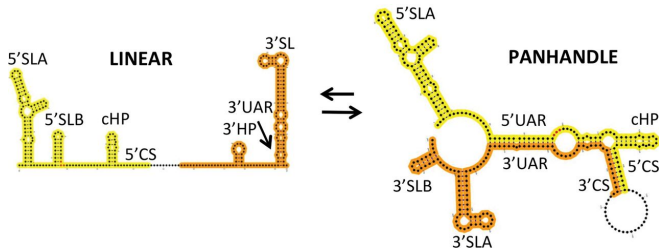
Flavivirus replication

- 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

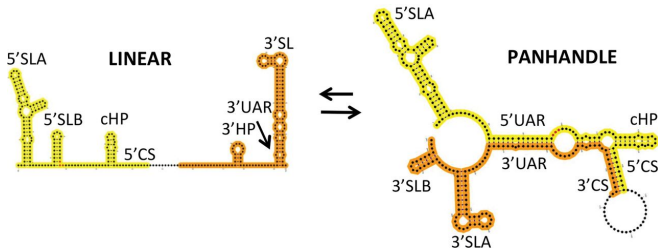
- FV RNA replication happens via asymmetric replication cycle
- Viral RNA-dependent RNA polymerase (RdRp) synthesizes a full-length minus-strand (-)gRNA
- (-)gRNA is a template for several rounds of (+)gRNA synthesis
- Ratio (-)gRNA / (+)gRNA between 1:10 and 1:100
- Genome cyclisation is required for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



- SLA acts as promoter / enhancer for viral RdRp

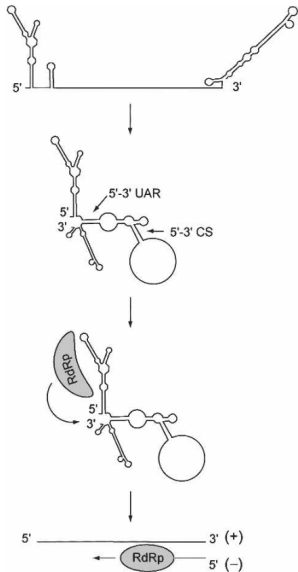
Flavivirus replication

- 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

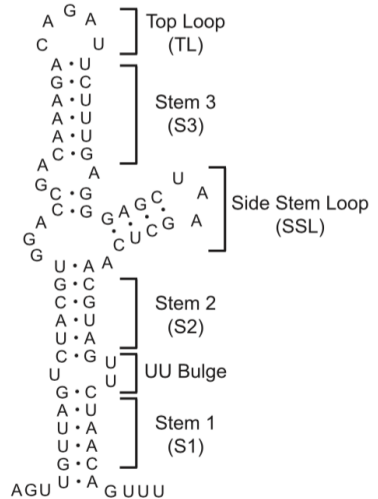
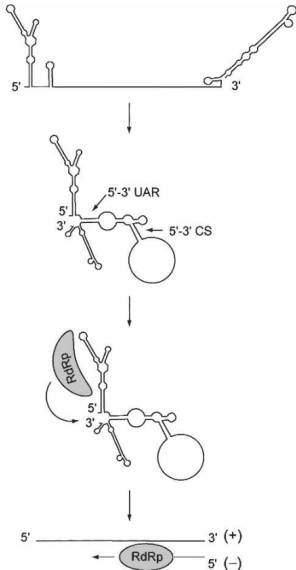


- SLA acts as promoter / enhancer for viral RdRp
- Viral RdRps are **error-prone**: Misincorporations every 40-70kb

Flavivirus replication



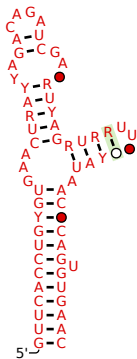
Flavivirus replication



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)



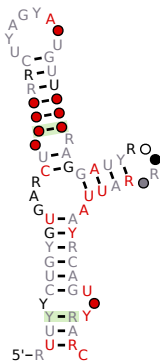
KOKV.1.4.1.1.1.1.1
NMV.1.4.Z.1.1.1
#=GC SS_cons

GUUCCACUGUGUAACUAACCAGACAGAUCGAAGUUAGGUGAUUA-CAUAACACAGUGUGAAC
GUUCCACUGCGUGAACUGAUUAGACAGAUCGACAUCAGAUAGUUUCUAUAACUCAGUGUGAAC
(((((((((((((((.....((((((..((.....)).....))))))((((.....))))..))))))..))))))

SLA in different FV groups

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

Japanese encephalitis virus group (JEVG)

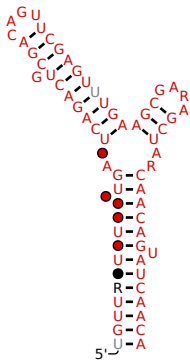


```
JEV.1.3.1.1.1.1.1 G-UUU--AUCUGUGUGA---ACUUCUUGGCUUAGUAUCGU-UGAGAAGAAUC-AGA-GAUUAGU-GCAGUUUAAAC-
JEV.1.3.1.163.1.1 G-UUU--AUCUGUGUGA---ACUUCUUGGCUUAGUAUCGU-UGAGAAGAAUCG-AGA-GAUUAGU-GCAGUUUAAAC-
JEV.1.3.1.130.1.1 G-UUU--AUCUGUGUGA---ACUUCUUGGCUUAGUAUCGU-UGAGAAGAAUCG-AGA-GAUUAGU-GCAGUUUAAAC-
ALFV.1.3.5.1.1.1.1 G-UUU--AUCUGUGUGA---ACUUUUUGACUCAGGAUUGU-UGGAAGGGAUUG-AAA-GAUUAAU-ACAGUUUAAAC-
JEV.1.3.1.165.1.1 G-UUU--AUCUGUGUGA---ACUUCUUGGUUUAGUAUCGU-UGAGAAGAAUCG-AAA-GAUUAGU-GCAGUUUAAAC-
JEV.1.3.1.75.1.1 G-UUUU--UAACGUGUGA---ACUUCUUGGCUUAGUAUCGU-CGAGAGAAUCG-AGA-GAUUAGU-GCAGUUUAAAC-
JEV.1.3.1.82.1.1 G-UUUU--UUUUUGUGU---ACUUCUGGCUUAGUAUCGU-UGAGAAGAAUCG-AGA-GAUUAGU-GCAGUUUAAAC-
JEV.1.3.1.84.1.1 G-UUUU--AACUGUGUG---ACUUCUUGGCUUAGUAUCGU-UGAGAAGAAUCG-AGA-GAUUAGU-GCAGUUUAAAC-
USUV.1.3.7.89.1.1 G-UUG--GCCUGUGUGA---GCUUCUACUUCUAGUAUUGU-UUUGGAGGAUCG-UUA-GAUUAAU-ACAGUUGUUCAG-
MVEV.1.3.6.2.1.1 G-UUC--AUCUGCGUGA---GCUUCGGAUCUCAGUAUUGU-UUGGAAGGAUCA--UU-GAUUAAAC-GCGGUUUGAAC-
MVEV.1.3.6.1.1.1 G-UUC--AUCUGCGUGA---GCUUCGGAUCUCAGUAUUGU-UUGGAAGGAUCA--UU-GAUUAAAC-GUGGUUUGAAC-
MVEV.1.3.6.4.1.1 G-UUC--AUCUGCGUGA---GCUUCGGAUCUCAGUAUUGU-UUGGAAGGAUCA--UUU-GAUUAAAC-GCGGUUUGAAC-
MVEV.1.3.6.9.1.1 G-UUC--AUCUGCGUGA---GCUUCGGAUCUCAGUAUUGU-UUGGAAGGAUCAUUUUU-GAUUAAAC-GCGGUUUGAAC-
WNV.1.3.2.1311.1.1 GUUUC--GCCUGUGUGA---GCUGACAAACUUGUAUGUGU-UUGUGAGGGAUUAACAAACAUUUAAC-ACAGUCCGAGC-
WNV.1.3.2.1370.1.1 GUUCU--AGCUGUGUGA---GCUGACAAACUUGUAUGUGU-UUGUGAGGGAUUAACAAACAUUUAAC-ACAGUCCGAGC-
WNV.1.3.2.1371.1.1 GUUUU--GACCCUGUGUGA---GCUGACAAACUUGUAUGUGU-UUGUGAGGGAUUAACAAACAUUUAAC-ACAGUCCGAGC-
WNV.1.3.2.1479.1.1 --GUCG--CCUGUGUGA---GCUGACAAACUUGUAUGUGU-UUGUGAGGGAUUAACAAACAUUUAAC-ACAGUCCGAGC-
WNV.1.3.2.973.1.1 G-UUCG--CCUGUGUGA---ACUGACAAACUUGUAUGUGU-UUGUGAGGGAUUAACAAACAUUUAAC-ACAGUCCGAGC-
KUNV.1.3.4.43.1.1 G-UUCG--CCUGUGUGA---GCUGACAAACUUGUAUGUGU-UUGUGAGGGAUUAACAAACAUUUAAC-ACAGUCCGAGC-
ACAA--CCUGUGUGA---GCUGACAAACUUGUAUGUGU-UUGUGAGGGAUUAACAAACAUUUAAC-ACAGUCCGAGC-
WNV.1.3.2.1683.1.1 --UUCG--CCUGUGUGA---GCUGACAAACUUGUAUGUGU-UUGUGAGGGAUUAACGAGGGAUUAU-ACAGUACGAGC-
WNV.1.3.2.1408.1.1 G-----CCUGUGUGA---GCUGACAAACUUGUAUGUGU-UUGUGAGGGAUUAACAAACAUUUAAC-ACAGUCCGAGC-
USUV.1.3.7.10.1.1 GUUCG--UUCUGUGUGA---GCUUCUACUUCUAGUAUUGUUUUUUGGAGGAUCGUGA--GAUUUAAAC-ACAGUCCGGC-
GUUCG--UUCUGUGUGA---GCUUCUACUUCUAGUAUUGUUUUUUGGAGGAUCGUGA--GAUUUAAAC-ACAGUCCGGC-
GUUCG--UUCUGUGUGA---GCUUCUACUUCUAGUAUUGUUUUUUGGAGGAUCGUGA--GAUUUAAAC-ACAGUCCGGC-
GUUCG--UUCUGUGUGA---GCUUCUACUUCUAGUAUUGUUUUUUGGAGGAUCGUGA--GAUUUAAAC-ACAGUCCGGC-
GUUCG--UUCUGUGUGA---GCUUCUACUUCUAGUAUUGUUUUUUGGAGGAUCGUGA--GAUUUAAAC-ACAGUCCGGC-
SLEV.1.3.7.9.1.1 UGUUCG--GUCGUGUGA---GCCAAGAGGAACAGAUUUC-UUUUUUUGGAGGAUUAACAAACUUAACUUGAUGCGAAC
SLEV.1.3.3.30.1.1 UGUUCG--CGUUGUGA---GCGAGGAAGAAACAGAUUUC-UUUUUUUGGAGGAUUAACAAACUUAACUUGAUGCGAAC
WNV.1.3.2.1616.1.1 UUCG--UCAGAAUC--CUGAAAGAAUUCUCAGUGU-UUGUGAGGGAUUAACAAACAUUUAAC-ACAGUCCGAGC-
#GC SS_cons ..(((.....(((((((.....(((.....(((.....(((.....(((.....(((.....(((.....(((.....)))))))))))))
```

SLA in different FV groups

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

Spondweni virus group (SPOVG)

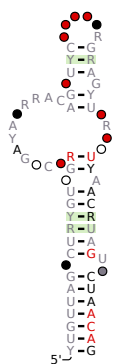


```
ZIKV.1.6.2.21.1.1 UGUUGAUCUGUGUGAGUCAGACUGCGACAGUUCGAGUUUGAAGCGAAAGCUAGCAACAGUAUCAACA
ZIKV.1.6.2.161.1.1 UGUUGAUCUGUGUGAAUCAGACUGCGACAGUUCGAGUUUGAAGCGAAAGCUAGCAACAGUAUCAACA
ZIKV.1.6.2.158.1.1 UGUUGAUCUGUGUGAGUCAGACUGCGACAGUUCGAGUCUGAAGCGAGAGCUAAACAACAGUAUCAACA
ZIKV.1.6.2.169.1.1 UGUUGAUCUGUGUGAAUCAGACUGCGACAGUUCGAGUUUGAAGCGAGAGCUAAACAACAGUAUCAACA
ZIKV.1.6.2.84.1.1 AGUUGAUCUGUGUGAAUCAGACUGCGACAGUUCGAGUUUGAAGCGAAAGCUAGCAACAGUAUCAACA
ZIKV.1.6.2.150.1.1 UGUU--UCUGUGUGAAUCAGACUGCGACAGUUCGAGUUUGAAGCGAAAGCUAGCAACAGUAUCAACA
ZIKV.1.6.2.141.1.1 UGUUACUGUUGCUGACUCAGACUGCGACAGUUCGAGUUUGAAGCGAAAGCUAGCAACAGUAUCAACA
#=GC SS_cons ((((((((((((((.....(((((((.(CCC...)))))))))))))(((((...)))..)))))..)))))
```

SLA in different FV groups

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

Dengue virus group (DENVG)



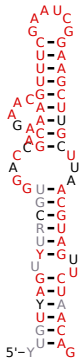
```
DENV4.1.2.4.15.1.1 CAG--UUAGAA-CUGGUUAAAC--AAGCUCC--GAUUUGGAACUGCCUUGCGG-----UCCACACAGA-CUAAAGUU
DENV2.1.2.2.51.1.1 CUG--UJAG-U-CUACGUUGAGCUUAGCUCC--UCAAGAAGAUUCUGUUCUUGCGG-----UCCACGUAGA-CUAAAGUU
DENV2.1.2.2.52.1.1 CUG--UJAGAA-CUACGUUGAGCUUAGCUCC--UCAAGAAGAUUCUG-UUCUUJAGGGAGGCUAAGC--UCCACGUAGUUCUAAACAG--
DENV4.1.2.4.11.1.1 CUG--UJAGAA-CUGGUUAAAG-----CA-AGCUCCGAUU-UUGAAGCUUGG-----UCCACACAGUUCUAAACAG--
DENV1.1.2.1.237.1.1 GCG-----CUUCGGAUCUGGACCGACAAGAA-CAGUUUCGAUUCGGAAGCUUGC-----UUAAACGUAGUUCUAAACAG--
DENV1.1.2.1.234.1.1 AUG-----UCC-----UUGGACCGACAAGAA-CAGUUUCGAUUCGGAAGCUUGC-----UUAAACGUAGUUCUAAACAG--
DENV1.1.2.1.311.1.1 UUGU--AAACCGAGACCCAGUGCCGACAAGAA-CAGUUUCGAUUCGGAAGCUUGC-----UUAAACGUAGUUCUAAACAG--
DENV4.1.2.4.66.1.1 GUG--UUUGUG-UUGUGUGGACC--GACAAGGA-CAGUUCCAAUUCGGAAGCUUGC-----UUAAACACAGUUCUAAACAG--
DENV4.1.2.4.65.1.1 UUG--UUJAGG-UUGUGUGGACC--GACAAGGA-CAGUUCCAAUUCGGAAGCUUGC-----UUAAACACAGUUCUAAACAG--
DENV3.1.2.3.251.1.1 UUG--UUJAG-U-UUUCUGGACC--GACAAGAA-CAGUUUCGAUUCGGAAGCUUGC-----UUAAACGUAGUUCUAAACAG--
DENV3.1.2.3.199.1.1 UUG--UUJAG-U-CUACGUGGACC--GACAAGAA-CAGUUUCGAUUCGGAAGCUUGC-----UUAAACGUAGUUCUAAACAG--
DENV3.1.2.3.800.1.1 UUG--UUJAG-U-CUACGUGGACC--GACAAGAA-CAGUUUCGACUUCGGAAGCUUGC-----UUAAACGUAGUUCUAAACAG--
DENV3.1.2.3.11.1.1 UUG--UUJAG-U-CUACGUGGACC--GACAAGAA-CAGUUUCGAUUCGGAAGCUUGC-----UUAAACGUAGUUCUAAACAG--
DENV3.1.2.3.56.1.1 UUG--UUJAG-U-CUACGUGGACC--GACAAGAA-CAGUUUCGACUUCGGAAGUUGC-----UUAAACGUAGUUCUAAACAG--
DENV3.1.2.3.796.1.1 UUG--UUJAG-U-CUACGUGGACC--GACAAGAA-CAGUUUCGACUUCGGAAGCUUGC-----UUAAACGUAGUUCUAAACAG--
DENV1.1.2.1.117.1.1 UUG--UUJAG-U-CUACGUGGACC--GACAAGAA-CAGUUUCGAUUCGGAAGCUUGC-----UUAAACGUAGUUCUAAACAG--
DENV1.1.2.1.1731.1.1 UUG--UUJAG-U-CUACGUGGACC--GACAAGAA-CAGUUUCGAUUCGGAAGCUUGC-----UUAAACGUAGUUCUAAACAG--
DENV4.1.2.4.185.1.1 UUG--UUJAG-U-CUGUGUGGACC--GACAAGGA-CAGUUCCAAUUCGGAAGCUUGC-----UUAAACACAGUUCUAAACAG--
DENV4.1.2.4.67.1.1 UUG--UUJAG-G-UUGUGUGGACC--GACAAGGA-CAGUUCCAAUUCGGAAGCUUGC-----UUAAACACAGUUCUAAACAG--
DENV1.1.2.1.376.1.1 UUG--UUAG-U-CUACGUGGGCC--GACAAGAA-CAGUUUCGAUUCGGAAGCUUGC-----UUAAACGUAGUUCUAAACAG--
DENV1.1.2.1.402.1.1 GAC--UJAG-A-UACACAGGACC--GACAAGAA-CAGUUUCGAUUCGGAAGCUUGC-----UUAAACGUAGUUCUAAACAG--
DENV2.1.2.2.386.1.1 UUG--UUAG-U-CUACGUGGACC--GACAAGAA-CAGAUUCUUGAGGGAGCUAAGC-----UCAAACGUAGUUCUAAACAG--
DENV2.1.2.2.367.1.1 UUC-----U-C-CUACGUUGGACCGACAAGA-CAGAUUCUUGAGGGAGCUAAGC-----UCAAACGUAGUUCUAAACAG--
DENV2.1.2.2.256.1.1 UUG--UUAG-U-CUACGUGGACC--GACAAGAA-CAGAUUCUUGAGGGAGCUAAGC-----UCAAACGUAGUUCUAAACAG--
DENV2.1.2.2.365.1.1 CAG--AGAC-A-CUACGUGGACC--GACAAGAA-CAGAUUCUUGAGGGAGCUAAGC-----UCAAACGUAGUUCUAAACAG--
DENV2.1.2.2.372.1.1 GUCC--UJAG-C-UAAACGUGGACC--GACAAGAA-CAGAUUCUUGAGGGAGCUAAGC-----UCAAACGUAGUUCUAAACAG--
DENV2.1.2.2.276.1.1 GGG--UJAG--UUACGUGGACC--GACAAGAA-CAGA-UUCUUGAGGGAGCUAAGC-----UCAAACGUAGUUCUAAACAG--
KEDV.1.2.5.1.1.1 UJ--UJAG-U-CG-GUGUGAUUGAGUACAAA-CAGAUUCUUGAGGGAGCUAULUC-----UCAAACUUCGUGUCAAACAG--
DENV2.1.2.2.389.1.1 GACUCAGACUCUCUACUGUGGACC--GACAAGAA-CAGAUUCUUGAGGGAGCUAAGC-----UCAAACGUAGUUCUAAACAG--
DENV1.1.2.1.29.1.1 CUG--UUJAG-----AACAGUUUCGAUUCGGAAGCUUGC-----UUAAACGUAGUUCUAAACAG--
#GC SS_cons (((((((((.....(((((((.....(((((.....)))))))))))))))))))))
```



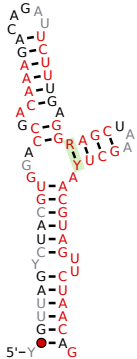
Dengue virus SLA

There are four DENV serotypes, probably five ...

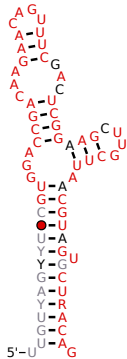
DENV1



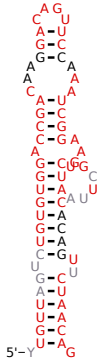
DENV2



DENV3



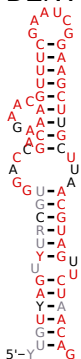
DENV4



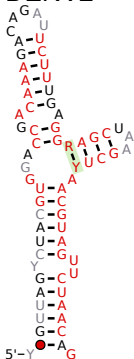
Dengue virus SLA

There are four DENV serotypes, probably five ...

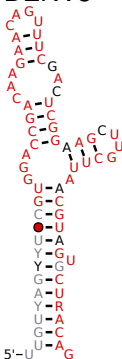
DENV1



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-CAGUUUCGAAUCGGAAGCUUGC---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---UUAACGUAGUUCUAACAG
DENV1.1.2.1.402.1.1 GACUCAGA-UACCACGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGUUCUAACAG
#=GC SS_cons .(((((((.((((((((.(.(((...(((.....)))...)))...)).....)))))..)))))
```

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-CAGUUUCGAAUCGGAAGCUUGC---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---UUAACGUAGUUCUAACAG
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

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-CAGUUUCGAAUCGGAAGCUUGC---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---UUAACGUAGUUCUAACAG
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

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-CAGUUUCGAAUCGGAAGCUUGC---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---UUAACGUAGUUCUAACAG
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---UUAACGUAGWUCUAACAG
DENV1.1.2.1.21.1.1 CUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---UUAACGUAGKUCUAACAG
DENV1.1.2.1.1731.1.1 UUGUUAGU-CUACGUGGACC-GACAAGAA-CAGUUUCGAAUCGGAAGCUUGC---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---UUAACGUAGUUCUAACAG
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

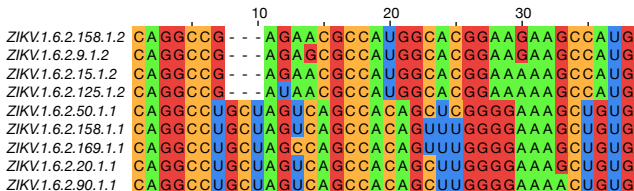
Structure conservation

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

Covariation

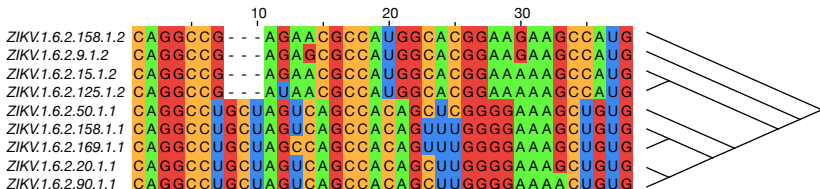
- compensatory mutations
- RNAalifold
- statistical significance
- R-scape

RNAaliSplit workflow



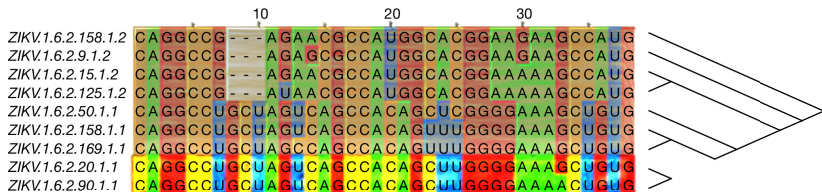
SVM prob	z-score	SCI	seq consensus structure	SP alignment
0.856846	-1.06	0.8613	9 (((((((((.....))).((((((((.....))))))))))..)))))	3 result.aln
0.939812	-1.17	0.8497	7 (((((((.....(((((.....))))))))..)))))	3 split7.set1.aln
0.036973	-0.65	0.9671	2 ((((((.....))((((((((.....))))))))..)))))	0 split7.set2.aln
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



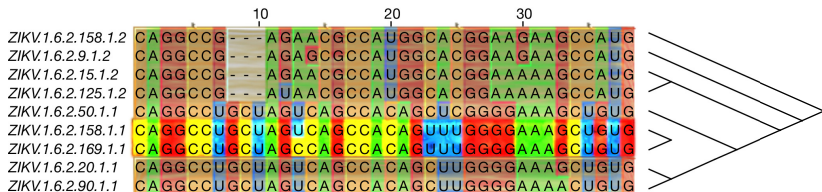
SVM prob	z-score	SCI	seq consensus structure	SP	alignment
0.856846	-1.06	0.8613	9 (((((((((.....))).(((((((.....))))))))..)))))) 3	result.aln	
0.939812	-1.17	0.8497	7 (((((((.....(((.....))))))))..)))))) 3	split7.set1.aln	
0.036973	-0.65	0.9671	2 ((((((.....))(((((((.....))))))))..)))))) 0	split7.set2.aln	
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



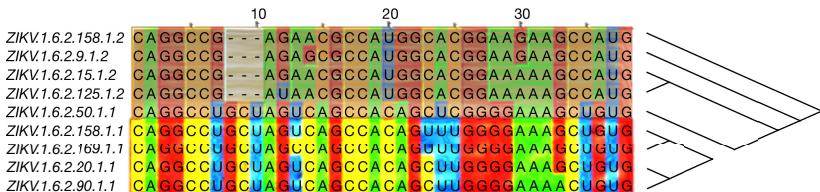
SVM prob	z-score	SCI	seq consensus	structure	SP	alignment
0.856846	-1.06	0.8613	9	((((((((.....))).(((((((.....)))))))).))))))	3	result.aln
0.939812	-1.17	0.8497	7	(((((.....(((.....)))))))).))))))	3	split7.set1.aln
0.036973	-0.65	0.9671	2	(((((.....(((.....)))))))).))))))	0	split7.set2.aln
0.051212	-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



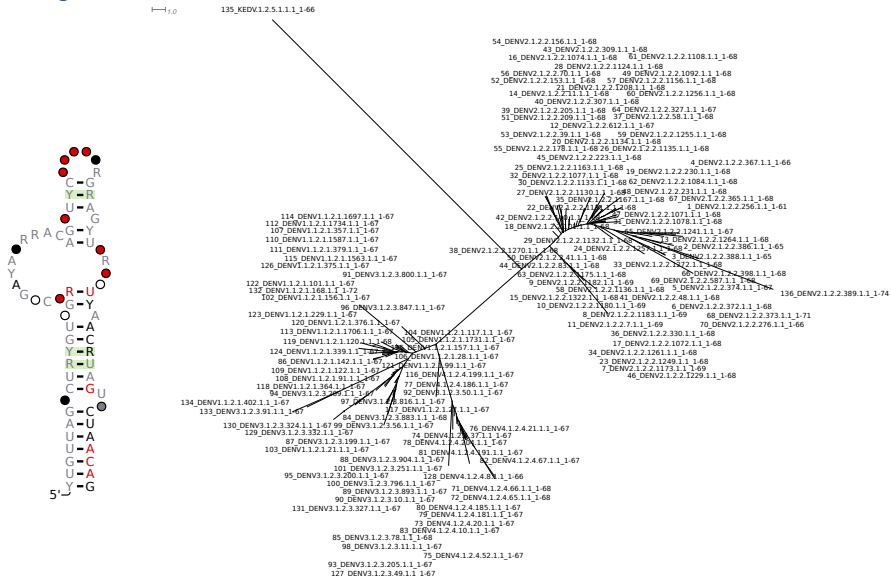
SVM prob	z-score	SCI	seq consensus	structure	SP	alignment
0.856846	-1.06	0.8613	9	((((((((.....))).(((((((.....)))))))).))))))	3	result.aln
0.939812	-1.17	0.8497	7	(((((.....(((.....)))))))).))))))	3	split7.set1.aln
0.036973	-0.65	0.9671	2	(((((.....(((.....)))))))).))))))	0	split7.set2.aln
0.051212	-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



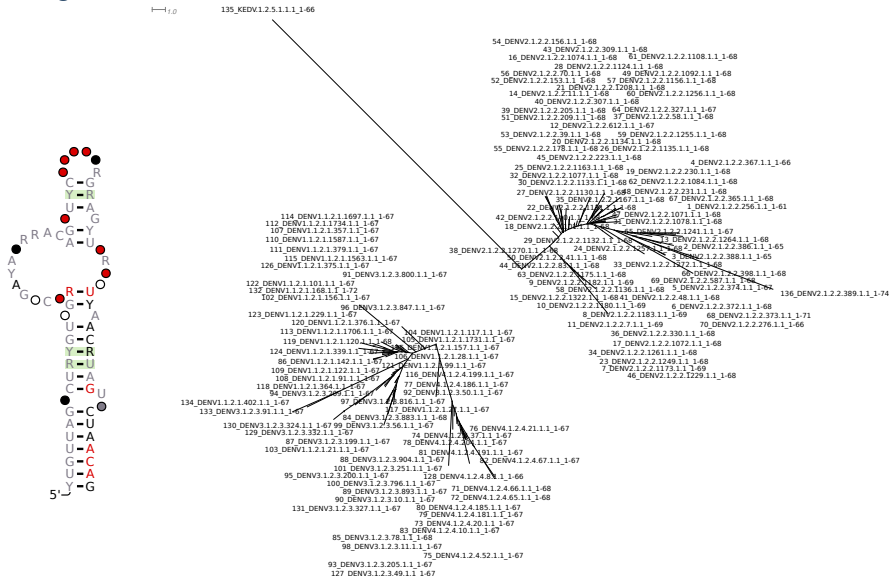
SVM prob	z-score	SCI	seq consensus	structure	SP	alignment
0.856846	-1.06	0.8613	9	((((((((.....))).(((((((.....)))))))).))))))	3	result.aln
0.939812	-1.17	0.8497	7	(((((.....(((.....)))))))).))))))	3	split7.set1.aln
0.036973	-0.65	0.9671	2	(((((.....(((.....)))))))).))))))	0	split7.set2.aln
0.051212	-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



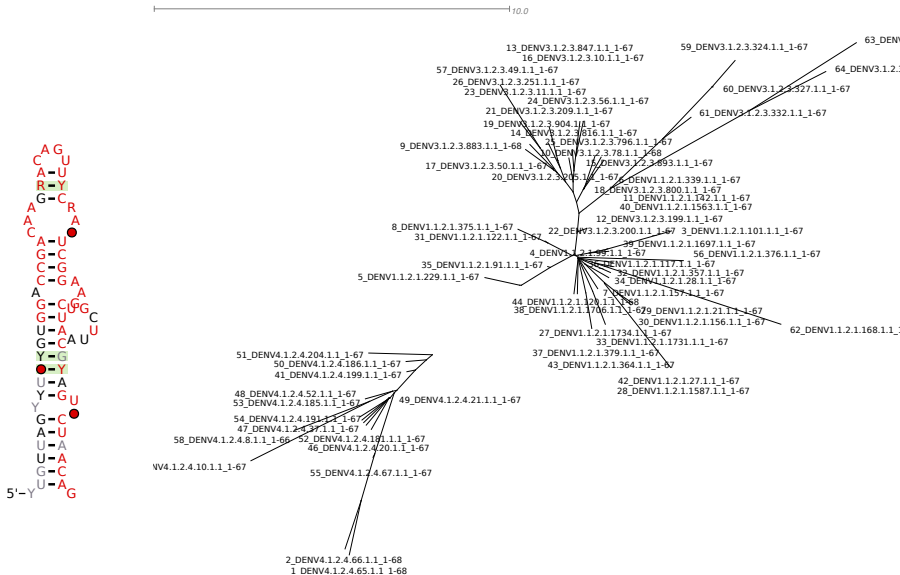
NJ tree; 146 pairwise different DENVG SLA sequences

Dengue virus SLA



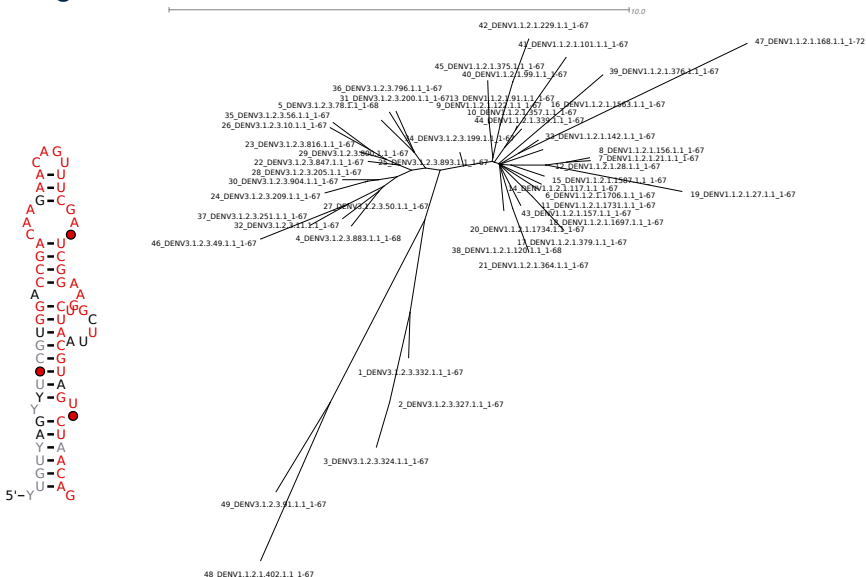
NJ tree; 146 pairwise different DENVG SLA sequences

Dengue virus SLA



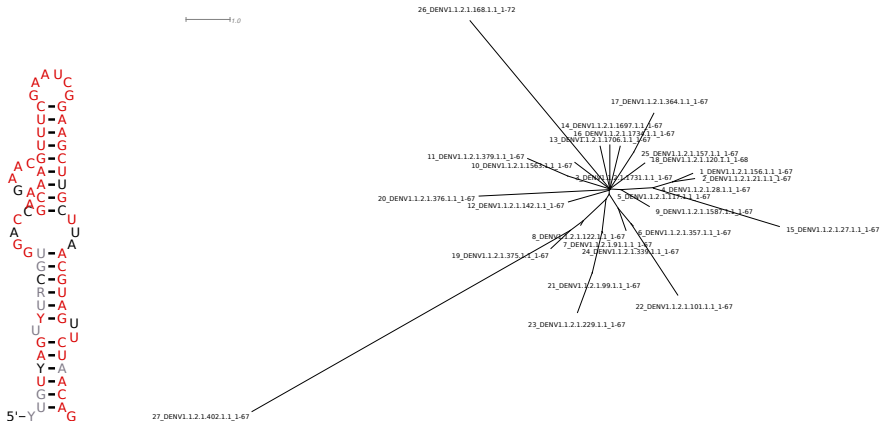
NJ tree; 64 pairwise different DENV1,3,4 SLA sequences

Dengue virus SLA



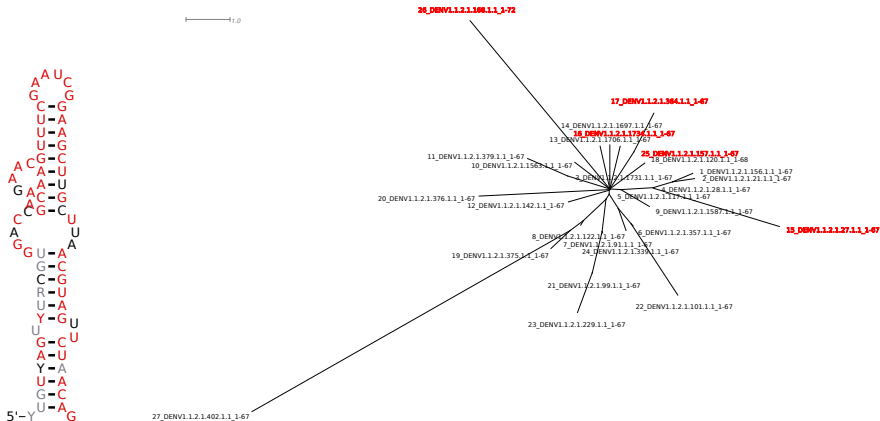
NJ tree; 49 pairwise different DENV1,3 SLA sequences

Dengue virus SLA



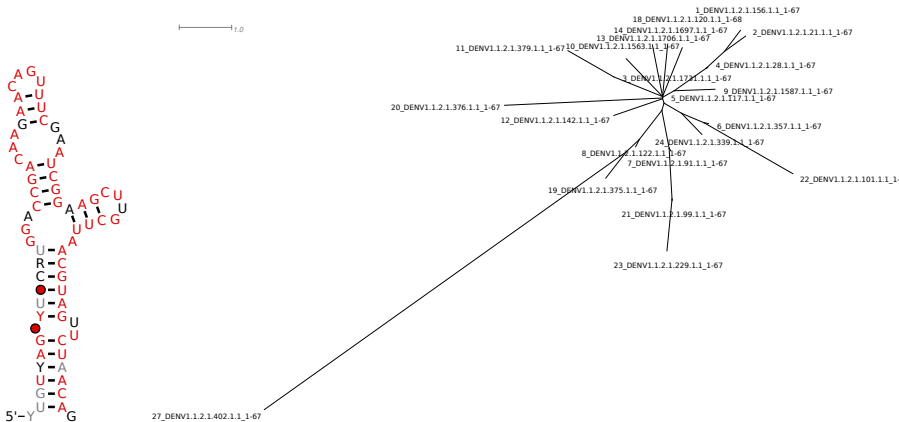
NJ tree; 27 pairwise different DENV1 SLA sequences

Dengue virus SLA



NJ tree; 27 pairwise different DENV1 SLA sequences

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
- 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)

FWF

Der Wissenschaftsfonds.

SFB RNA regulation of the
transcriptome (FWF-F43)



universität
wien



MEDIZINISCHE
UNIVERSITÄT
WIEN

Thank you!