The quest for conserved RNAs in viral genomes

Michael T. Wolfinger

Research Group Bioinformatics and Computational Biology
& Department of Theoretical Chemistry
University of Vienna
Austria

universität wien

University of Kent
15 October 2019

The World's Deadliest Animals

Number of people killed by animals, 2015

gates notes

830,000
Mosquito

580,000
Human

- 60,000 Snake
- 24,200 Sandfly
- 17,400 Dog
- 8,000 Kissing bug
- 4,400 Freshwater snail
- 3,500 Scorpion
- 3,500 Tsetse fly
- 1,600 Tapeworm
- 500 Hippopotamus
- 100 Elephant
- 60 Bee
- 40 Jellyfish
- 6 Shark

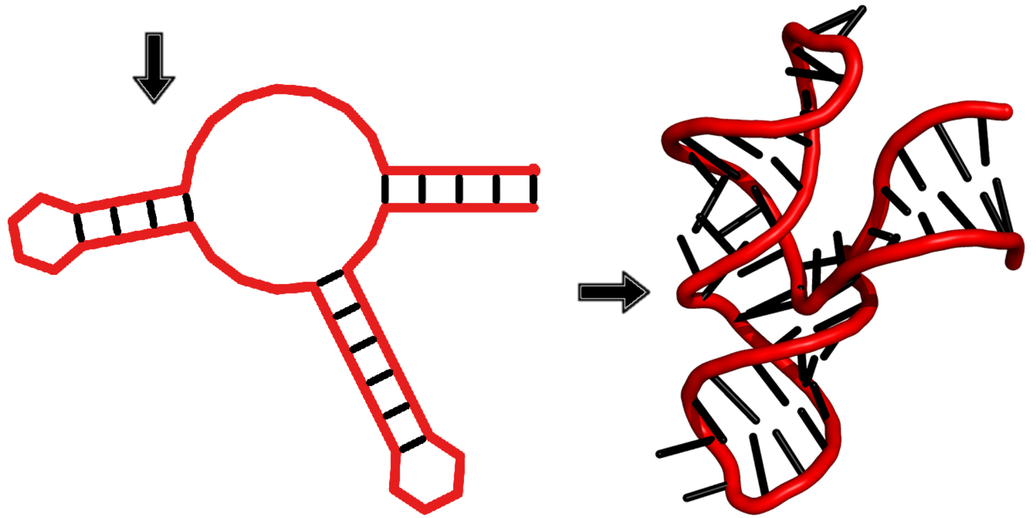
Sources: IHME, WHO, CrocBITE, FAO, Norwegian Institute for Nature Research, International Shark Attack File, National Geographic, PBS, National Science Foundation, CDC, WWF, *Wilderness & Environmental Medicine*, *Nature*, French Institute of Research for Development. All calculations have wide error margins.

Part I:

RNA structure prediction

The RNA Folding Problem

GCGCUCUGAUGAGGCCGCAAGGCCGAAACUGCCGCAAGGCAGUCAGCGC

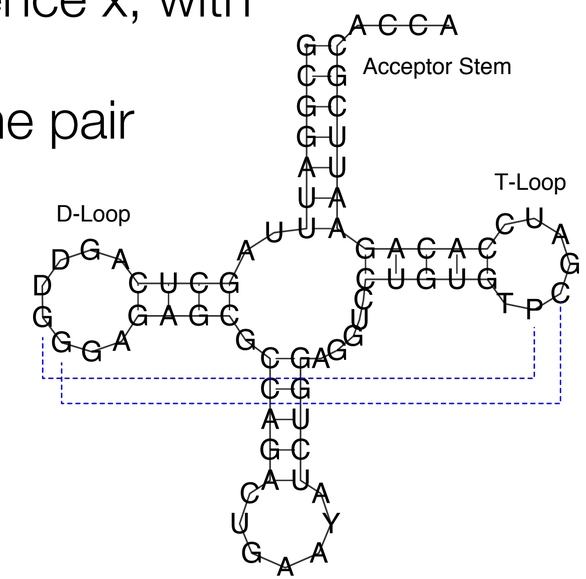


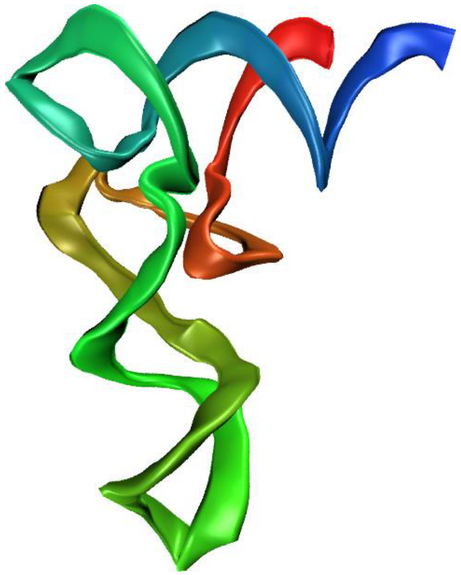
- Hierarchical folding: Secondary structure forms first then helices arrange to form tertiary structure
- Secondary structures cover most most of the folding energy
- Convenient and biologically useful description
- Computationally easy to handle
- Tertiary structure prediction needs knowledge of secondary structure

Secondary Structures

A secondary structure is a list of base pairs (i,j) on a sequence x, with

- Any nucleotide (sequence position) can form at most one pair
- No pseudo-knots: No pairs (i,j) and (k,l) with $i < k < j < l$
- If (i,j) is a pair then $x_i x_j \in \{GC,CG,AU,UA,GU,UG\}$
- If (i,j) is a base pair, then $j - i > 3$

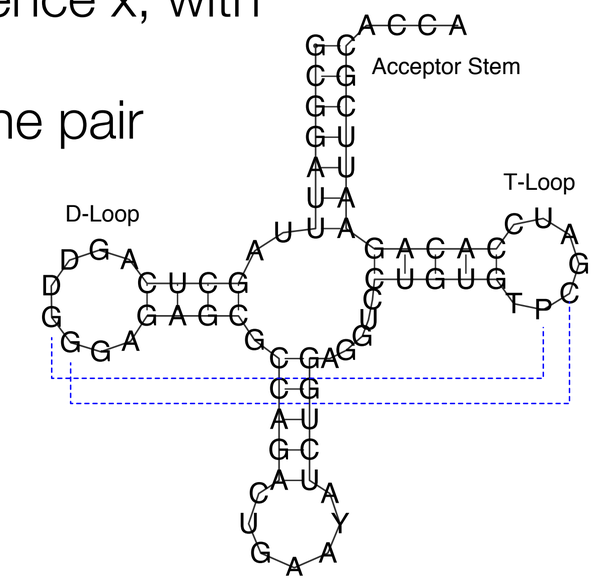


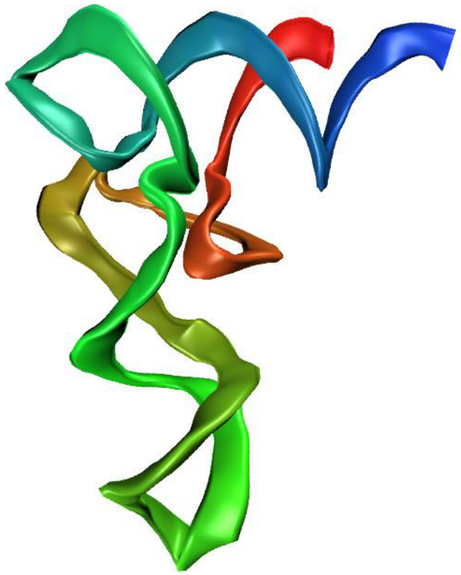


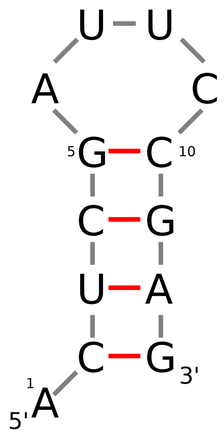
Secondary Structures

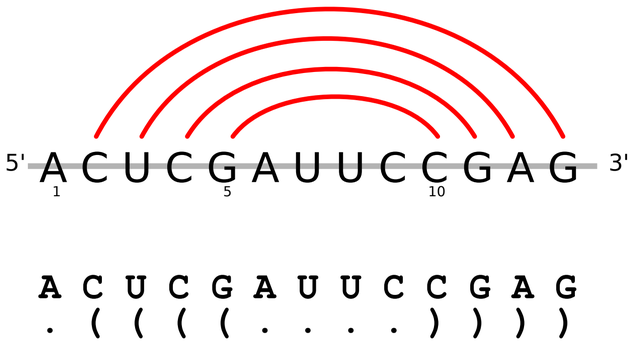
A secondary structure is a list of base pairs (i,j) on a sequence x, with

- Any nucleotide (sequence position) can form at most one pair
- No pseudo-knots: No pairs (i,j) and (k,l) with $i < k < j < l$
- If (i,j) is a pair then $x_i x_j \in \{GC,CG,AU,UA,GU,UG\}$
- If (i,j) is a base pair, then $j - i > 3$





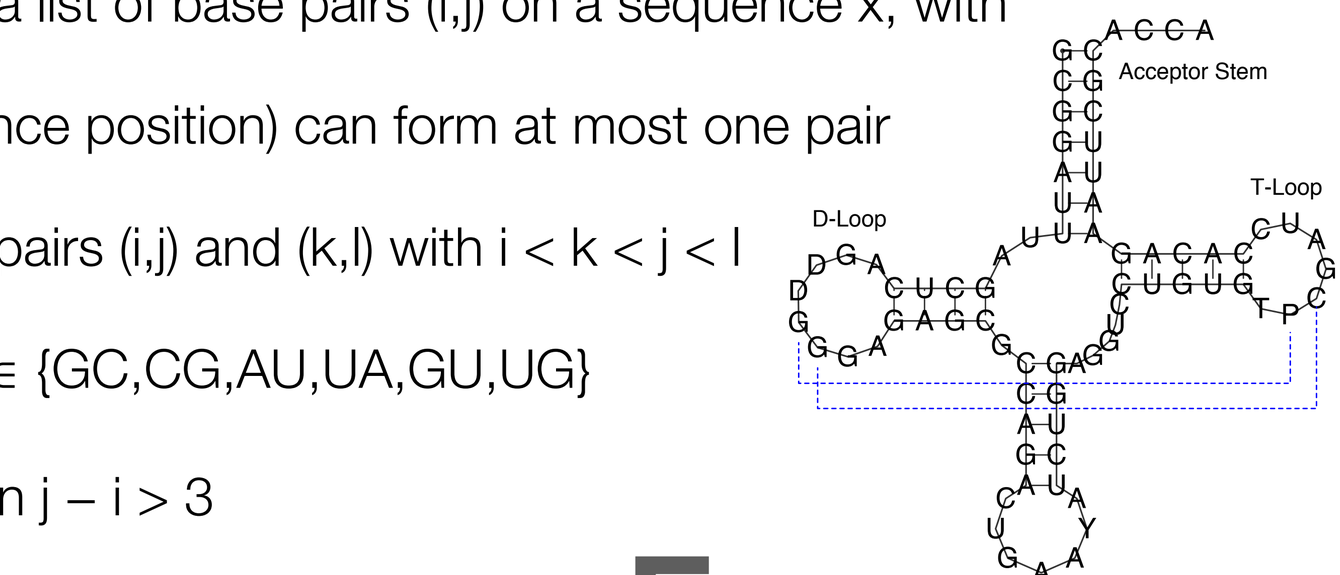


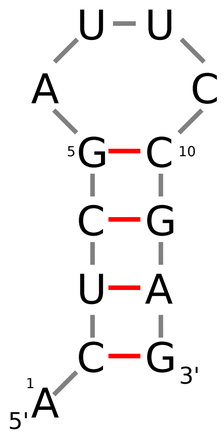


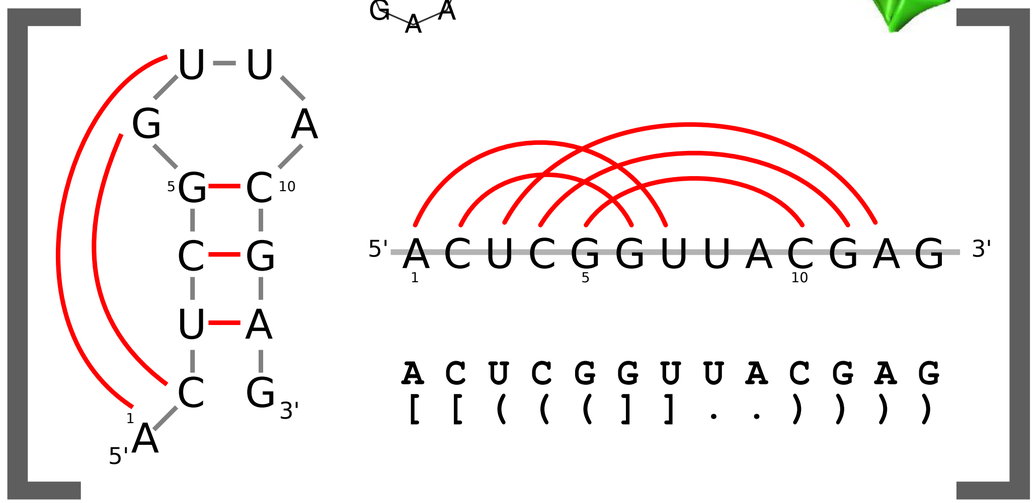
Secondary Structures

A secondary structure is a list of base pairs (i,j) on a sequence x, with

- Any nucleotide (sequence position) can form at most one pair
- No pseudo-knots: No pairs (i,j) and (k,l) with $i < k < j < l$
- If (i,j) is a pair then $x_i x_j \in \{GC,CG,AU,UA,GU,UG\}$
- If (i,j) is a base pair, then $j - i > 3$

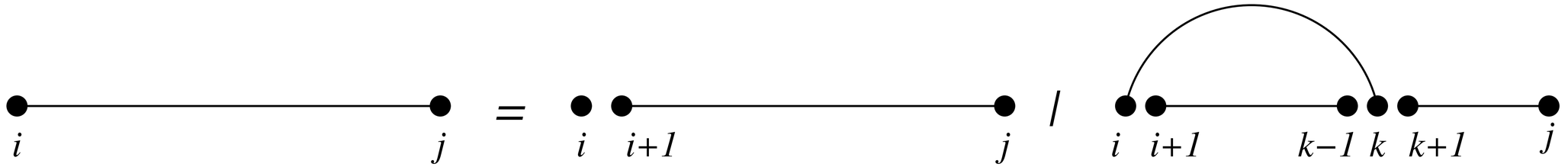






Conformation Space

The number of secondary structures for a sequence $x = x_1 \ldots x_n$ can be computed recursively



$$S_{ij} = S_{i+1,j} + \sum_{k=i+m}^{j} S_{i+1,k-1} S_{k+1,j} \Pi_{kj}$$

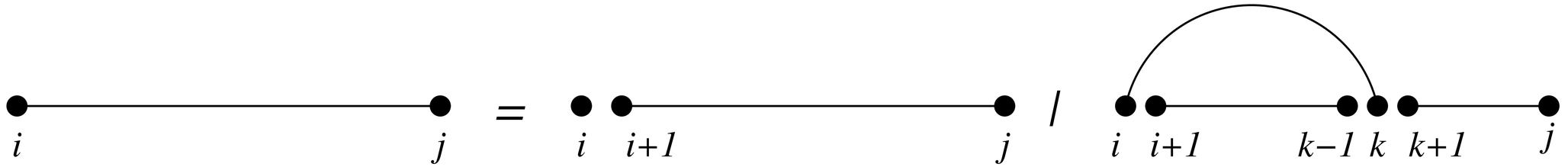
$\Pi_{ik} = 1$ if $x_i x_k \in \{GC, CG, AU, UA, GU, UG\}$, otherwise $\Pi_{ik} = 0$

For sequences with equal {A,U,G,C} content, the number of conformations grows asymptotically with sequence length

$$\bar{S}(n) \sim n^{-\frac{3}{2}} 1.85^n$$

Conformation Space

The number of secondary structures for a sequence $x = x_1 \ldots x_n$ can be computed recursively



$$S_{ij} = S_{i+1,j} + \sum_{k=i+m}^{j} S_{i+1,k-1} S_{k+1,j} \Pi_{kj}$$

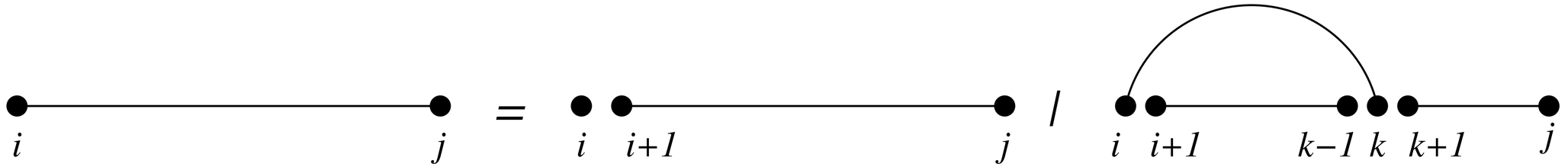
$\Pi_{ik} = 1$ if $x_i x_k \in \{GC,CG,AU,UA,GU,UG\}$, otherwise $\Pi_{ik} = 0$

For sequences with equal {A,U,G,C} content, the number of conformations grows asymptotically with sequence length

$$\bar{S}(n) \sim n^{-\frac{3}{2}} 1.85^n$$

Conformation Space

The number of secondary structures for a sequence $x = x_1 \ldots x_n$ can be computed recursively



$$S_{ij} = S_{i+1,j} + \sum_{k=i+m}^{j} S_{i+1,k-1} S_{k+1,j} \Pi_{kj}$$

$\Pi_{ik} = 1$ if $x_i x_k \in \{GC,CG,AU,UA,GU,UG\}$, otherwise $\Pi_{ik} = 0$

For sequences with equal {A,U,G,C} content, the number of conformations grows asymptotically with sequence length

$$\bar{S}(n) \sim n^{-\frac{3}{2}} 1.85^n$$

Conformation Space

The number of secondary structures for a sequence $x = x_1 \ldots x_n$ can be computed recursively

Many sequences fold into the same structure

$$S_{ij} = S_{i+1,j} + \sum_{k=i+m}^{j} S_{i+1,k-1} S_{k+1,j} \Pi_{kj}$$

$\Pi_{ik} = 1$ if $x_i x_k \in \{GC, CG, AU, UA, GU, UG\}$, otherwise $\Pi_{ik} = 0$

For sequences with equal {A,U,G,C} content, the number of conformations grows asymptotically with sequence length

$$\bar{S}(n) \sim n^{-\frac{3}{2}} 1.85^n$$

Conformation Space

The number of secondary structures for a sequence $x = x_1 \ldots x_n$ can be computed recursively

Many sequences fold into the same structure

Nature ‘exploits’ this property

$\Pi_{ik} = 1$ if $x_i x_k \in \{GC, CG, AU, UA, GU, UG\}$, otherwise $\Pi_{ik} = 0$

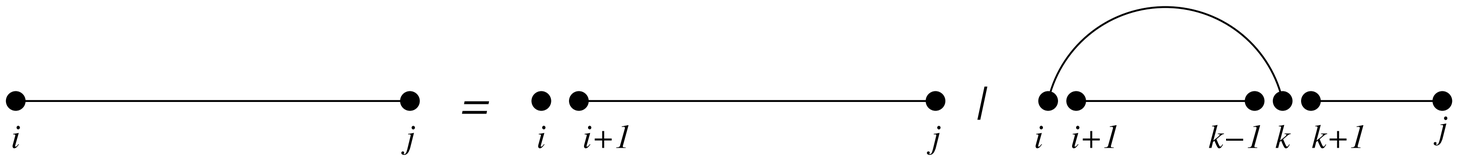
For sequences with equal {A,U,G,C} content, the number of conformations grows asymptotically with sequence length

$$\bar{S}(n) \sim n^{-\frac{3}{2}} 1.85^n$$

Solving the RNA Folding Problem

Toy model for RNA folding: assign energies to base pairs $\varepsilon(x, y)$

Easily solved by Dynamic Programming: recursive computation with tabulation of intermediate results



$$E_{ij} = \min_{i<k\leq j} \left\{ E_{i+1,j} \; ; \quad \left(E_{i+1,k-1} + E_{k+1,j} \; + \varepsilon(x_i, x_k) \right) \right\}$$

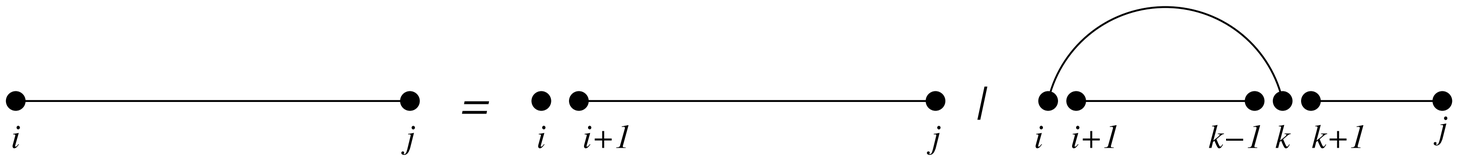
- E_{1n} is the best possible energy for our sequence
- Backtracing through the E table yields the corresponding structure
- The algorithms requires $\mathcal{O}(n^2)$ memory and $\mathcal{O}(n^3)$ CPU time

In practice this toy model is not good enough !
We need loop-dependent energies for serious predictions

Solving the RNA Folding Problem

Toy model for RNA folding: assign energies to base pairs $\varepsilon(x, y)$

Easily solved by Dynamic Programming: recursive computation with tabulation of intermediate results

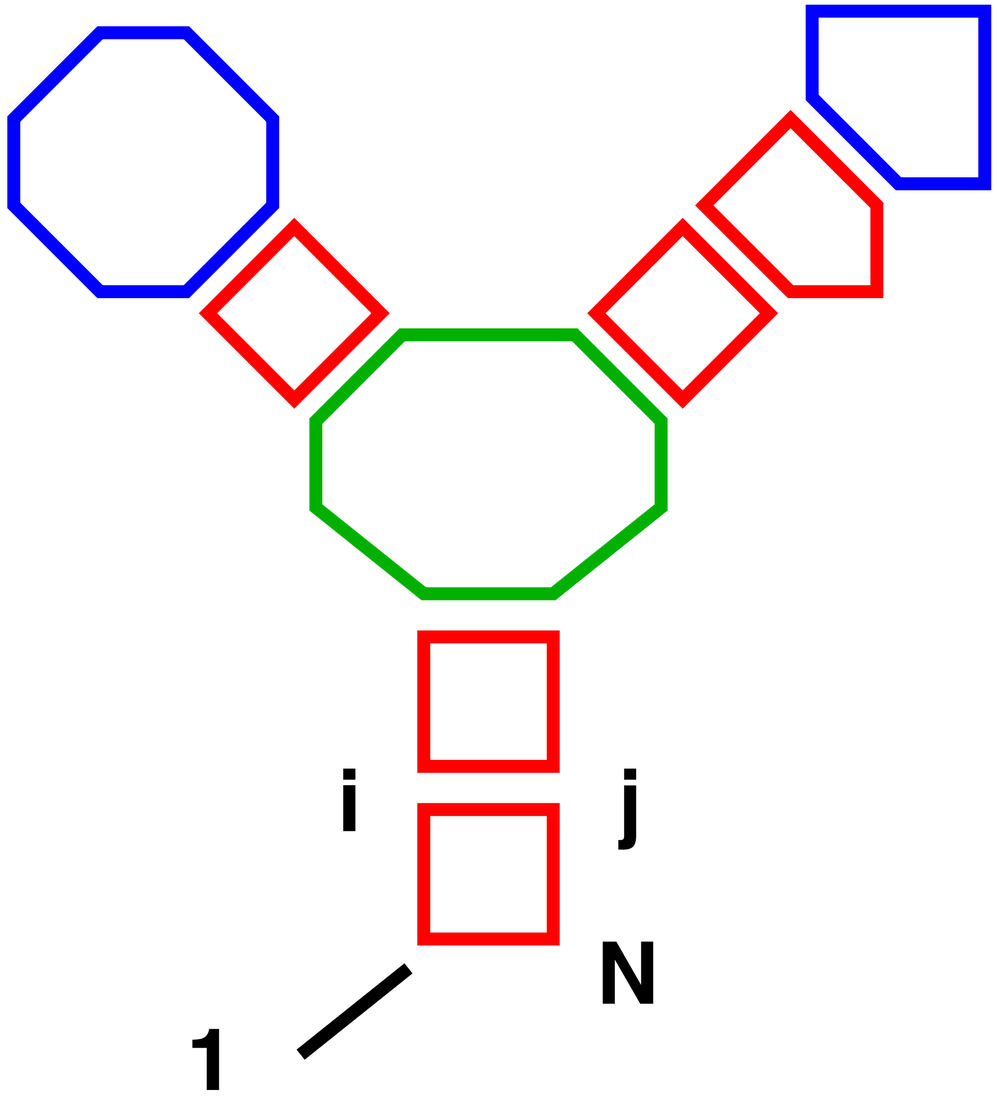


$$E_{ij} = \min_{i<k\leq j} \left\{ E_{i+1,j} \; ; \quad \left(E_{i+1,k-1} + E_{k+1,j} \; + \varepsilon(x_i, x_k) \right) \right\}$$

- E_{1n} is the best possible energy for our sequence
- Backtracing through the E table yields the corresponding structure
- The algorithms requires $\mathcal{O}(n^2)$ memory and $\mathcal{O}(n^3)$ CPU time

In practice this toy model is not good enough !
We need loop-dependent energies for serious predictions

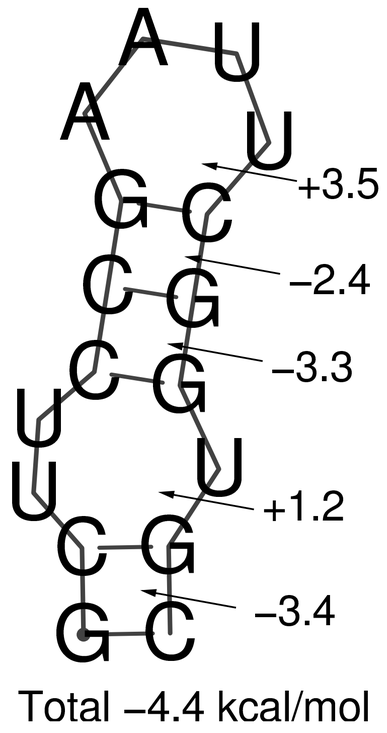
Loop Decomposition



Nearest Neighbour Model

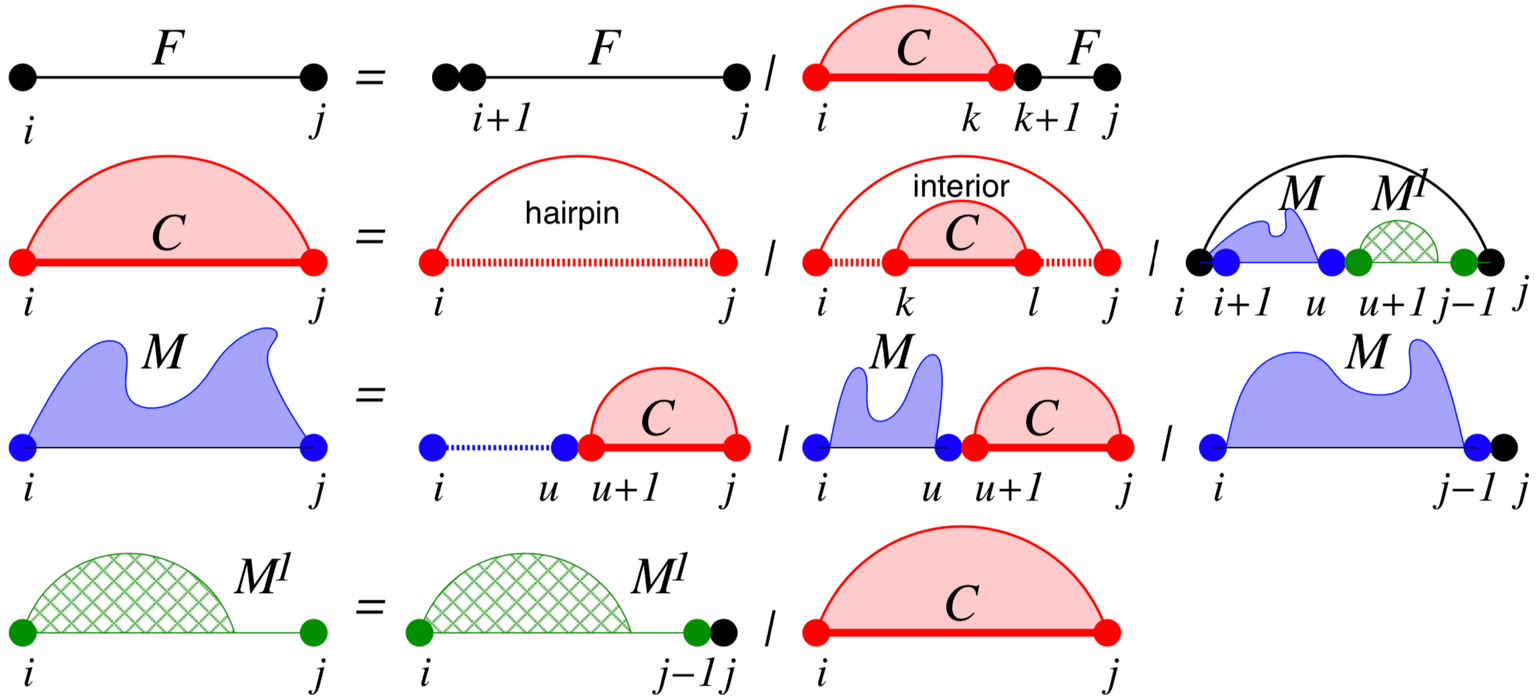
The standard energy model expresses the free energy of a structure as the sum over its loop energies

$$E(S) = \sum_{l \in S} E(l)$$



- Good approximation for most oligonucleotides
- Loop energies depend on loop type/size and some sequence dependence
- Most relevant parameters are experimentally measured; some still guesswork
- Secondary structures are macro-states, hence energies are temperature-dependent free energies
- Training parameters is becoming a viable alternative to experiment

Folding with Loop Based Energies



F_{ij} free energy of the optimal substructure on the subsequence $x[i..j]$.

C_{ij} optimal free energy on $x[i..j]$, where (i, j) pair.

M_{ij} $x[i..j]$ is part of a multiloop and contains at least one pair.

M^1_{ij} same as M_{ij} but contains exactly one component closed by (i, h).

Partition Function

Recall: $\bar{S}(n) \sim n^{-\frac{3}{2}} 1.85^n$

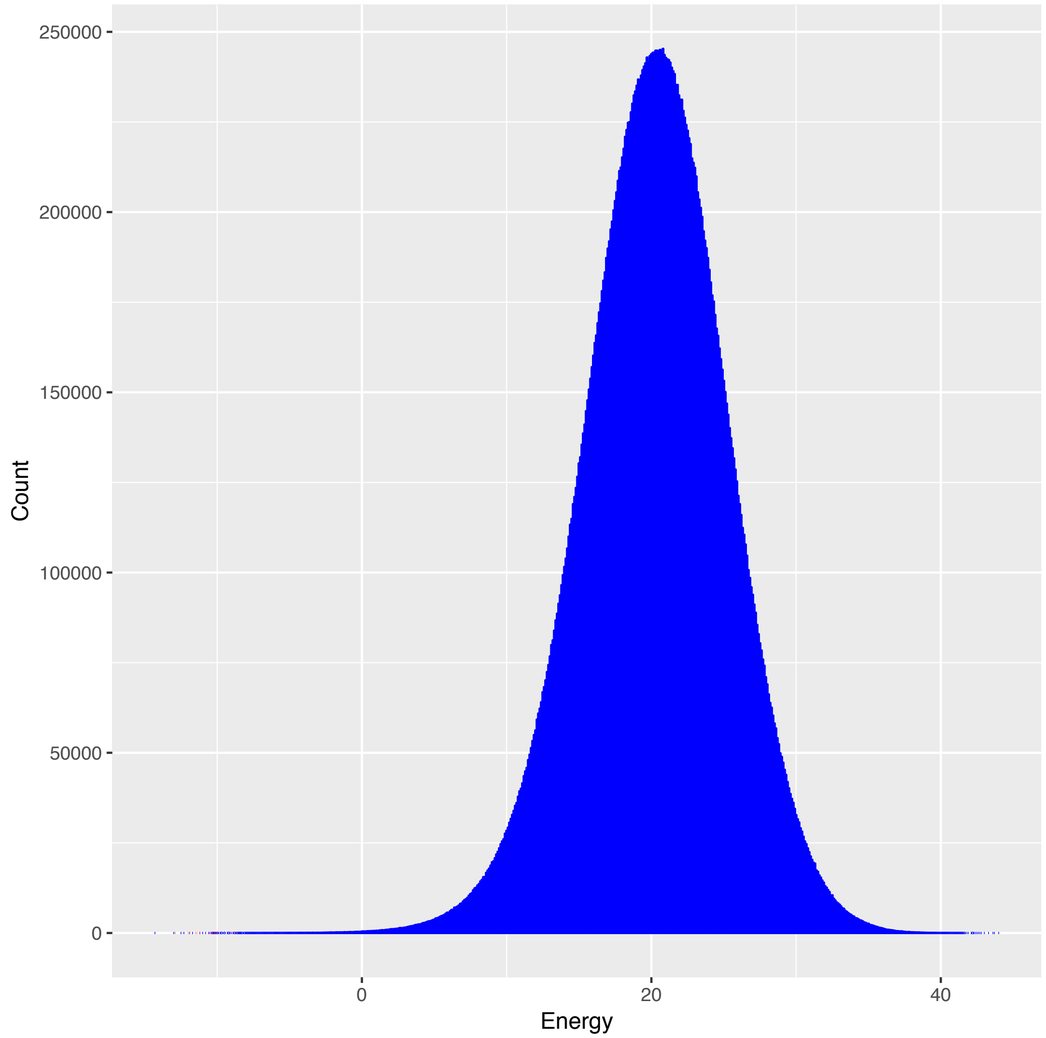
RNA is a biopolymer and ruled by thermodynamics

The partition function is the fundamental quantity of statistical mechanics and all thermodynamic properties can be derived from it

$$Z = \sum_{\Psi} exp\left(-\frac{E(\Psi)}{RT}\right)$$

E.g. the free energy of formation is given by

$$\Delta G = -RT \ln Z$$



Partition Function

Recall: $\bar{S}(n) \sim n^{-\frac{3}{2}} 1.85^n$

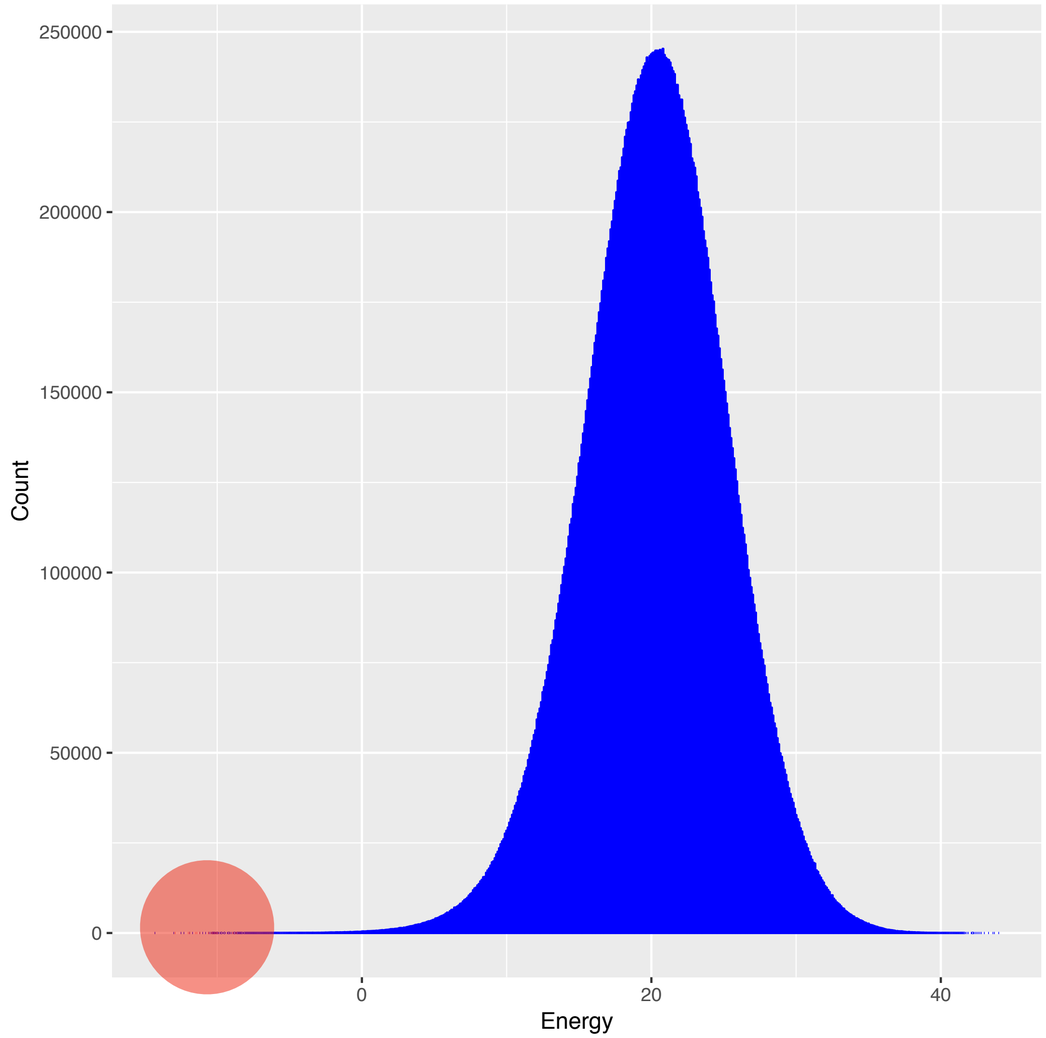
RNA is a biopolymer and ruled by thermodynamics

The partition function is the fundamental quantity of statistical mechanics and all thermodynamic properties can be derived from it

$$Z = \sum_{\Psi} exp\left(-\frac{E(\Psi)}{RT}\right)$$

E.g. the free energy of formation is given by

$$\Delta G = -RT \ln Z$$



Computing the Partition Function

The recursion has the same structure as for energy minimisation, with two differences

- replace minimum operation by sums
- addition of energies by products of partition functions

$$E_{ij} = \min_{i<k\leq j} \left\{ E_{i+1,j}\,;\ \ \left(E_{i+1,k-1} + E_{k+1,j} + \varepsilon(x_i, x_k) \right) \right\}$$

$$Z_{ij} = Z_{i+1,j} + \sum_{k,\,(i,k)\ \text{pairs}} Z_{i+1,k-1} Z_{k+1,j} \exp(-\varepsilon(x_i, x_k)/RT)$$

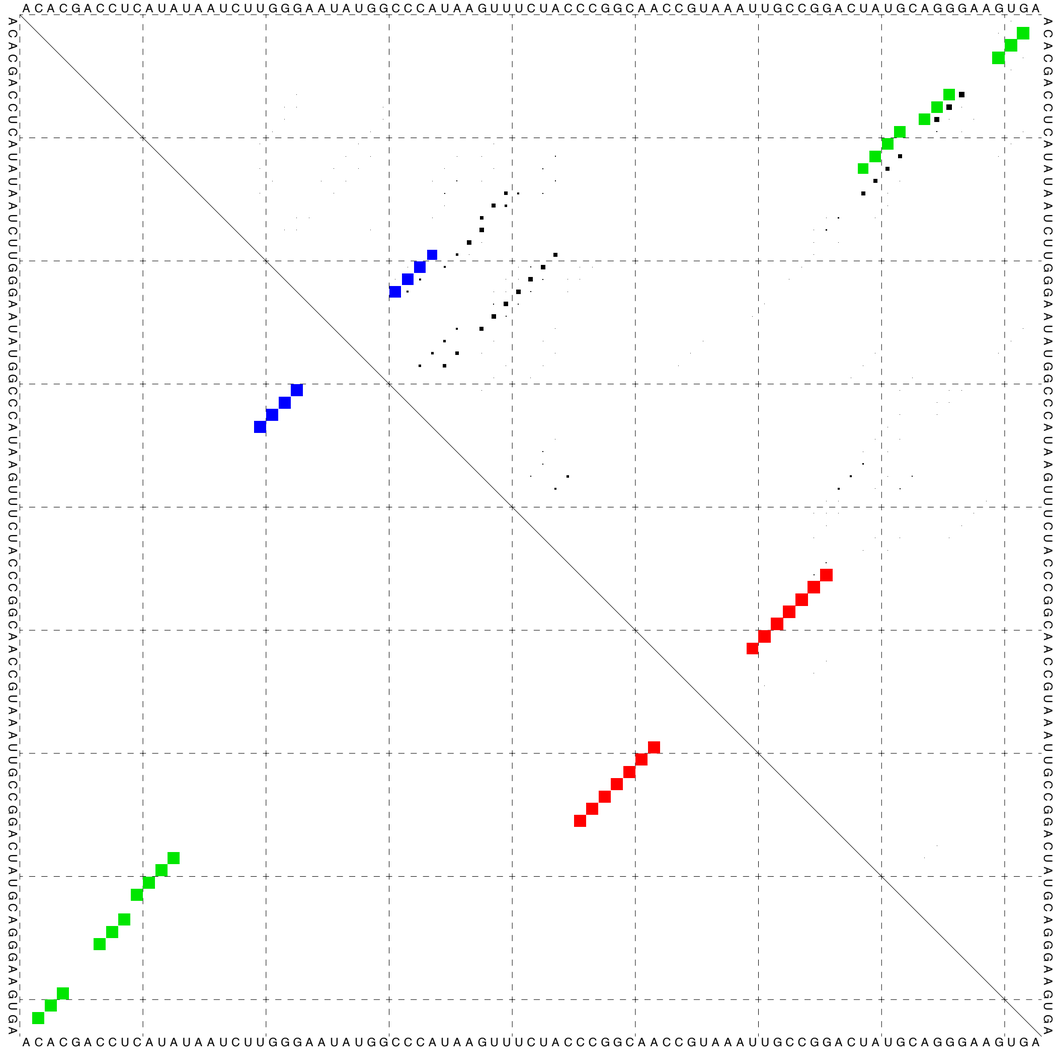
The probability of structure features can be computed from Z, e.g. the probability that a pair is formed

$$p_{ij} = \sum_{\Psi,(i,j)\in\Psi} p(S)$$

For efficient computation define the partition function $\widehat{Z}_{ij}$ for structures outside the subsequence x[i..j]

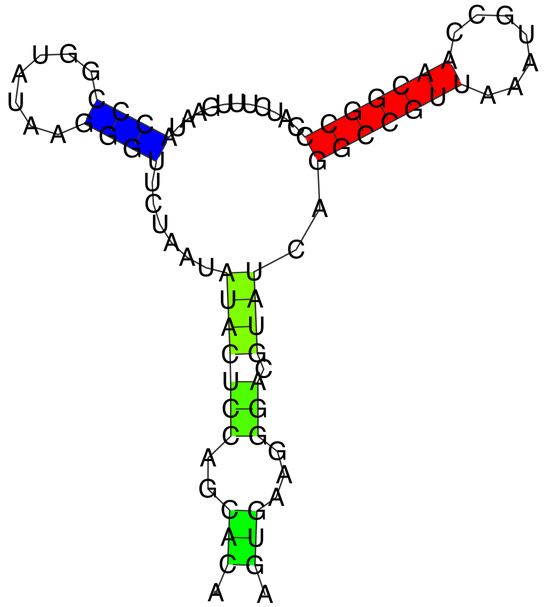
$$p_{ij} = \widehat{Z}_{ij} Z_{i+1,j-1} \exp(-\varepsilon_{ij}/RT)/Z$$

Representing Ensembles of RNA Structures



Ensembles of structures (thermodynamic equilibrium) are best represented by base pair probabilities.

A pair (i , j) with probability p is represented by a square in row i and column j with area p.



The Vienna RNA Package

- Minimum free energy and partition function folding
- Complete suboptimal folding
- Inverse folding / RNA design
- Comparison of secondary structures
- Specific heat curves
- Inclusion of structure probing data
- Analysis of folding kinetics / co-transcriptional folding
- Utilities for plotting and annotation structures
- 2.5D prediction: G-quadruplexes and pseudo-knots
- Prediction of consensus structures

For the programmer:

- A *C* library to link against your programs
- Python/Perl scripting language interface

The Vienna RNA Package

- Minimum free energy and partition function folding
- Complete suboptimal folding
- Inverse folding / RNA design
- Comparison of secondary structures
- Specific heat curves
- Inclusion of structure probing data
- Analysis of folding kinetics / co-transcriptional folding
- Utilities for plotting and annotation structures
- Prediction of consensus structures

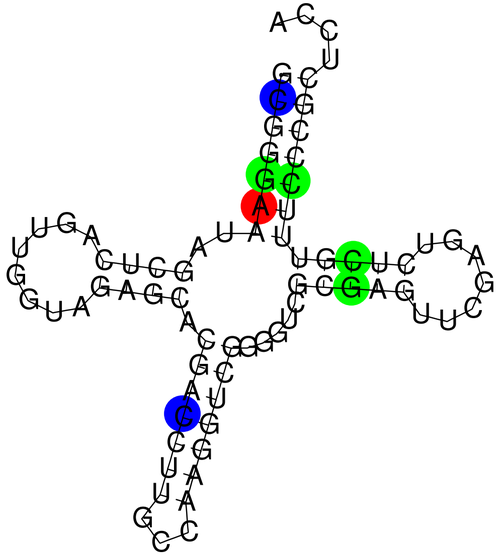
For the programmer:

- A *C* library to link against your programs
- Python/Perl scripting language interface

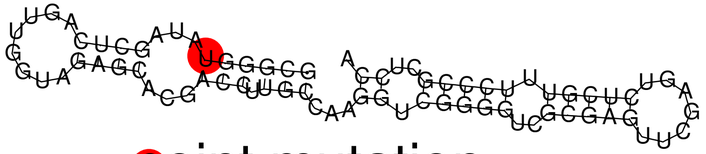
Free software, C source code and fold server available at

http://www.tbi.univie.ac.at/RNA/

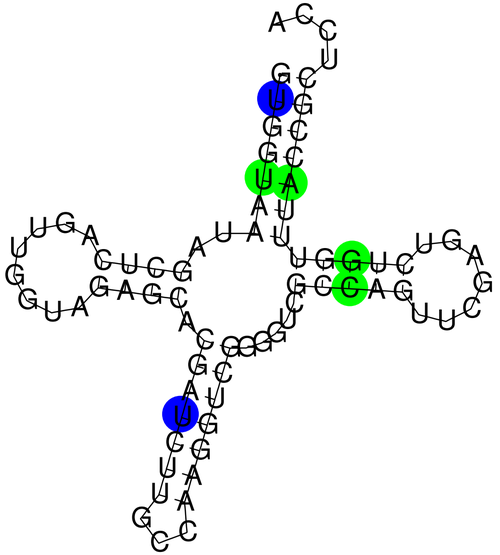
Functional Structures: Point Mutations



original structure

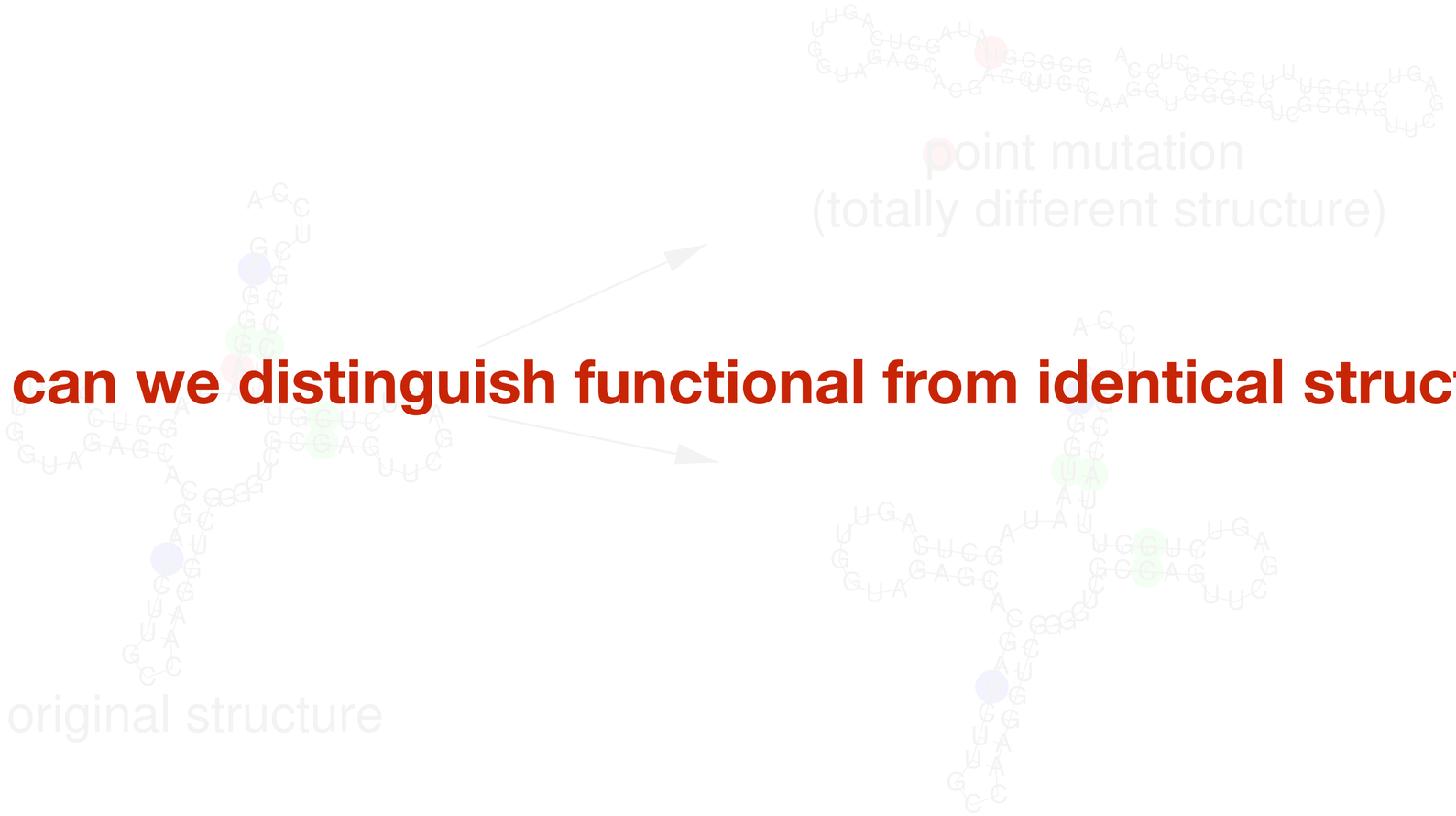


point mutation
(totally different structure)



compensatory and consistent mutations
(no structural change)

Functional Structures: Point Mutations



How can we distinguish functional from identical structures?

Consensus Structures: Alignment Folding

Combine covariance analysis and folding into one DP algorithm

- Apply conventional folding algorithm to alignment
- Use a modified energy function that includes covariance score

$$E_c(A, \Psi) = \sum_k E(A_k, \Psi) + cv \cdot \sum_{(i,j) \in \Psi} B_{ij}$$

- Can be used for all variants: MFE, partition function, …
- Efficient: $\mathcal{O}(N \cdot n^2 + n^3)$ CPU and $\mathcal{O}(n^2)$ memory for alignment length n and N sequences
- Same results as RNAfold for single sequences

Consensus Structures: Alignment Folding

Combine covariance analysis and folding into one DP algorithm

RNAalifold

- Apply conventional folding algorithm to alignment
- Use a modified energy function that includes covariance score

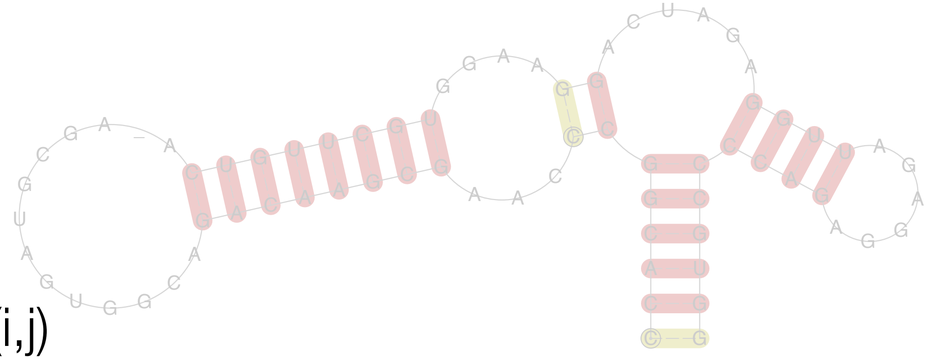
$$E_c(A, \Psi) = \sum_k E(A_k, \Psi) + cv \cdot \sum_{(i,j)\in\Psi} B_{ij}$$

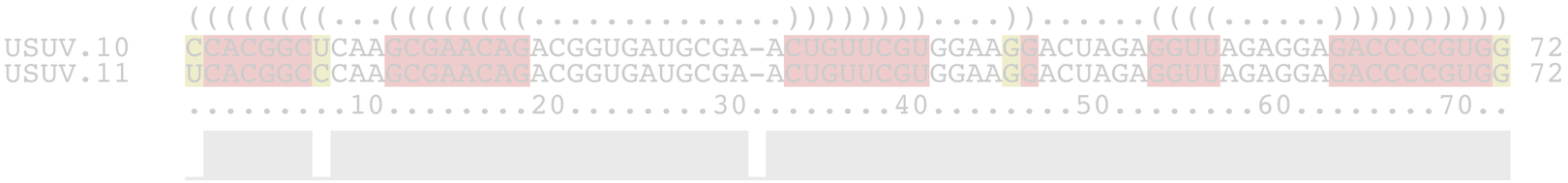
- Can be used for all variants: MFE, partition function, …
- Efficient: $\mathcal{O}(N \cdot n^2 + n^3)$ CPU and $\mathcal{O}(n^2)$ memory for alignment length n and N sequences
- Same results as RNAfold for single sequences

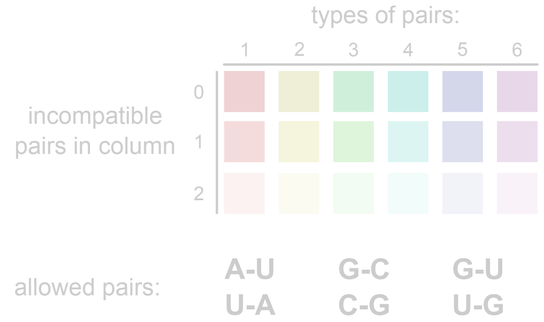
```
GGGCCGGUAGCUCAUUUAGGCAGAGCGUCUGACUCUUAAUCAGACGGUCGCGUGUUCGAAUCGCGUCCGGCCCA
GCCGAAAUAGCUCAGUU-GGGAGAGCGUUAGACUGAAGAUCUAAAGGUCCCCGGUUCAAUCCCGGGUUUCGGCA
GCGGAUUUAGCUCAGUU-GGGAGAGCGCCAGACUGAAGAUUUGGAGGUCCUGUGUUCGAUCCACAGAAUUCGCA
CAGAGUGUAGCUUA---ACACAAAGCACCCAACUUACACUUAGGAGAU-UUCAACU-UAACUUGACCGCUCUGA

(((((((..((((.........)))).(((((.......)))))).....(((((.......)))))))))))).
mfe: -29 kcal/mol = (-21.3 +  -7.7) kcal/mol
```

RNA Covariation as Evolutionary Trait

- High mutation rate in RNA viruses due to error-prone RdRP
- For base pair (i,j): GC/CG/AU/UA/GU/UG
- Consistent mutation: different standard combinations
- Compensatory mutation: both positions are mutated
- Presence of both strongly supports predicted base pair (i,j)

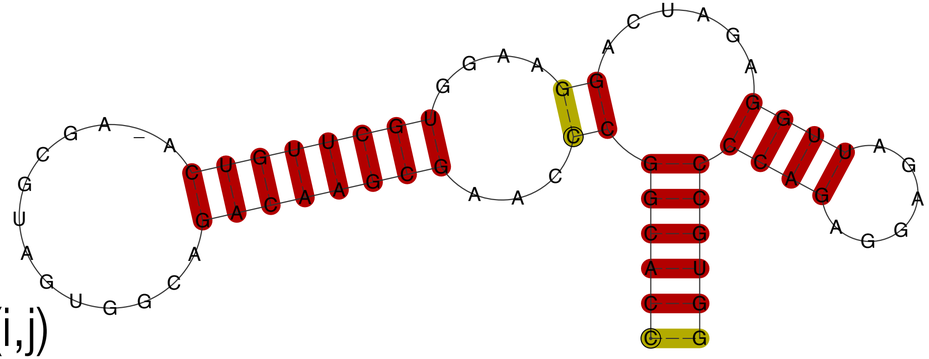




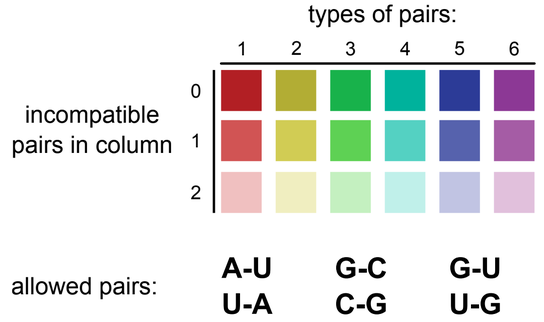


RNA Covariation as Evolutionary Trait

- High mutation rate in RNA viruses due to error-prone RdRP
- For base pair (i,j): GC/CG/AU/UA/GU/UG
- Consistent mutation: different standard combinations
- Compensatory mutation: both positions are mutated
- Presence of both strongly supports predicted base pair (i,j)

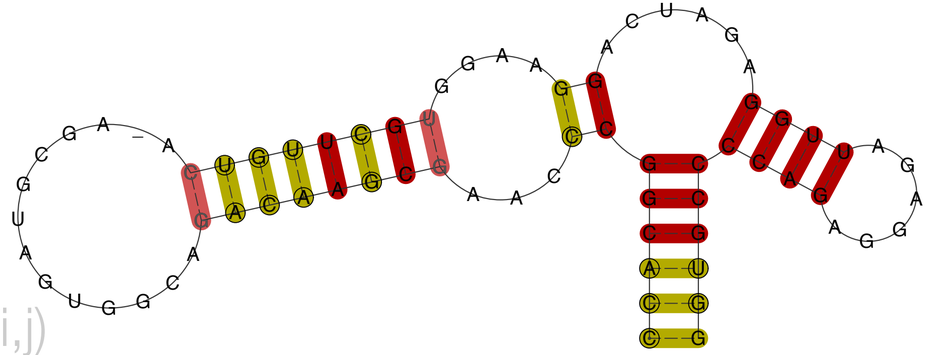


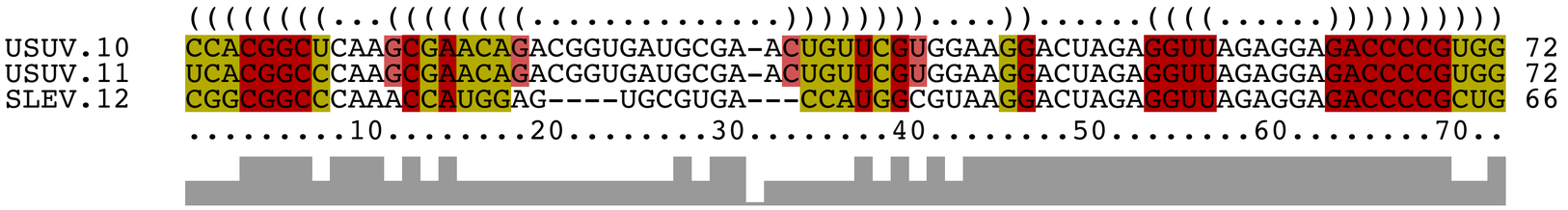
```
                 (((((((((...(((((((((..............)))))))))....)).....(((((.....)))))))))))
USUV.10          CCACGGCUCAAGCGAACAGACGGUGAUGCGA-ACUGUUCGUGGAAGGACUAGAGGUUAGAGGAGACCCCGUGG 72
USUV.11          UCACGGCCCAAGCGAACAGACGGUGAUGCGA-ACUGUUCGUGGAAGGACUAGAGGUUAGAGGAGACCCCGUGG 72
                 .........10........20........30........40........50........60........70..
```

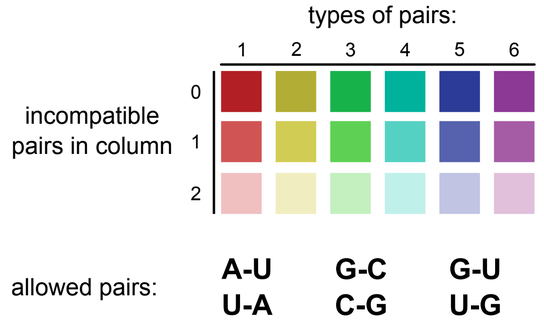


RNA Covariation as Evolutionary Trait

- High mutation rate in RNA viruses due to error-prone RdRP
- For base pair (i,j): GC/CG/AU/UA/GU/UG
- Consistent mutation: different standard combinations
- Compensatory mutation: both positions are mutated
- Presence of both strongly supports predicted base pair (i,j)

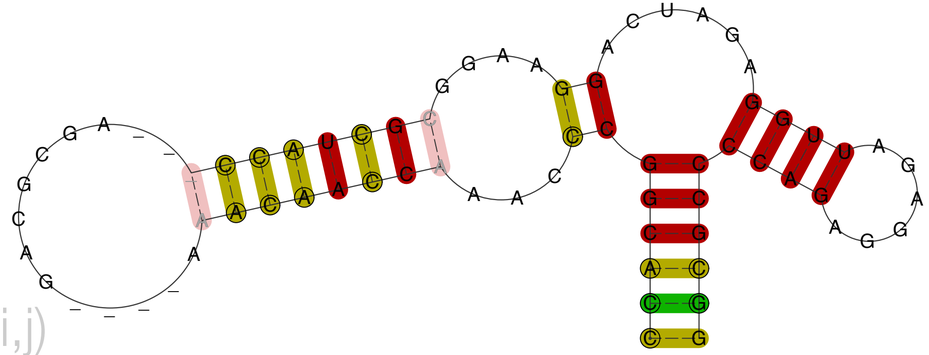


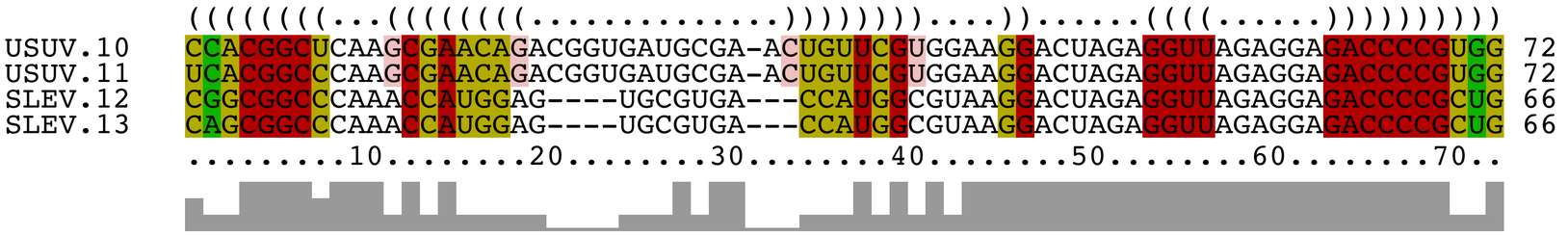


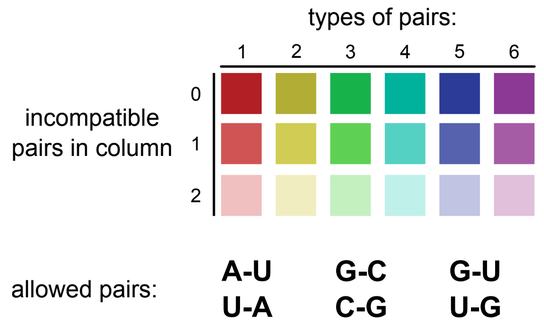


RNA Covariation as Evolutionary Trait

- High mutation rate in RNA viruses due to error-prone RdRP
- For base pair (i,j): GC/CG/AU/UA/GU/UG
- Consistent mutation: different standard combinations
- Compensatory mutation: both positions are mutated
- Presence of both strongly supports predicted base pair (i,j)

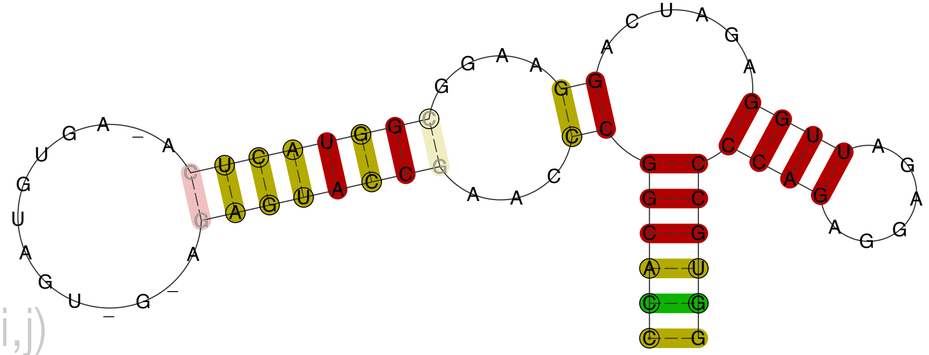


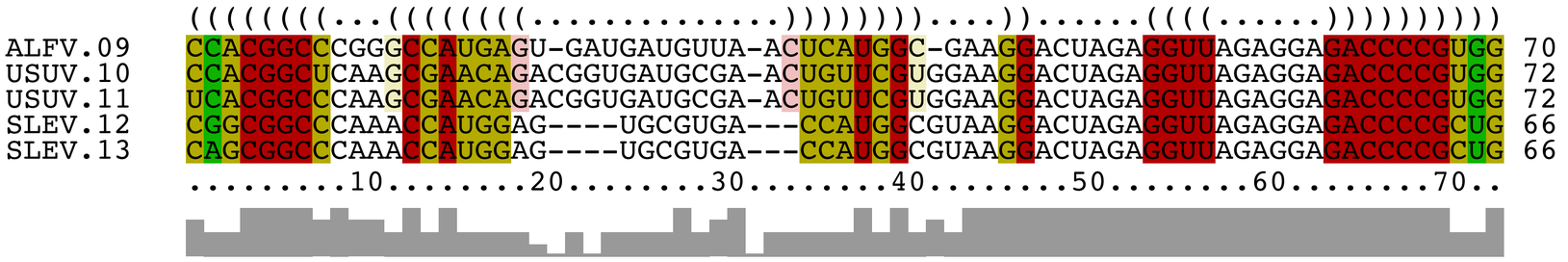


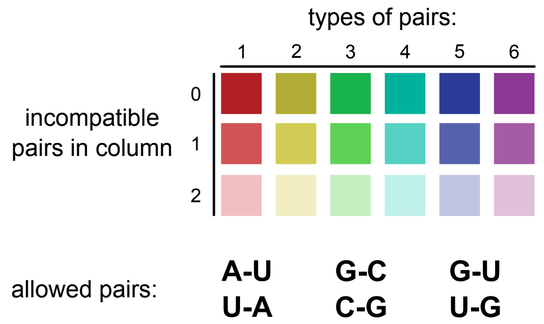


RNA Covariation as Evolutionary Trait

- High mutation rate in RNA viruses due to error-prone RdRP
- For base pair (i,j): GC/CG/AU/UA/GU/UG
- Consistent mutation: different standard combinations
- Compensatory mutation: both positions are mutated
- Presence of both strongly supports predicted base pair (i,j)

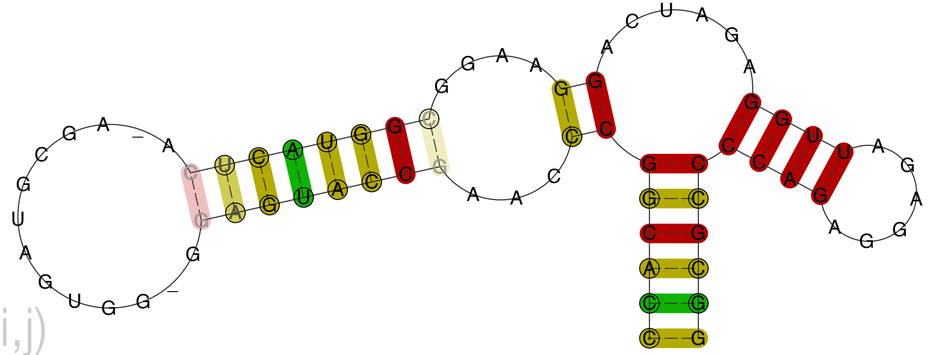


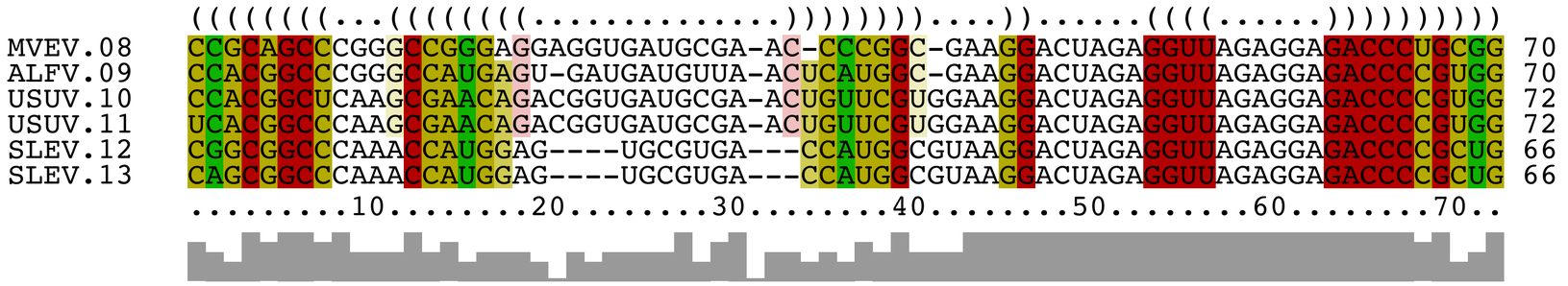


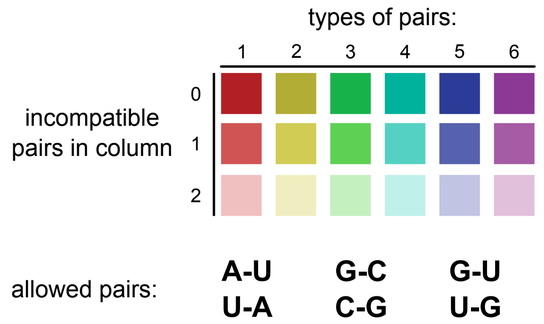


RNA Covariation as Evolutionary Trait

- High mutation rate in RNA viruses due to error-prone RdRP
- For base pair (i,j): GC/CG/AU/UA/GU/UG
- Consistent mutation: different standard combinations
- Compensatory mutation: both positions are mutated
- Presence of both strongly supports predicted base pair (i,j)

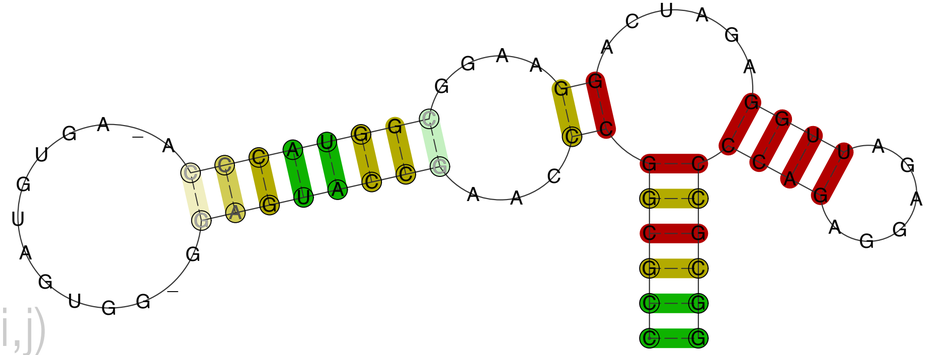


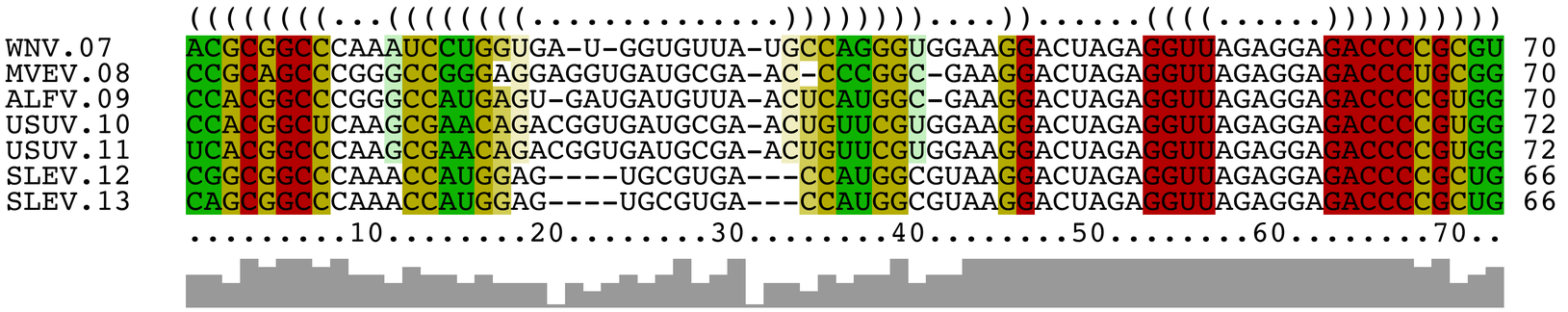


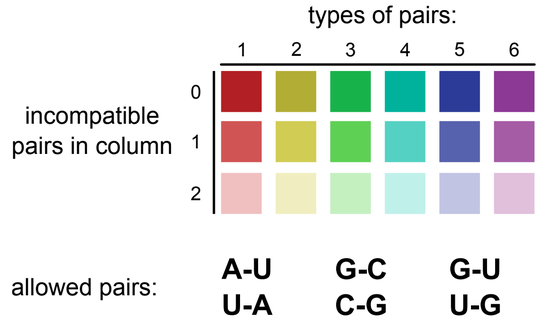


RNA Covariation as Evolutionary Trait

- High mutation rate in RNA viruses due to error-prone RdRP
- For base pair (i,j): GC/CG/AU/UA/GU/UG
- Consistent mutation: different standard combinations
- Compensatory mutation: both positions are mutated
- Presence of both strongly supports predicted base pair (i,j)

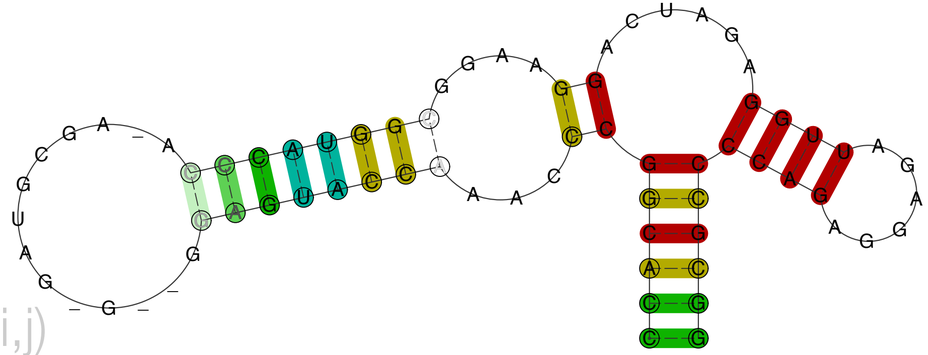


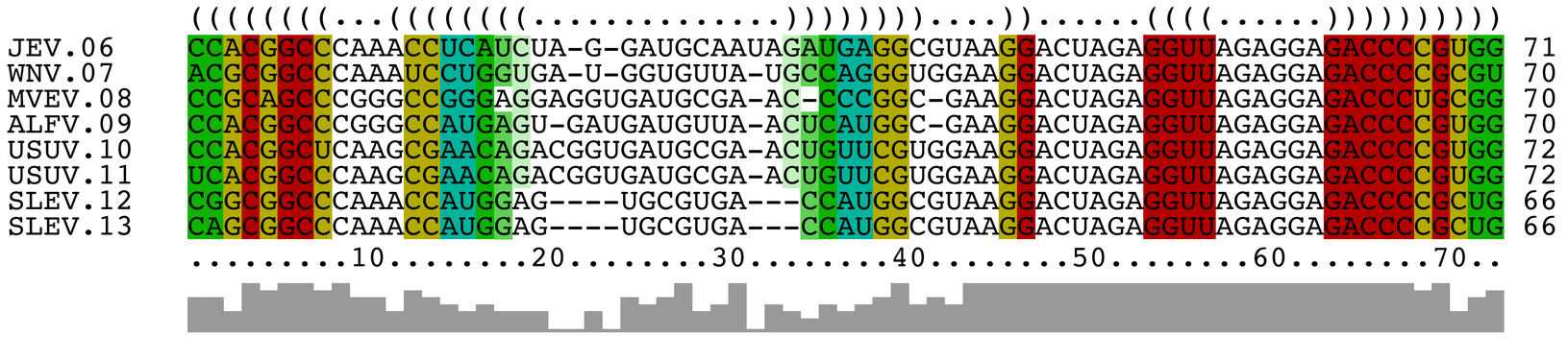


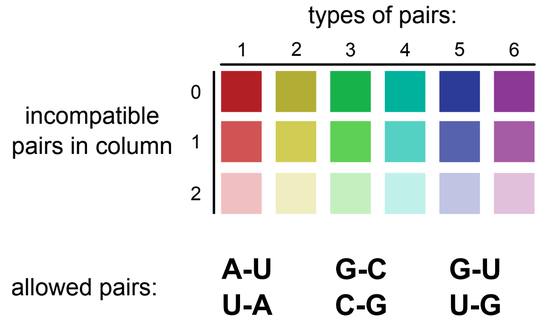


RNA Covariation as Evolutionary Trait

- High mutation rate in RNA viruses due to error-prone RdRP
- For base pair (i,j): GC/CG/AU/UA/GU/UG
- Consistent mutation: different standard combinations
- Compensatory mutation: both positions are mutated
- Presence of both strongly supports predicted base pair (i,j)

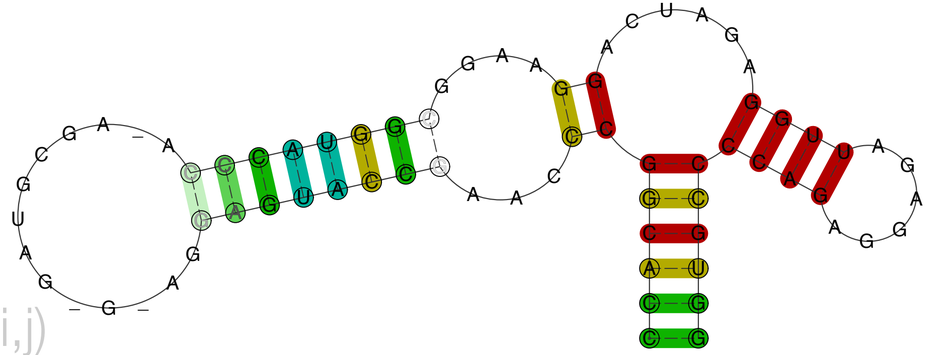


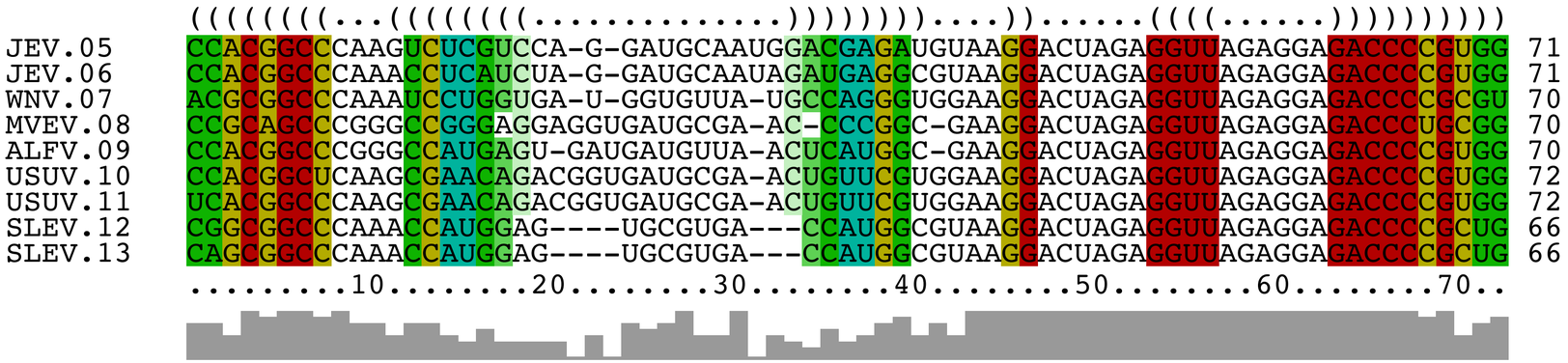


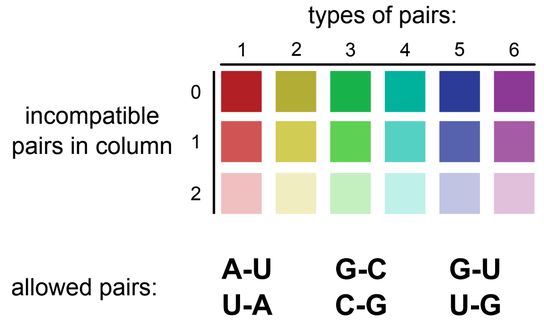


RNA Covariation as Evolutionary Trait

- High mutation rate in RNA viruses due to error-prone RdRP
- For base pair (i,j): GC/CG/AU/UA/GU/UG
- Consistent mutation: different standard combinations
- Compensatory mutation: both positions are mutated
- Presence of both strongly supports predicted base pair (i,j)

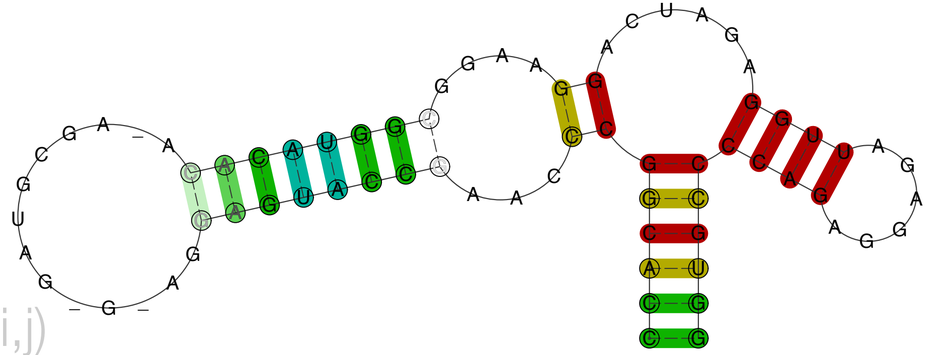


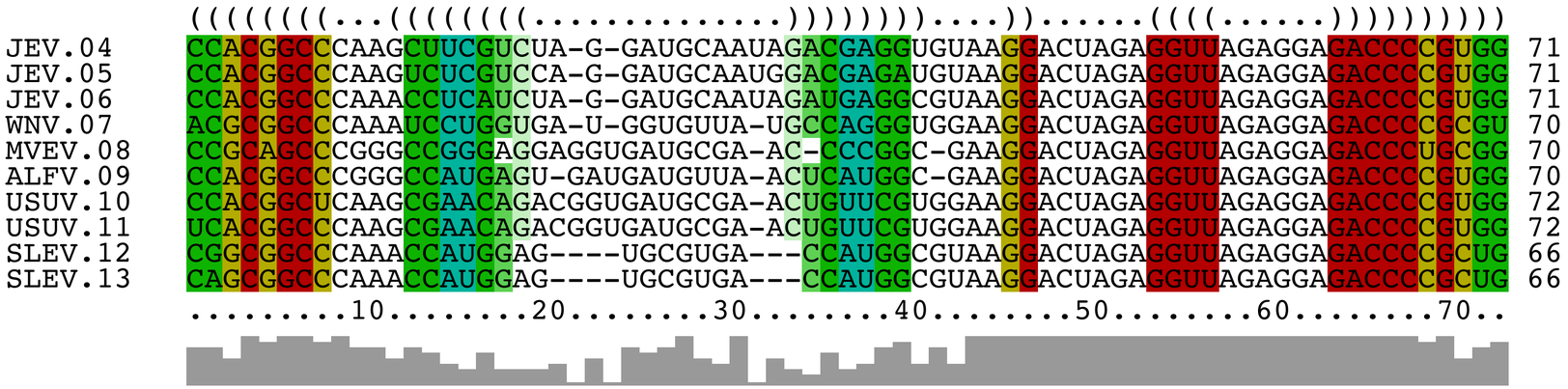


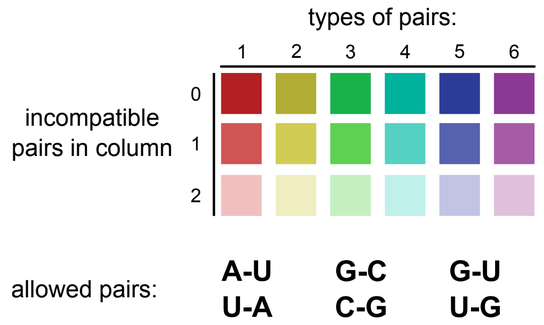


RNA Covariation as Evolutionary Trait

- High mutation rate in RNA viruses due to error-prone RdRP
- For base pair (i,j): GC/CG/AU/UA/GU/UG
- Consistent mutation: different standard combinations
- Compensatory mutation: both positions are mutated
- Presence of both strongly supports predicted base pair (i,j)

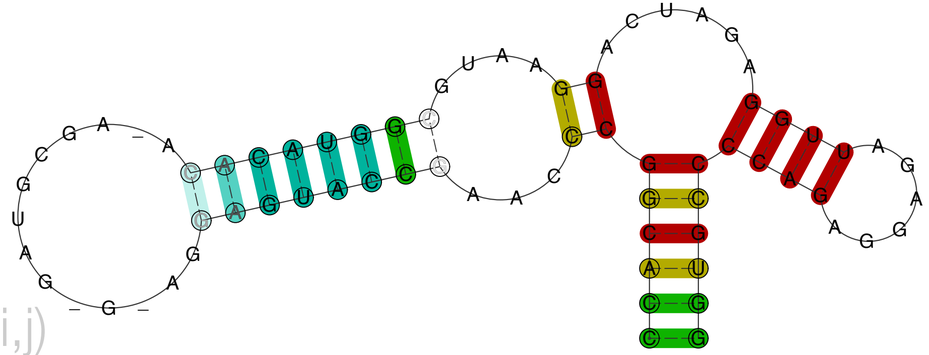


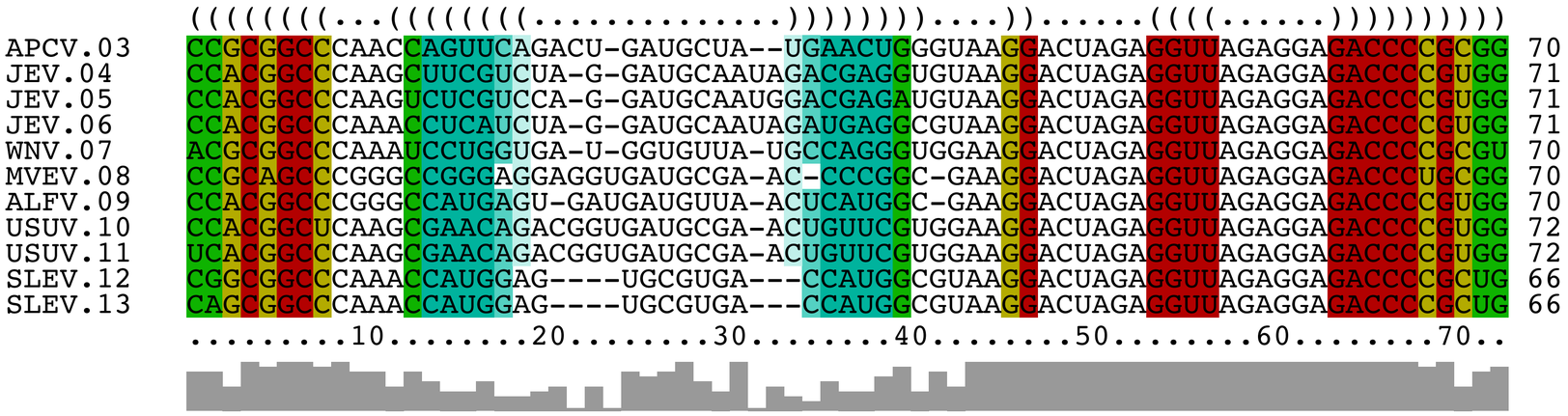


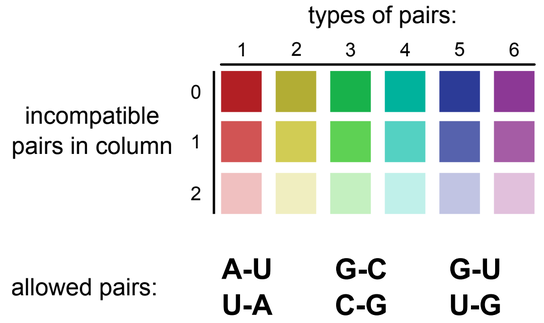


RNA Covariation as Evolutionary Trait

- High mutation rate in RNA viruses due to error-prone RdRP
- For base pair (i,j): GC/CG/AU/UA/GU/UG
- Consistent mutation: different standard combinations
- Compensatory mutation: both positions are mutated
- Presence of both strongly supports predicted base pair (i,j)

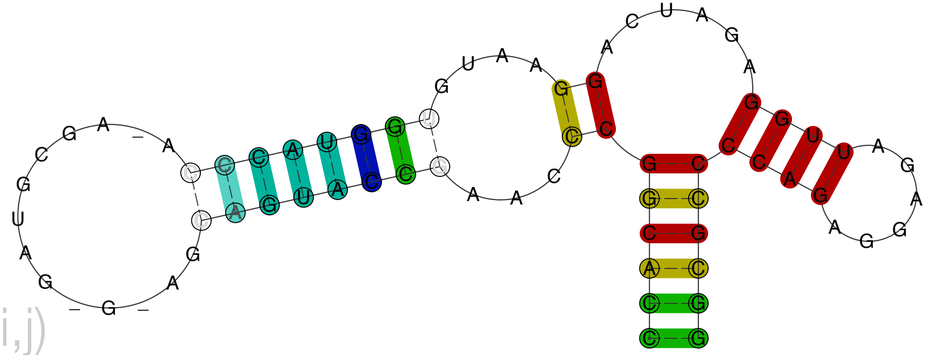


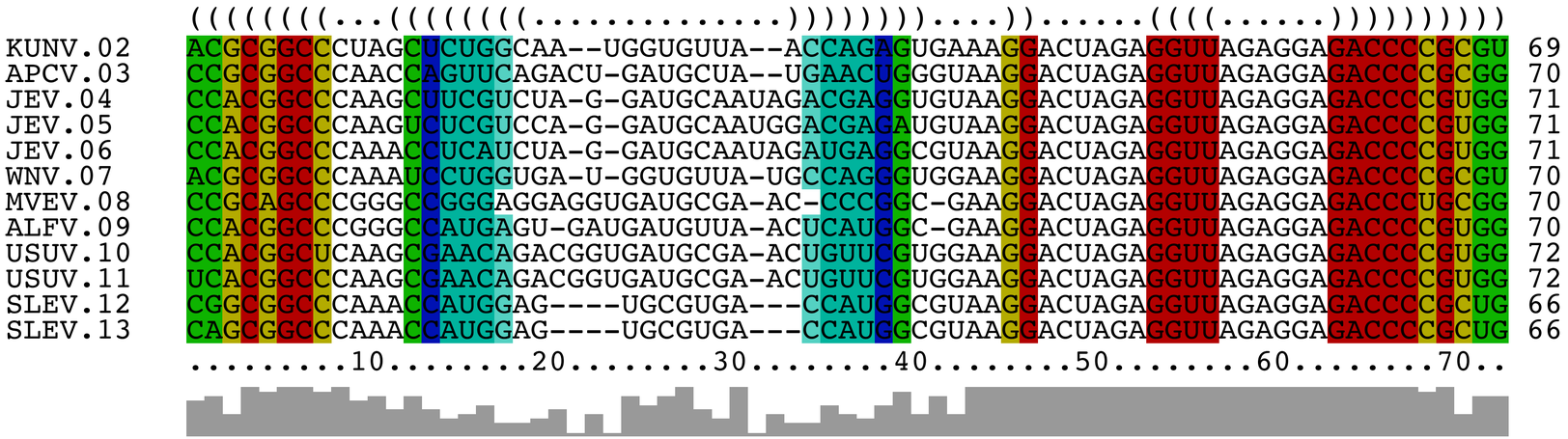


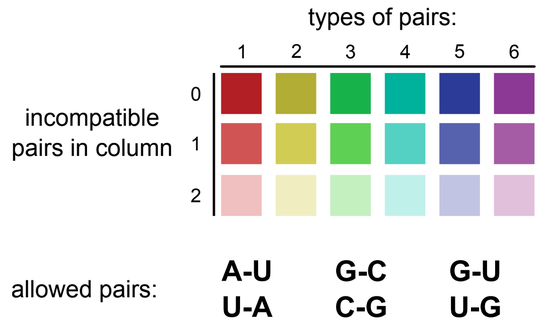


RNA Covariation as Evolutionary Trait

- High mutation rate in RNA viruses due to error-prone RdRP
- For base pair (i,j): GC/CG/AU/UA/GU/UG
- Consistent mutation: different standard combinations
- Compensatory mutation: both positions are mutated
- Presence of both strongly supports predicted base pair (i,j)

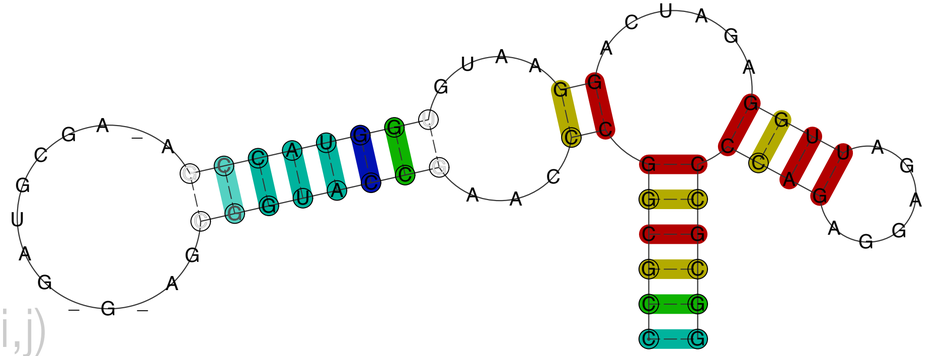


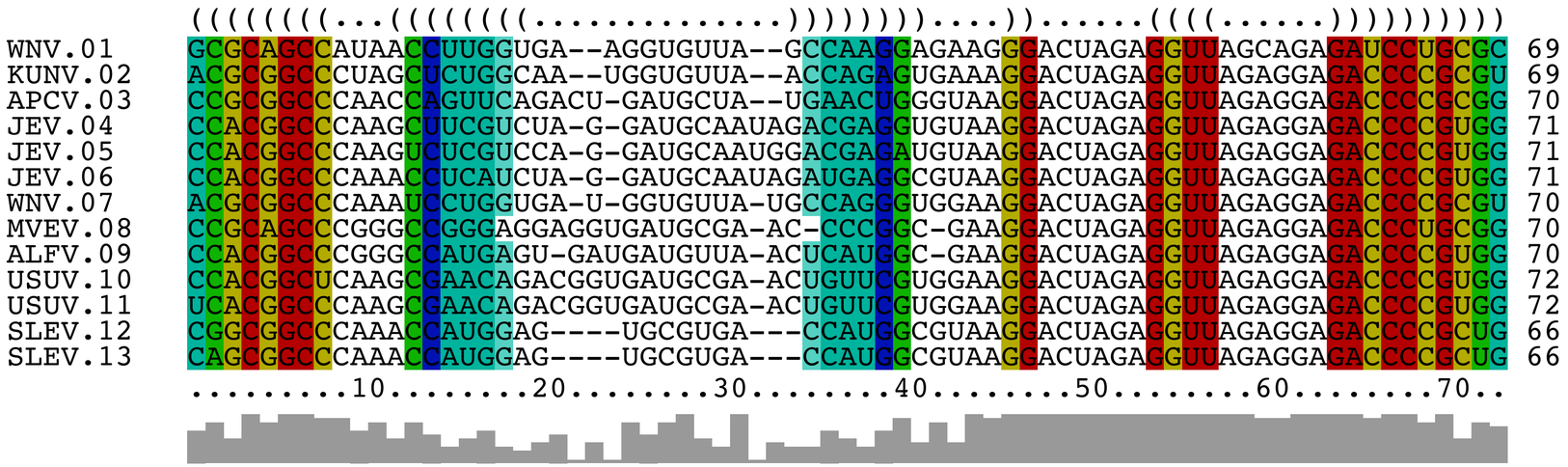


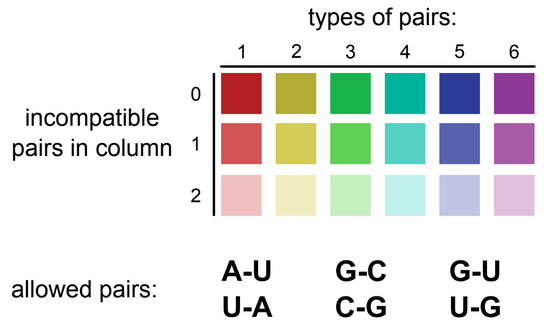


RNA Covariation as Evolutionary Trait

- High mutation rate in RNA viruses due to error-prone RdRP
- For base pair (i,j): GC/CG/AU/UA/GU/UG
- Consistent mutation: different standard combinations
- Compensatory mutation: both positions are mutated
- Presence of both strongly supports predicted base pair (i,j)

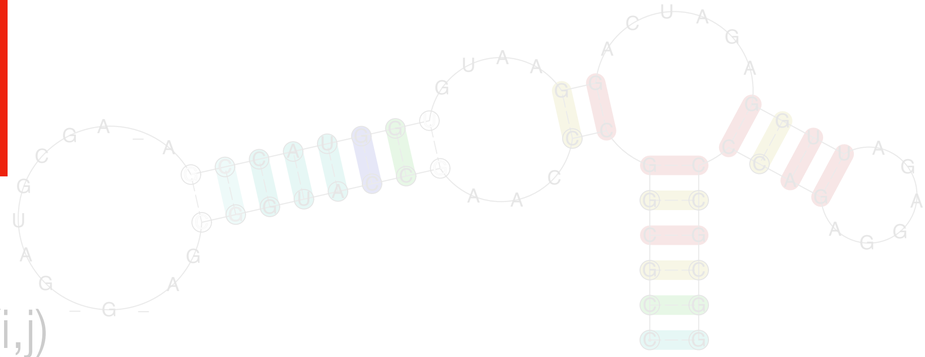


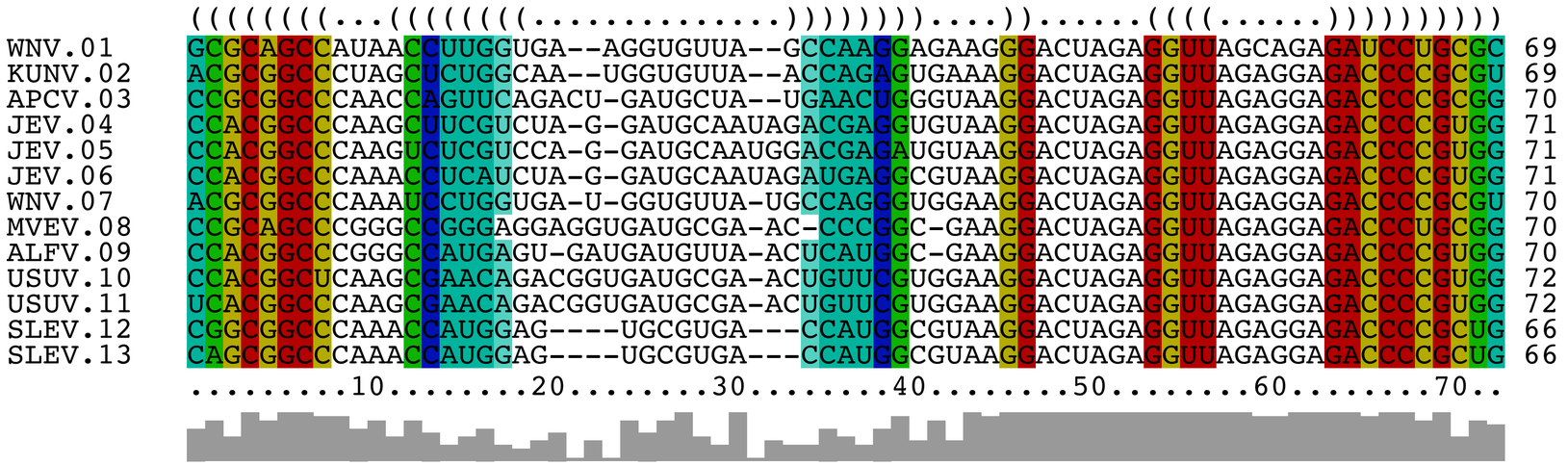


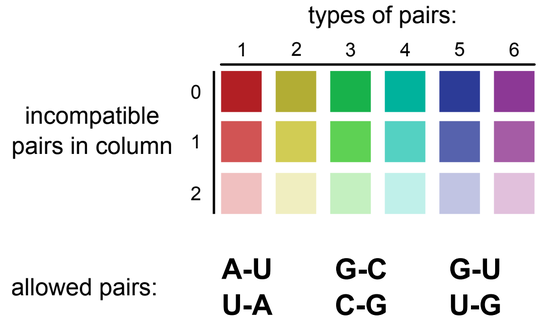


RNA Covariation as Evolutionary Trait

- High mutation rate in RNA viruses due to error-prone RdRP
- **We use structural RNA alignments and Covariance Models to find conserved elements**
- Compensatory mutation: both positions are mutated
- Presence of both strongly supports predicted base pair (i,j)

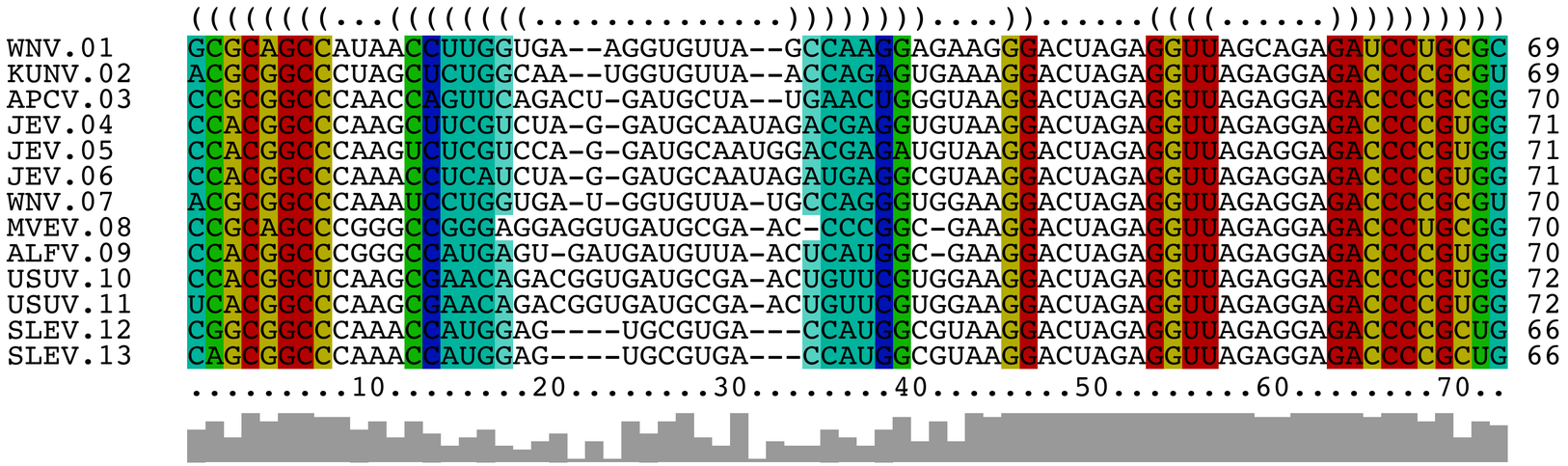


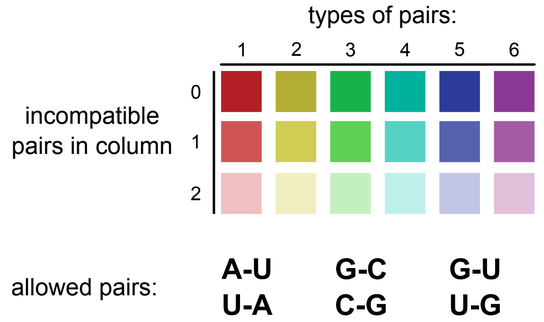




RNA Covariation as Evolutionary Trait

- High mutation rate in RNA viruses due to error-prone RdRP
- **We use structural RNA alignments and Covariance Models to find conserved elements**
- Compensatory mutation: both positions are mutated
- Presence of both strongly supports predicted base pair (i,j)



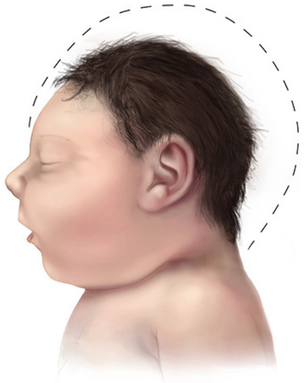


Part II:

Conserved RNA structures in flaviviruses

Flaviviruses as Global Health Threat

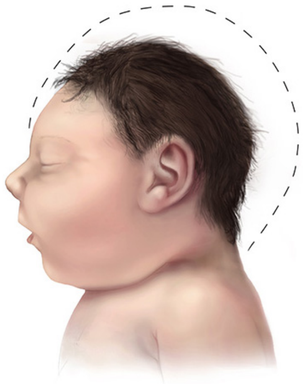
- Family *Flaviviridae* / genus *Flavivirus*; arthropod-borne (mainly mosquitoes & ticks)
- FV include (re-)emerging human pathogens like YFV, DENV, JEV, WNV, TBEV
- Clinical manifestations: headache, rash, (hemorrhagic) fever, meningitis, encephalitis
- FV neurotropism reported for YFV, WNV, TBEV, DENV, ZIKV
- ZIKV outbreak in the Americas 2015-2017
- Congenital neurotropism: high increase in microcephaly cases in newborns



Source: Wikipedia

Flaviviruses as Global Health Threat

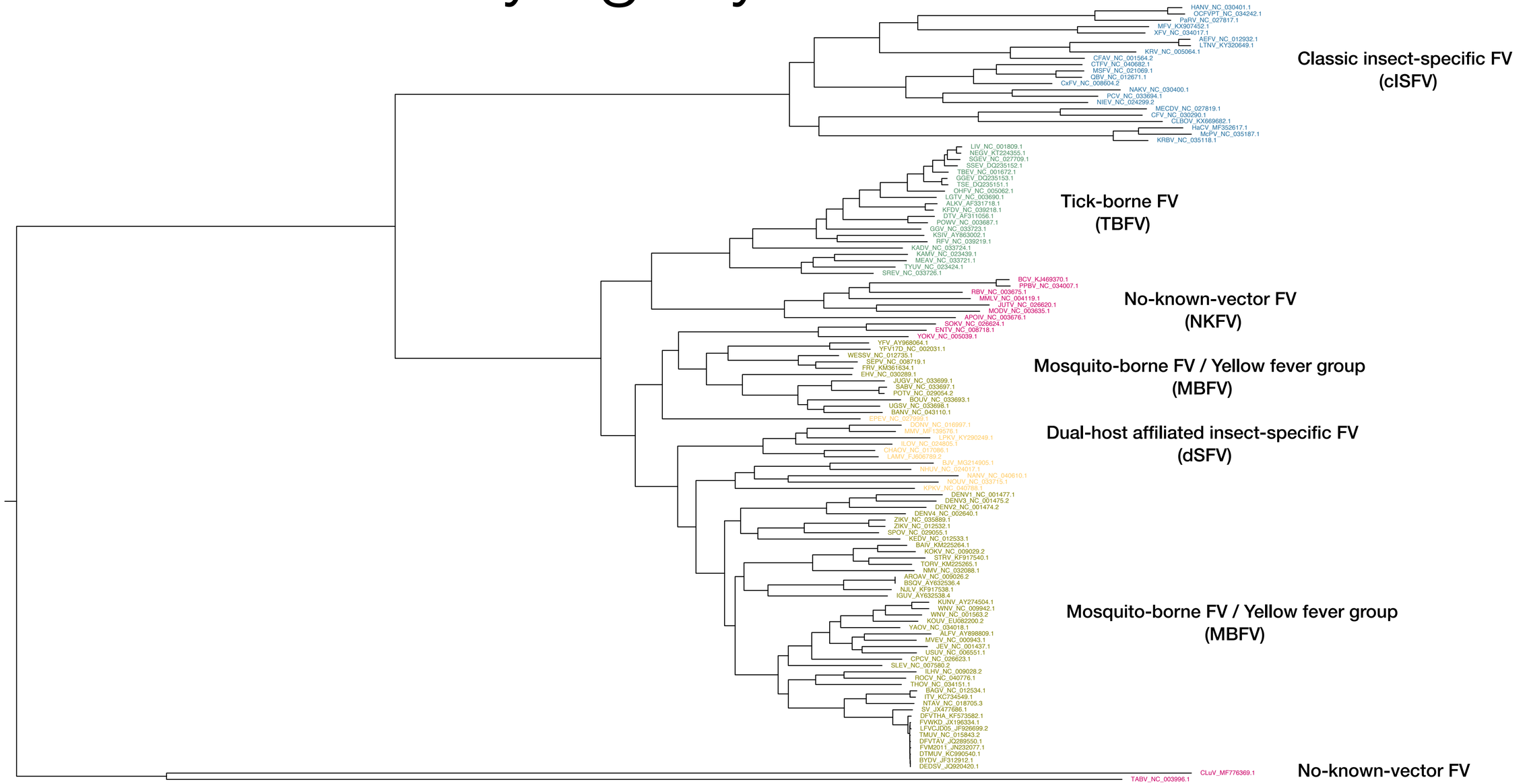
- Family *Flaviviridae* / genus *Flavivirus*; arthropod-borne (mainly mosquitoes & ticks)
- FV include (re-)emerging human pathogens like YFV, DENV, JEV, WNV, TBEV
- Clinical manifestations: headache, rash, (hemorrhagic) fever, meningitis, encephalitis
- FV neurotropism reported for YFV, WNV, TBEV, DENV, ZIKV
- ZIKV outbreak in the Americas 2015-2017
- Congenital neurotropism: high increase in microcephaly cases in newborns



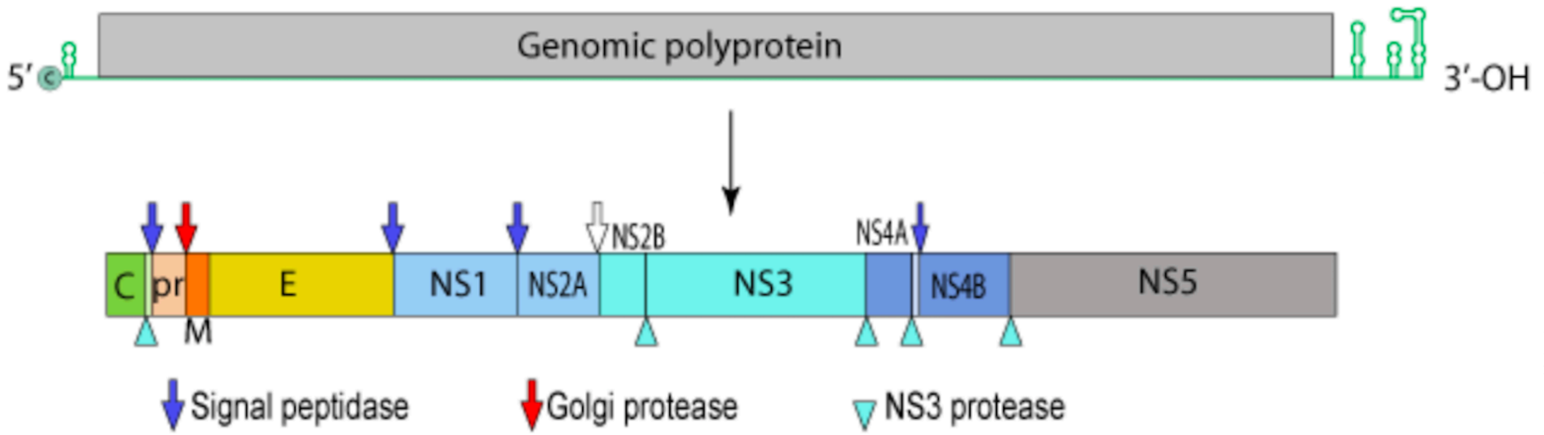
Source: Wikipedia

Vaccine only available for YFV

Flavivirus Phylogeny



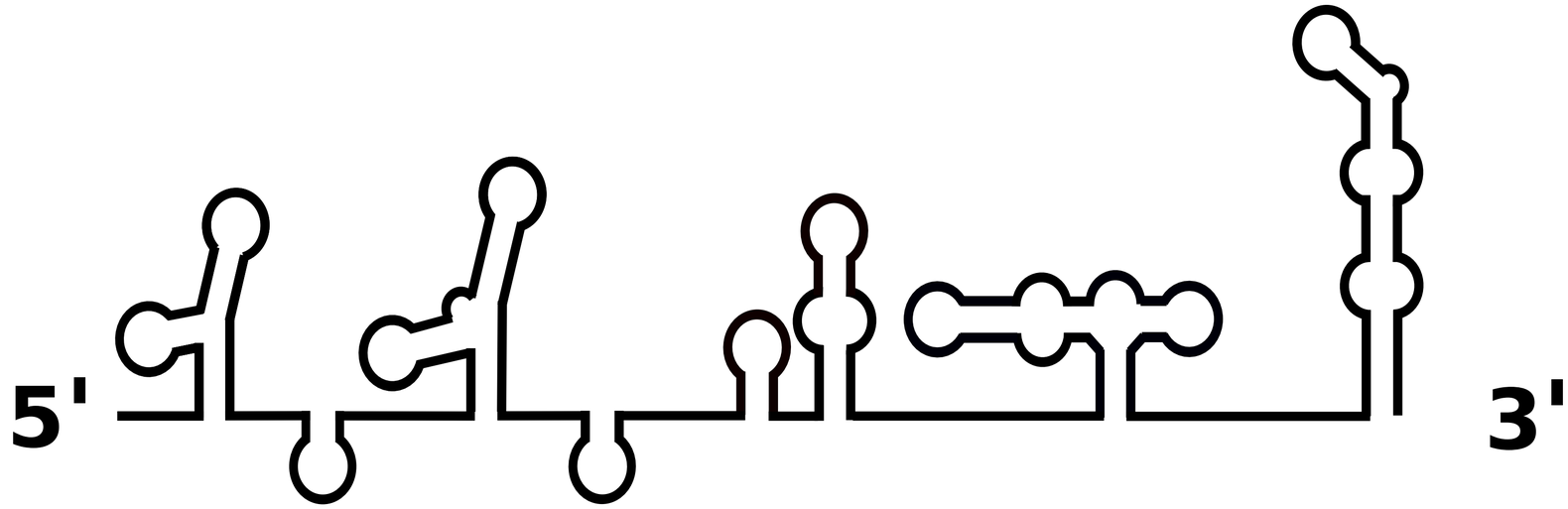
Flavivirus Genome Organization



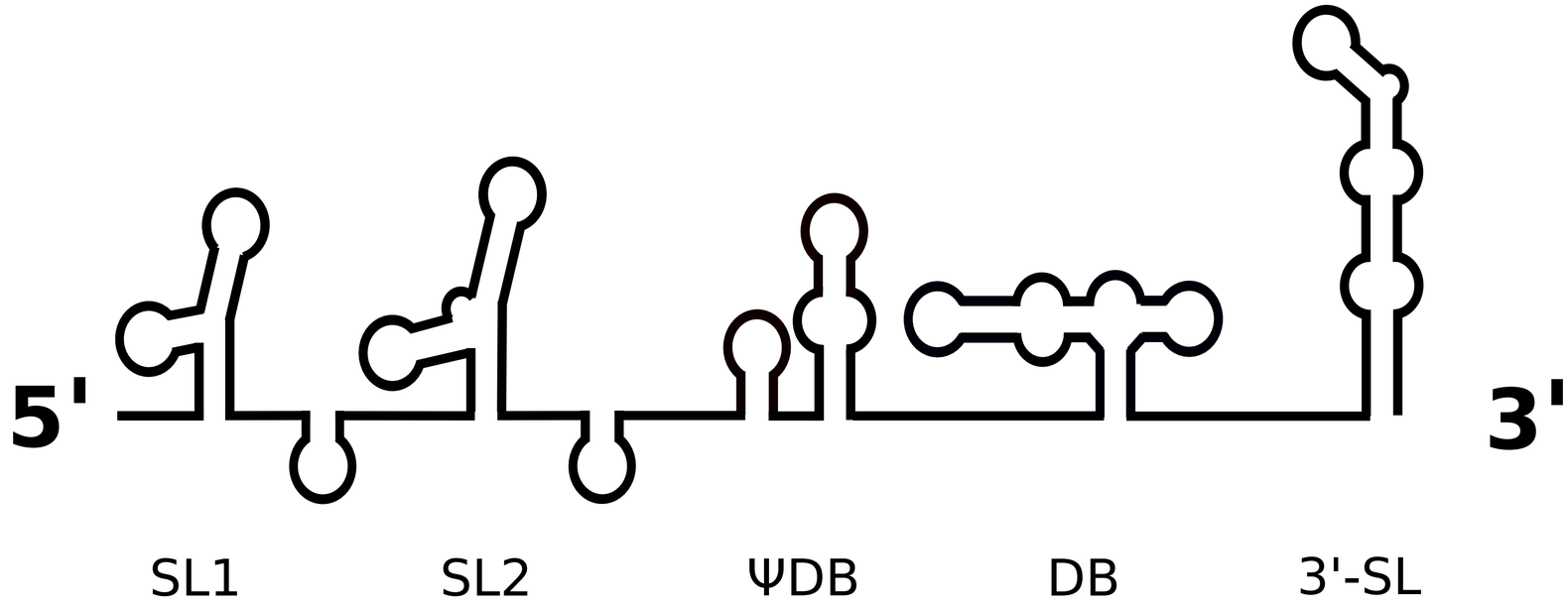
https://viralzone.expasy.org/

- Non-segmented, single-stranded, (+)-sense RNA genomes of 10-12kB length
- Capped, non-polyadenylated
- Encode a single ORF, flanked by highly structures 5’UTR and 3’UTR

Conserved RNAs in Viral 3’UTR: ZIKV

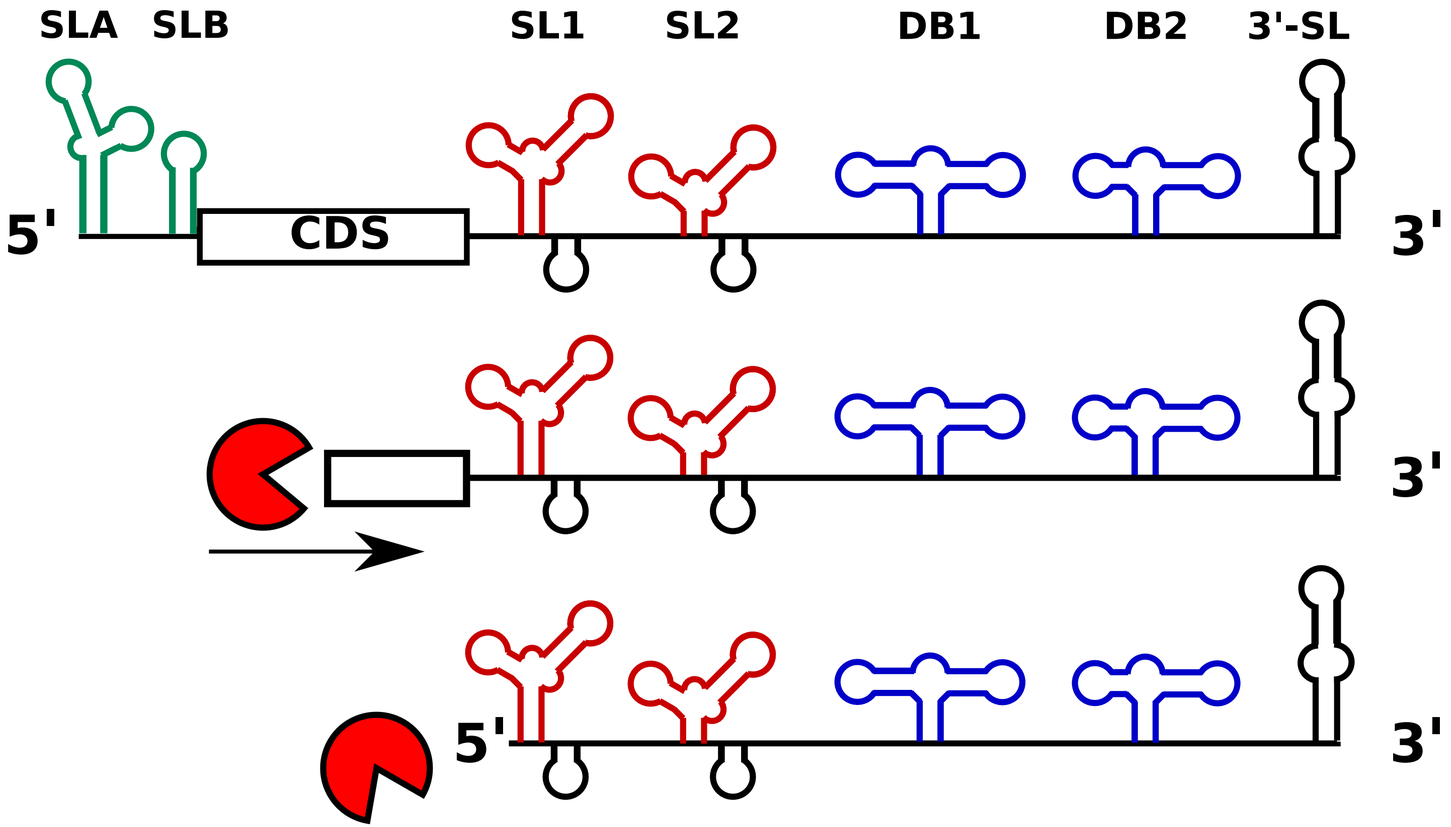


Conserved RNAs in Viral 3’UTR: ZIKV



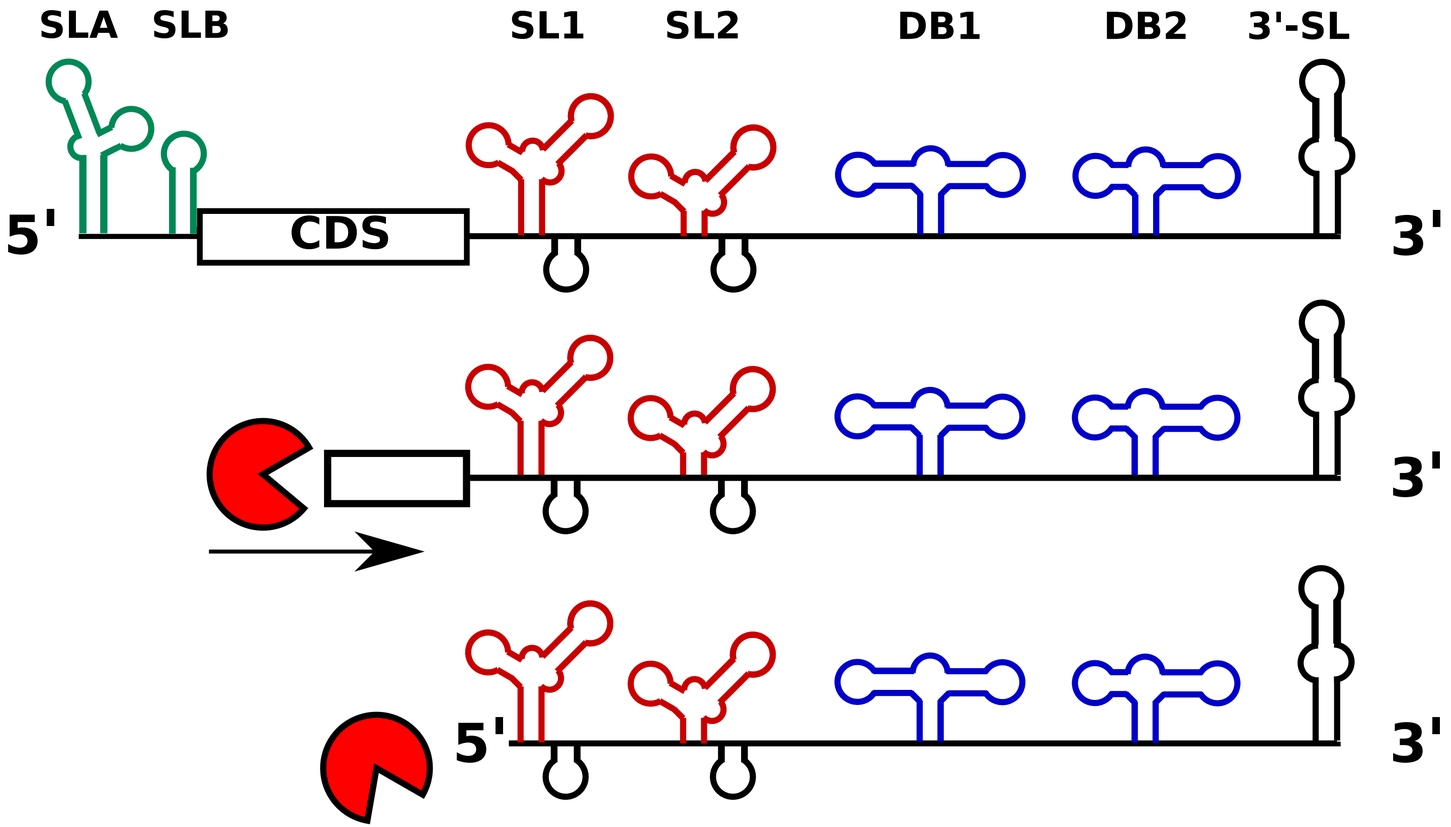
- Accumulation of short flavivirus RNA (sfRNA) upon infection
- Stable decay intermediates produced by partial exoribonuclease degradation
- Xrn1 is efficiently stalled at conserved xrRNA structures

Conserved RNAs in Viral 3'UTR



- Accumulation of short flavivirus RNA (sfRNA) upon infection
- Stable decay intermediates produced by partial exoribonuclease degradation
- Xrn1 is efficiently stalled at conserved xrRNA structures

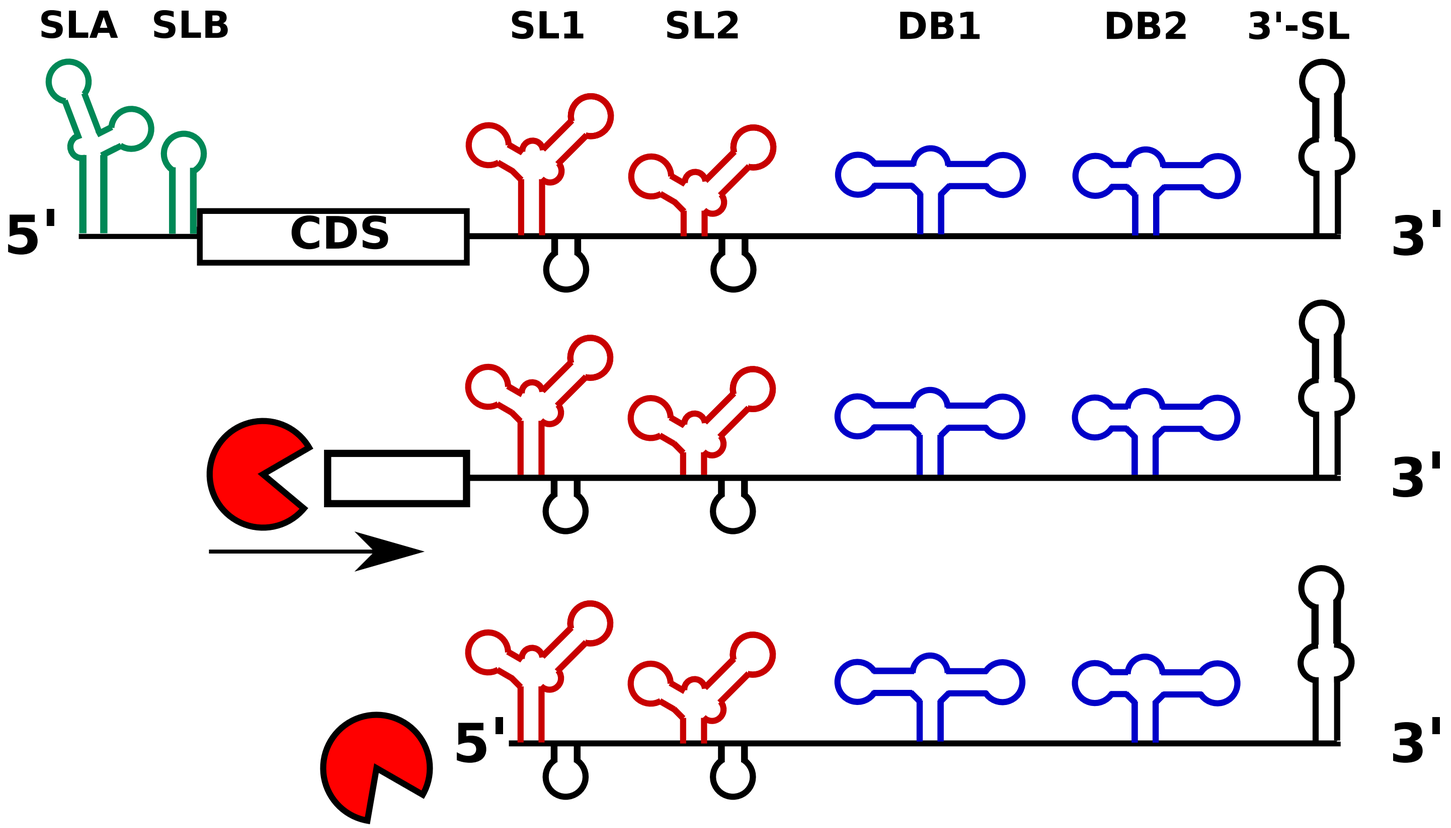
Conserved RNAs in Viral 3'UTR



sfRNA mediates pathogenicity !

- Accumulation of short flavivirus RNA (sfRNA) upon infection
- Stable decay intermediates produced by partial exoribonuclease degradation
- Xrn1 is efficiently stalled at conserved xrRNA structures

Conserved RNAs in Viral 3'UTR

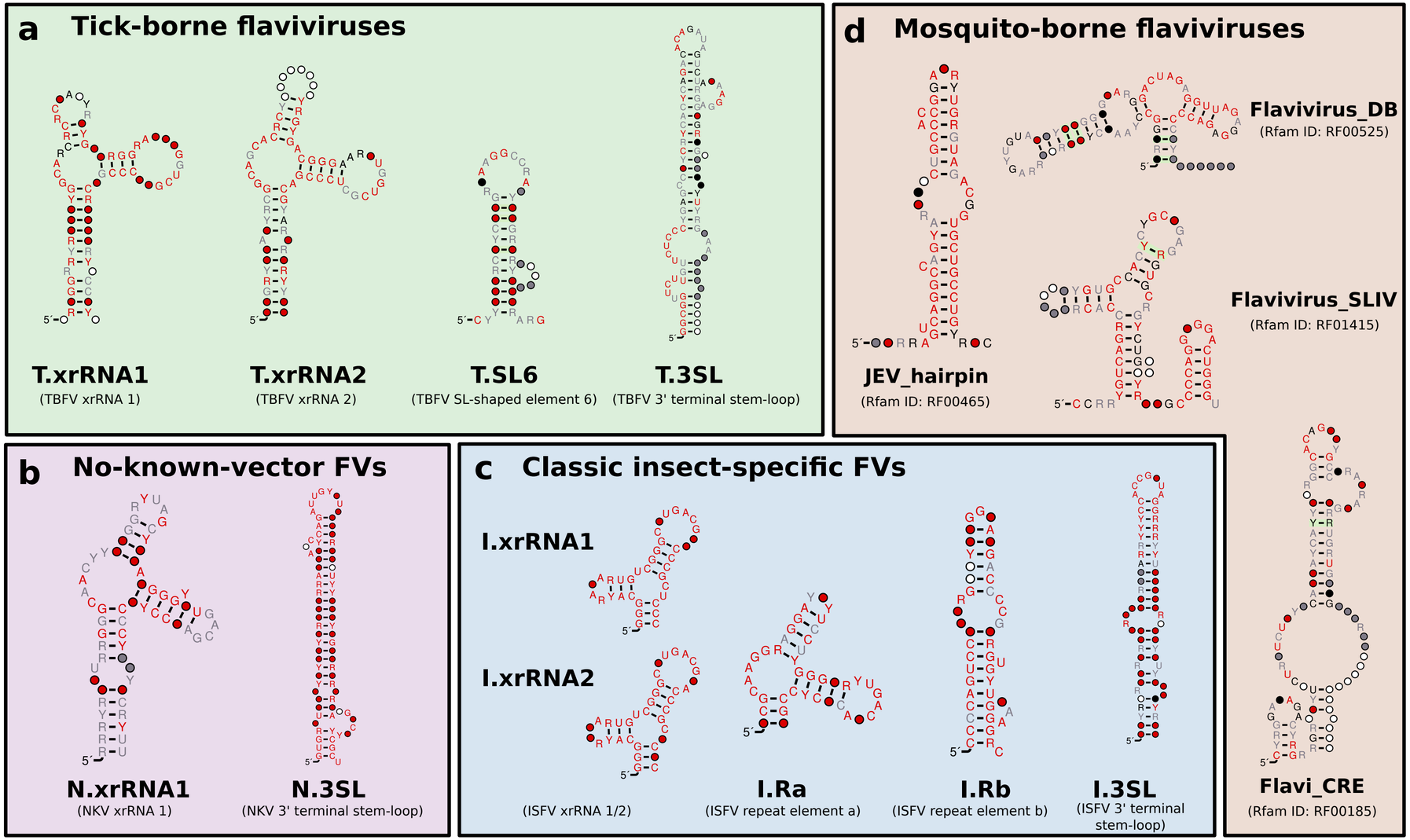


sfRNA mediates pathogenicity !

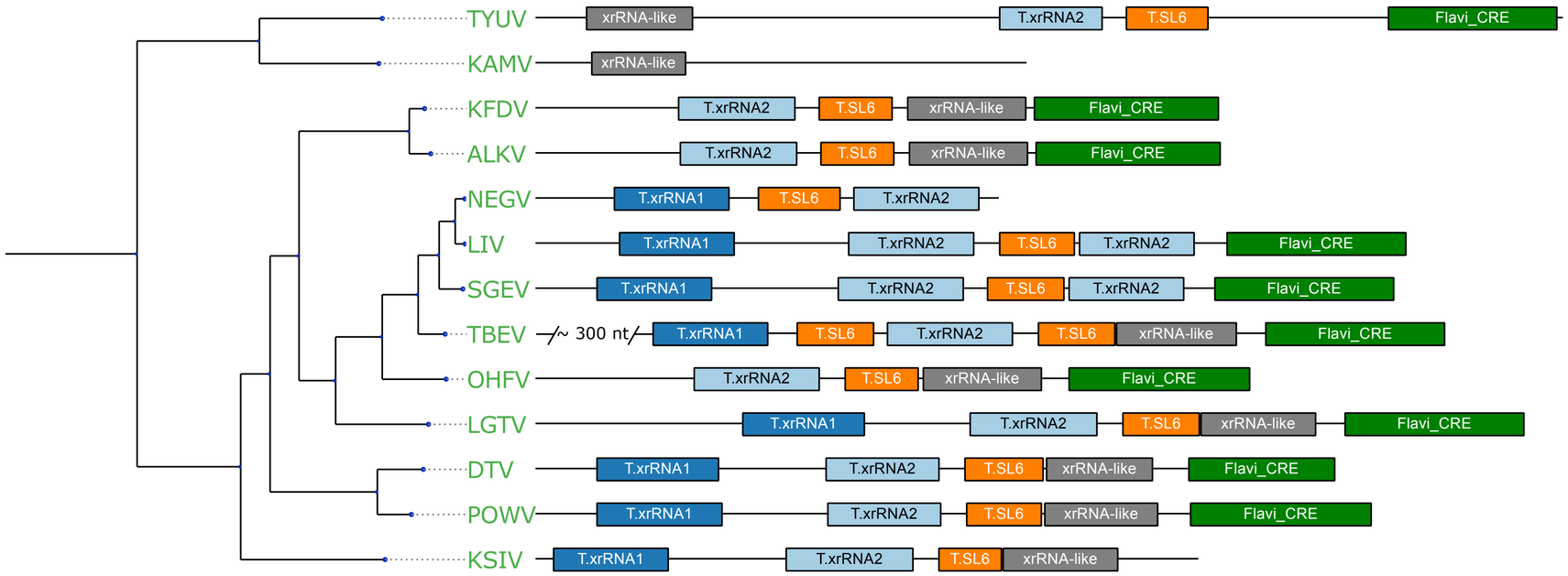
How to find them?

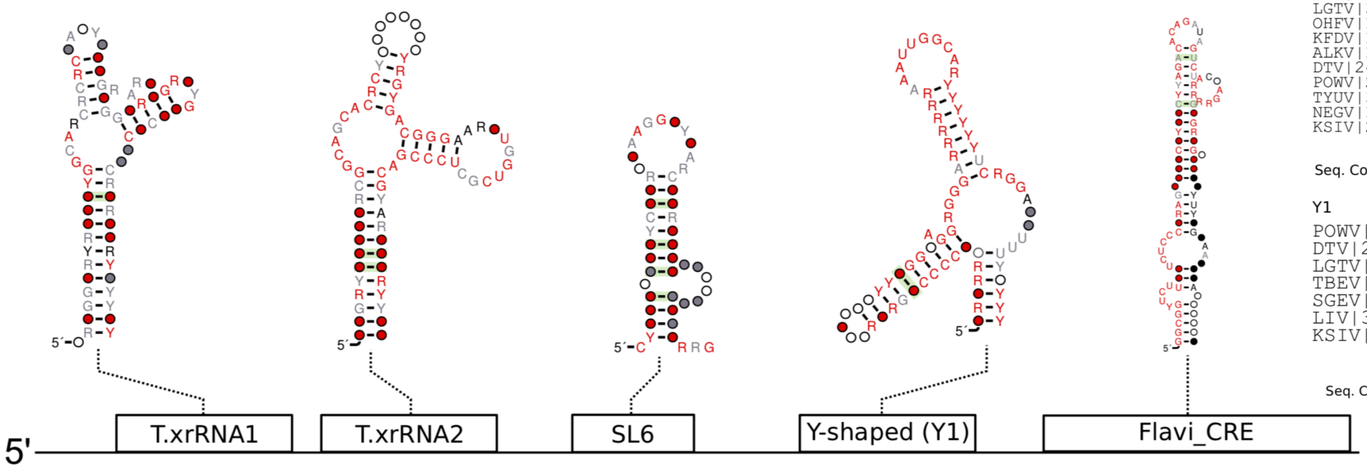
- Accumulation of short flavivirus RNA (sfRNA) upon infection
- Stable decay intermediates produced by partial exoribonuclease degradation
- Xrn1 is efficiently stalled at conserved xrRNA structures

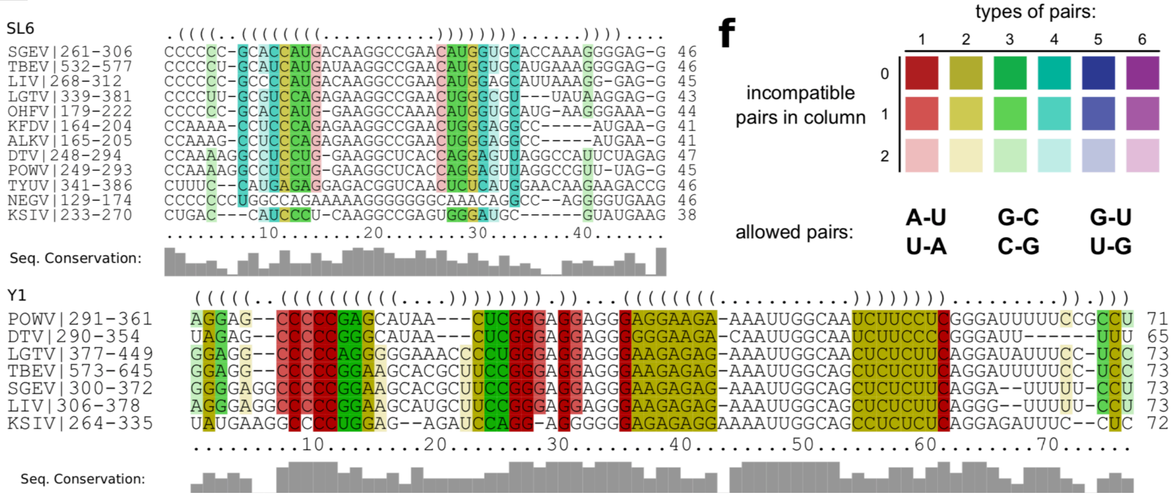
Consensus Structures & Covariance Models



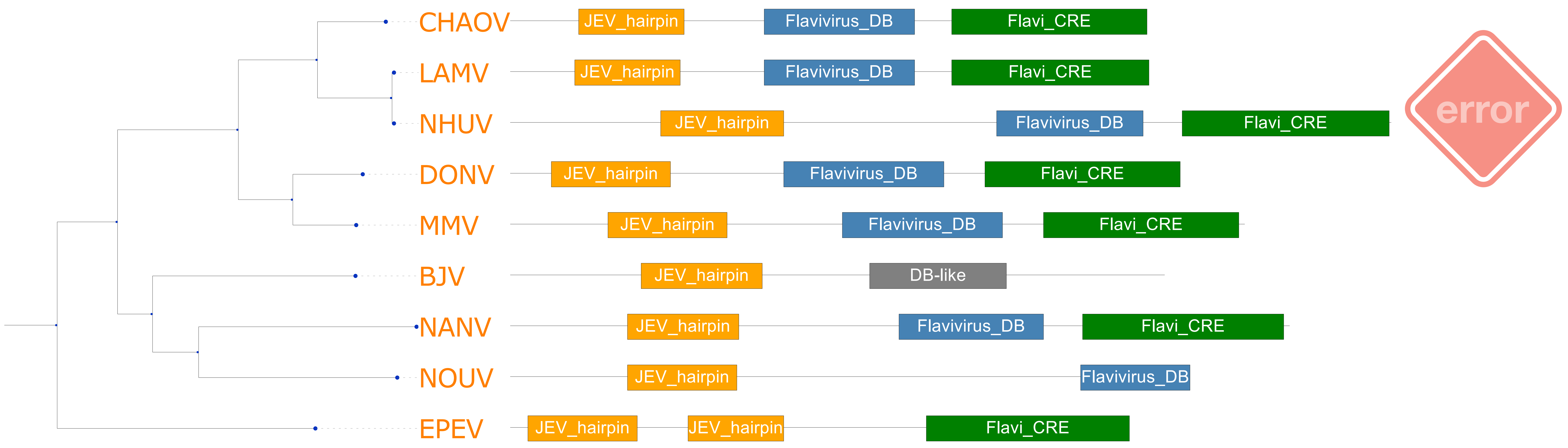
Tick-Borne Flaviviruses



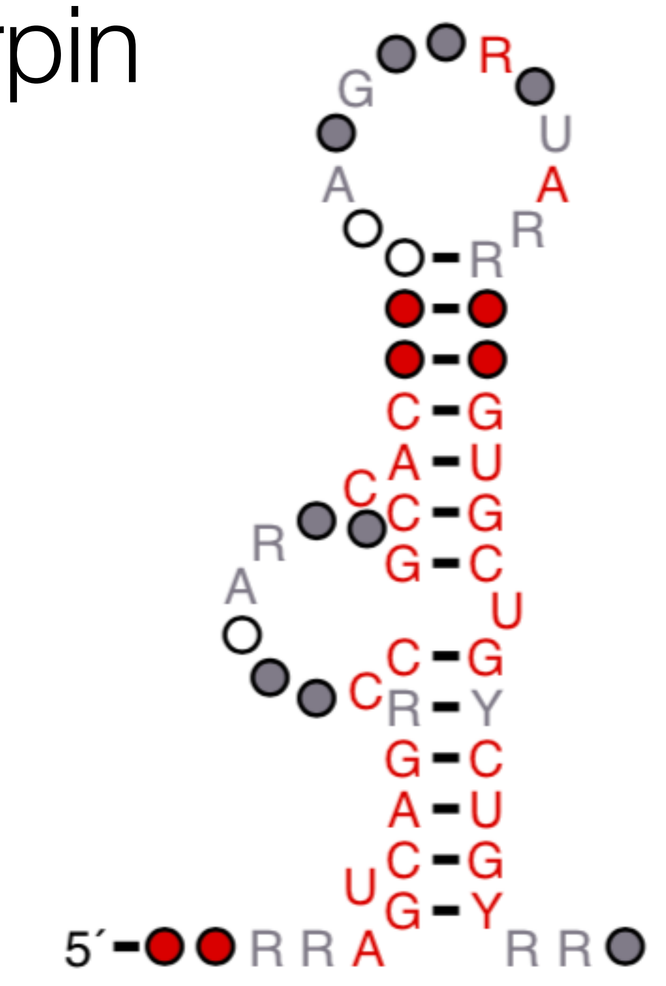


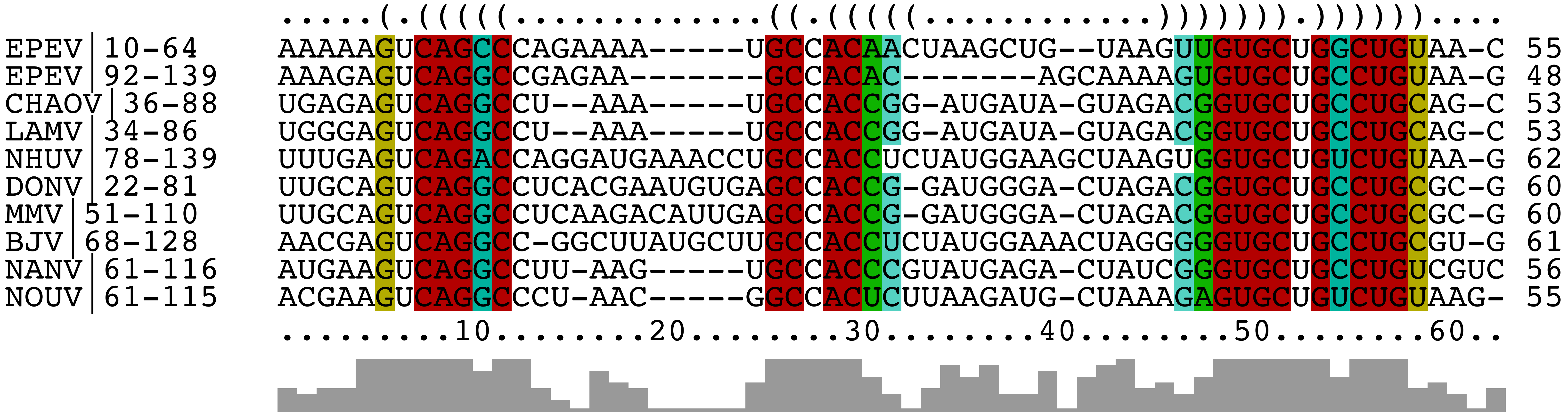


Dual-Host Affiliated Insect-Specific FVs

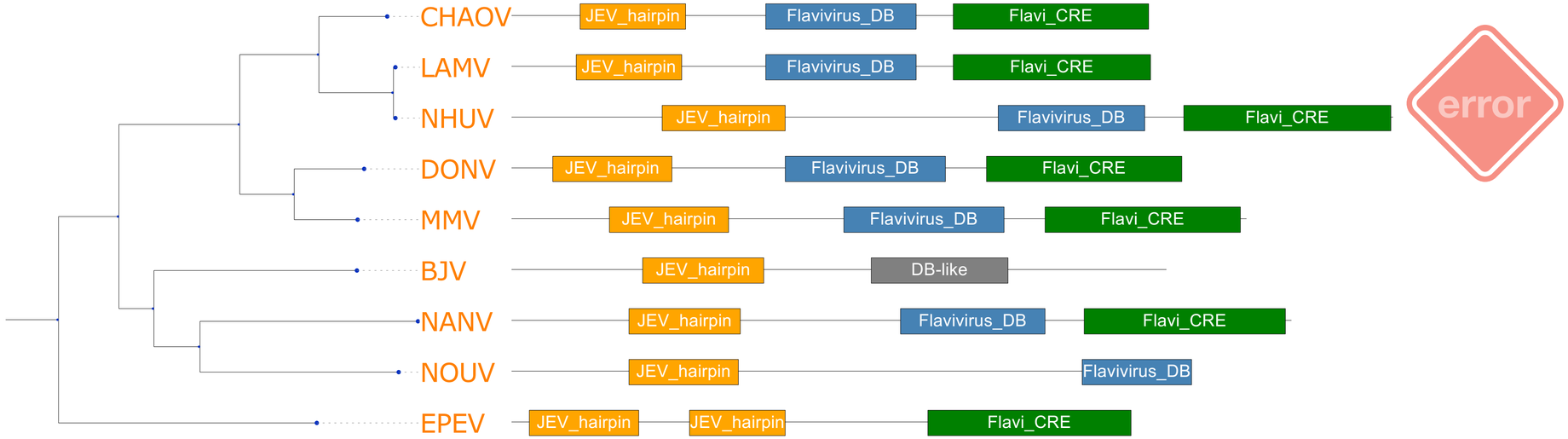


JEV_hairpin

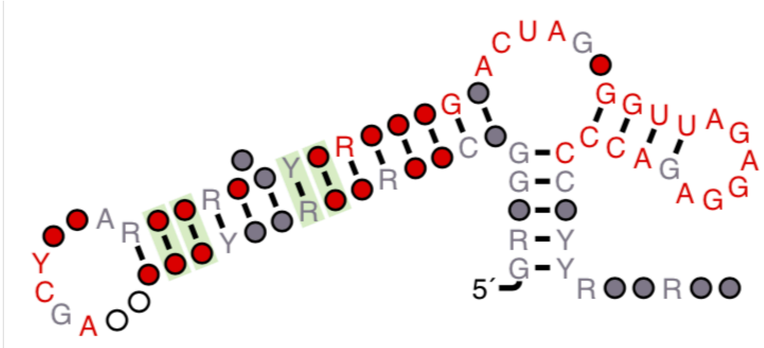


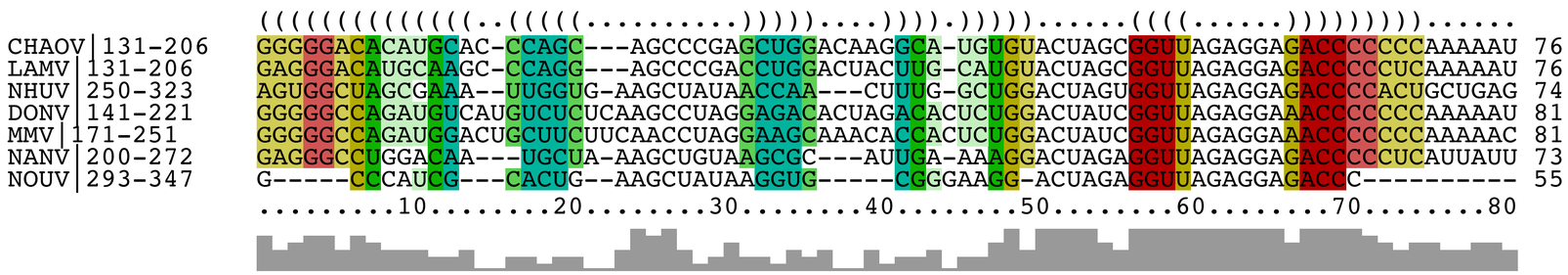


Dual-Host Affiliated Insect-Specific FVs

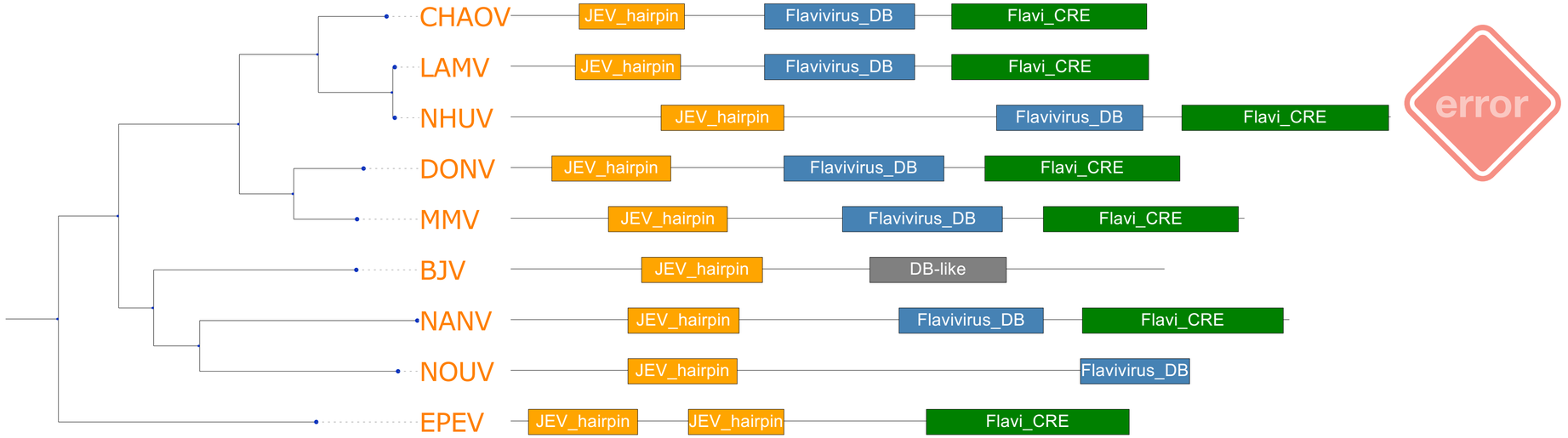


Flavivirus_DB

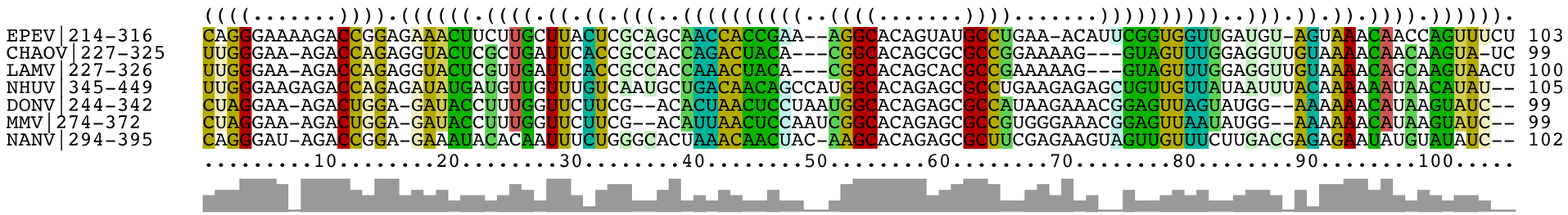


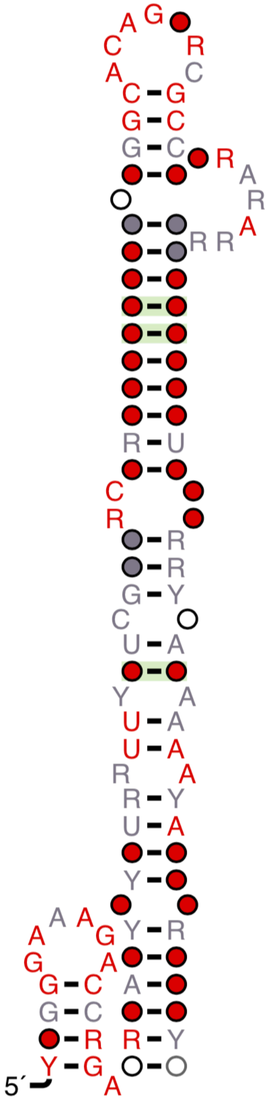


Dual-Host Affiliated Insect-Specific FVs

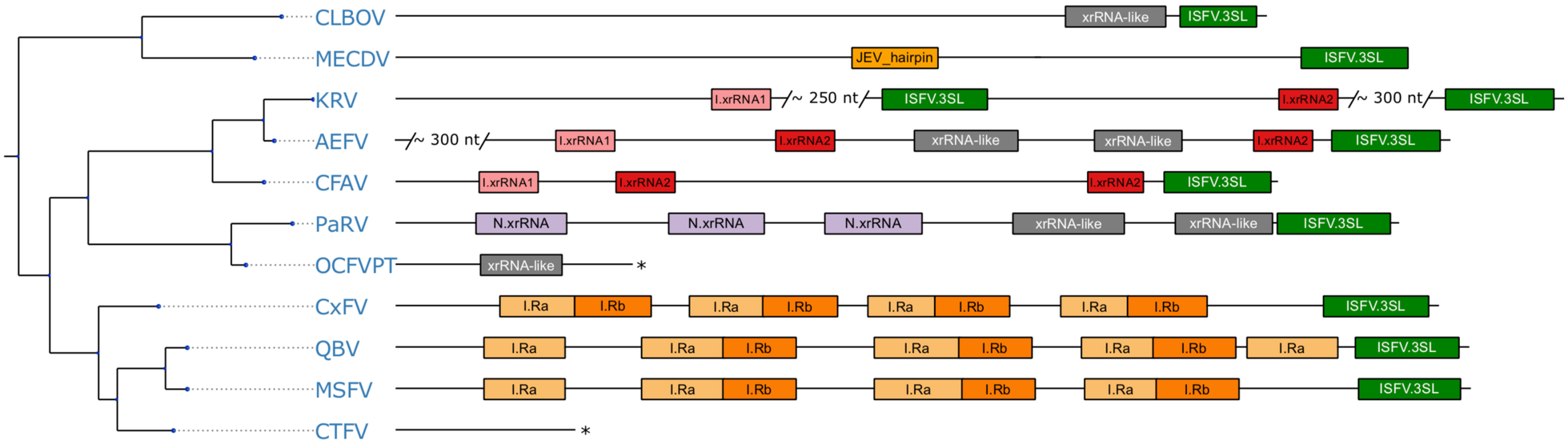


Flavi_CRE

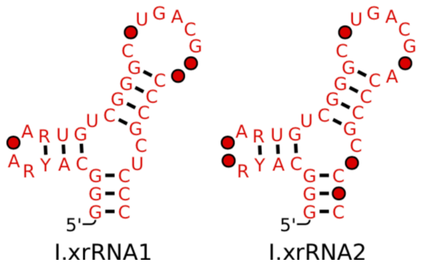


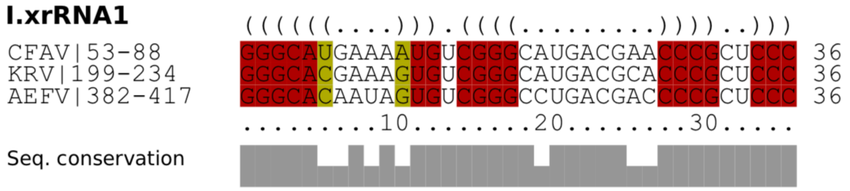


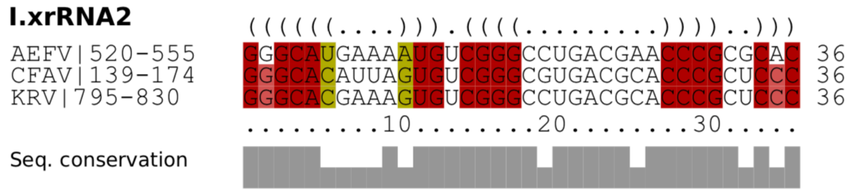
Classic Insect-Specific FVs



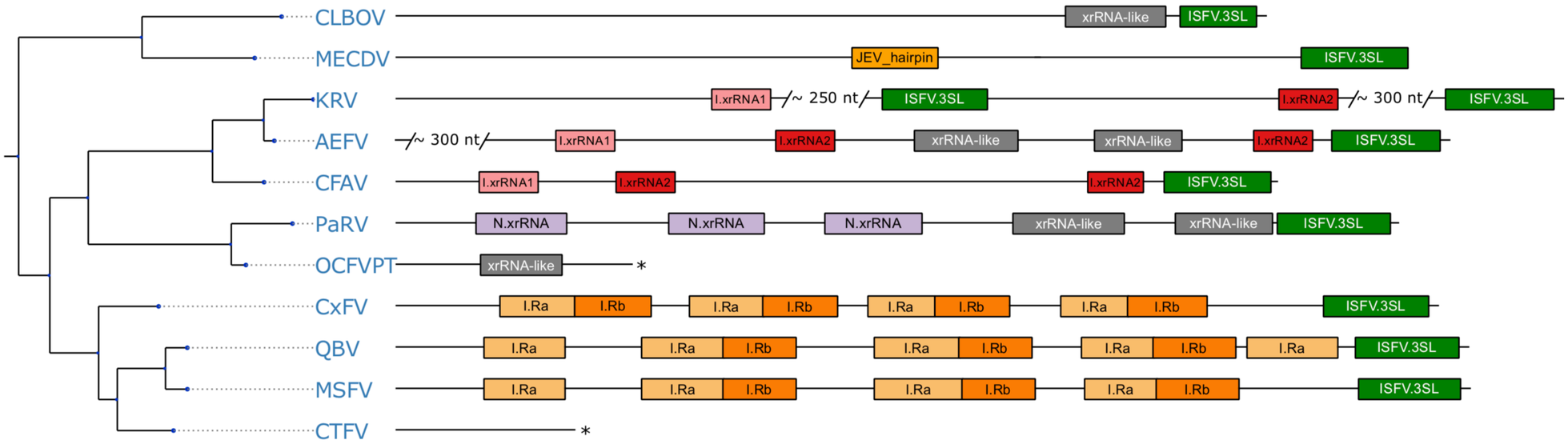
KRV / AEFV / CFAV



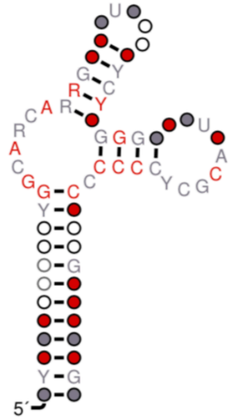


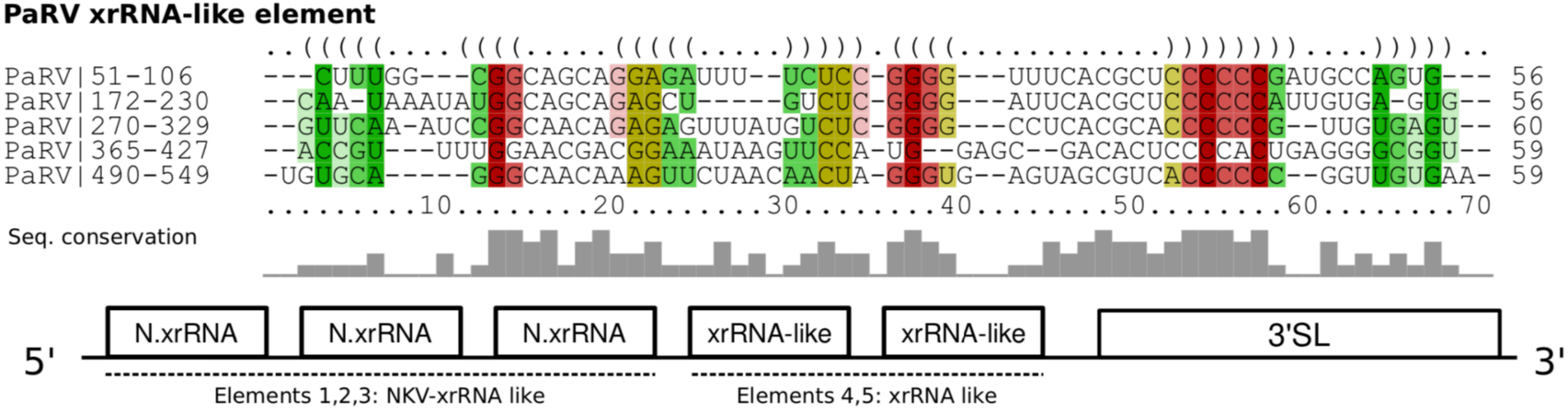


Classic Insect-Specific FVs

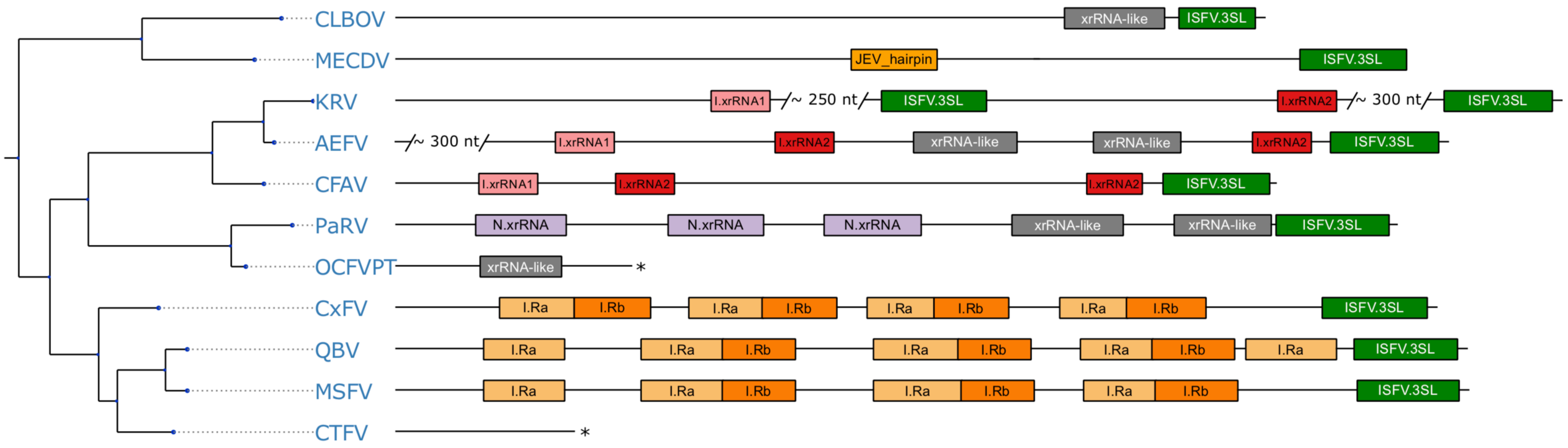


PaRV

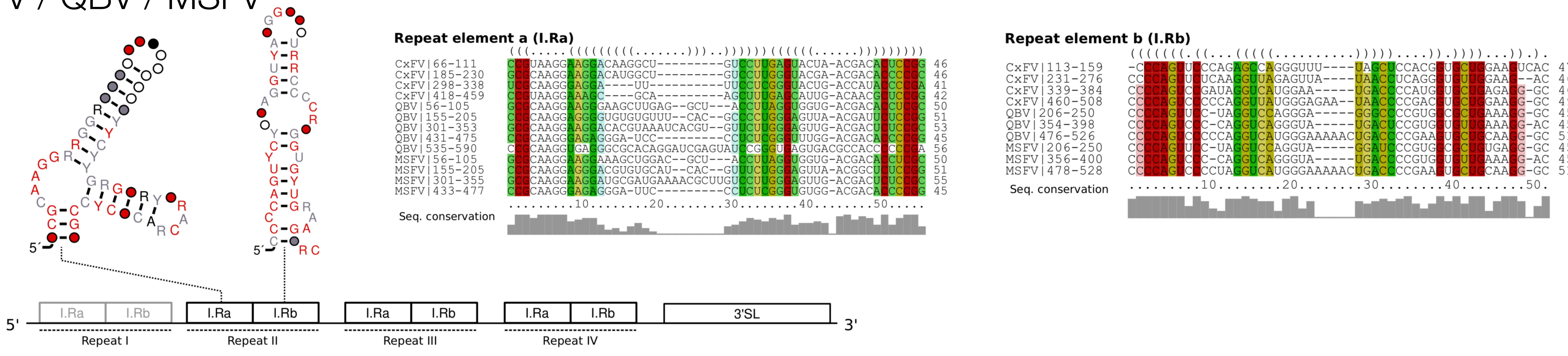




Classic Insect-Specific FVs



CxFV / QBV / MSFV

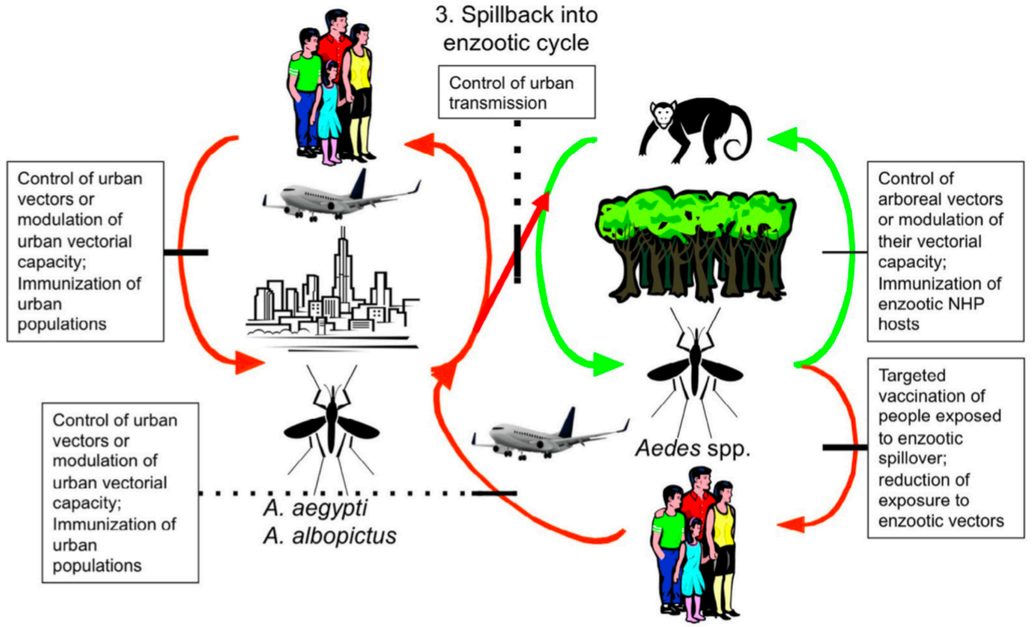


Part II:

Conserved RNA structures in alphaviruses

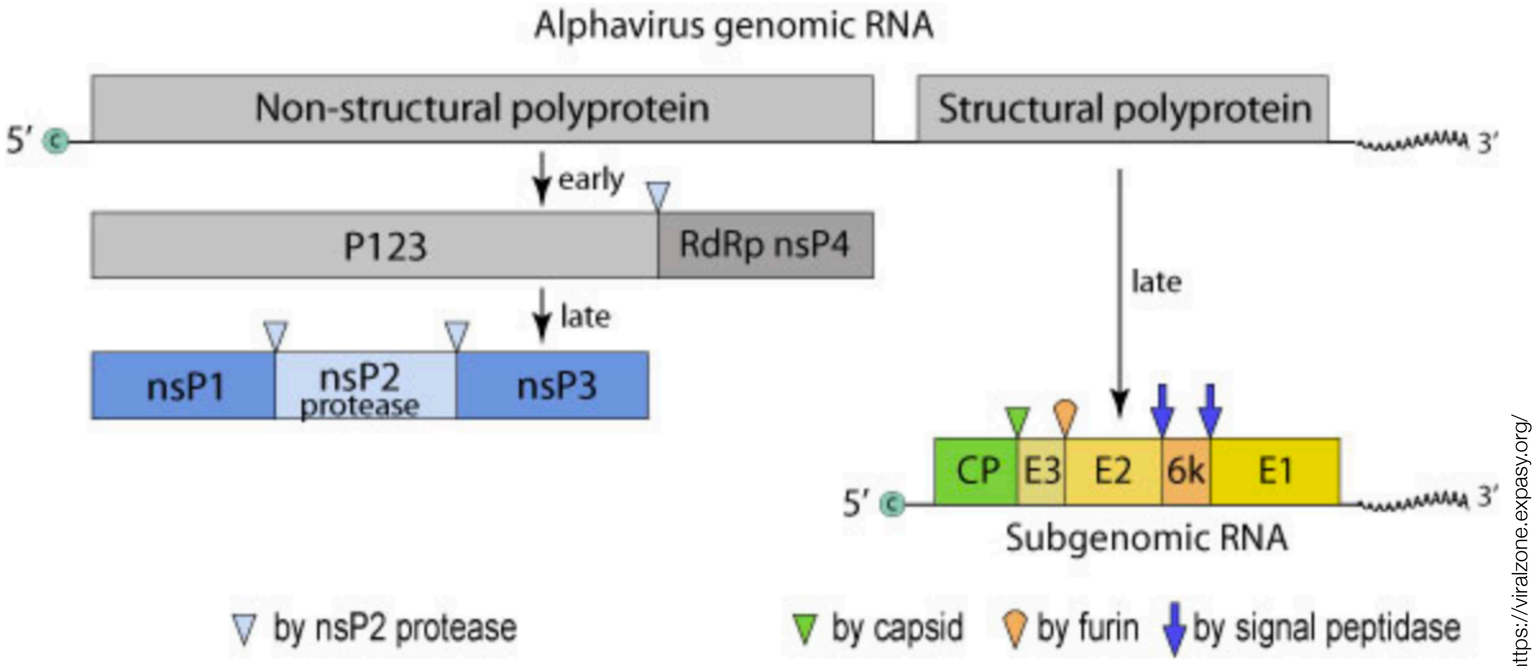
Chikungunya Virus (CHIKV)

- Family *Togaviridae* / genus *Alphavirus*; mosquito-borne (*Aedes* spp.)
- Single-stranded (+) sense RNA virus
- Chikungunya fever: febrile illness, arthralgia, rash, rarely causes hemorrhagic complications
- Enzootic in tropical and subtropical regions of Africa
- First outbreak described 1952 in Tanzania
- No vaccine available



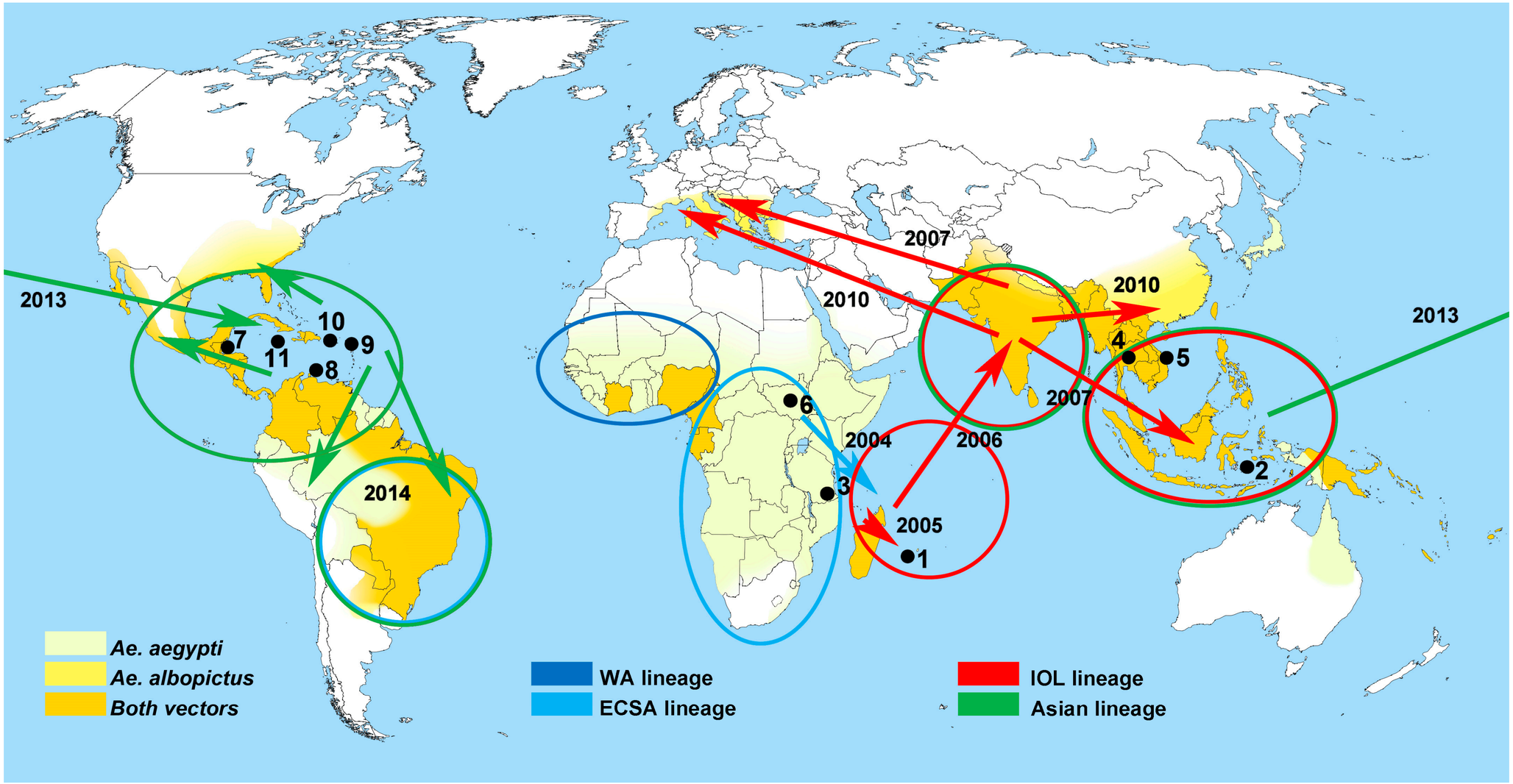
Weaver and Forrester, Antivir. Res. (2015)

Alphavirus Genome Organization



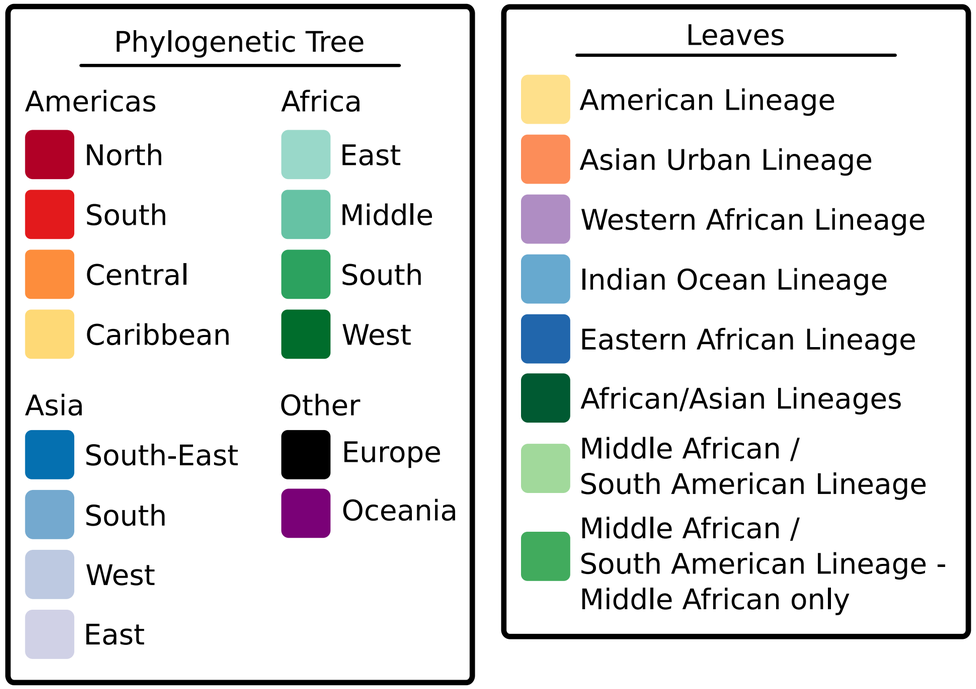
- Non-segmented, single-stranded, (+)-sense RNA genomes of 11-12kB length
- Capped and polyadenylated
- AV genomes appear to host cells as mRNA for immediate translation upon entry into the cytoplasm

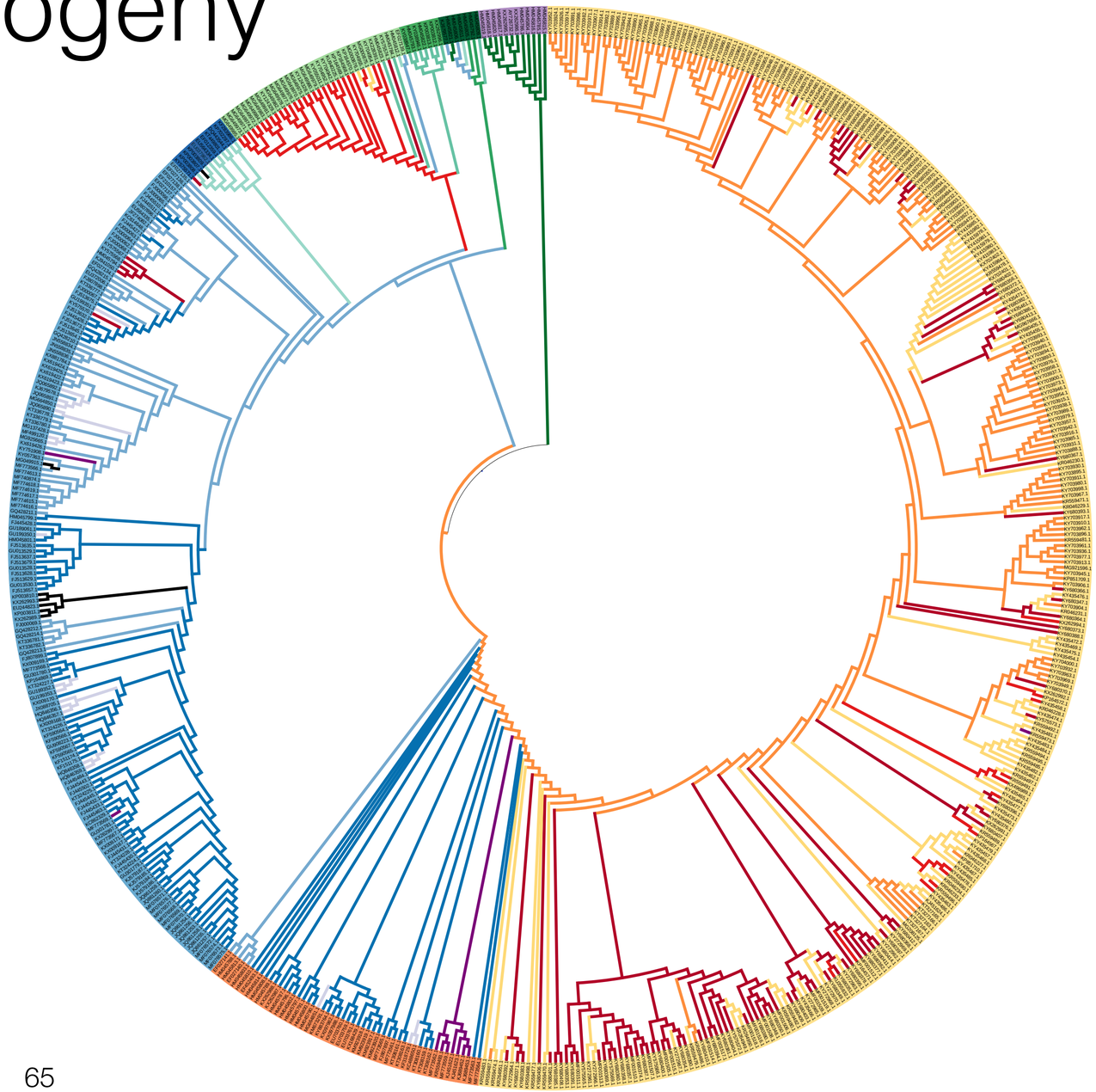
CHIKV Epidemic Spread



Frickmann et al., Viruses (2019)

Updated CHIKV Phylogeny





- 590 CHIKV genomes
- iq-tree / SH-aLRT

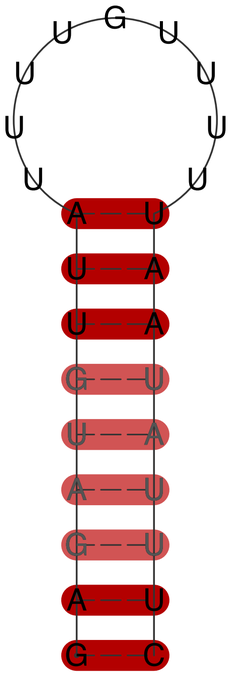
CHIKV 3’UTR

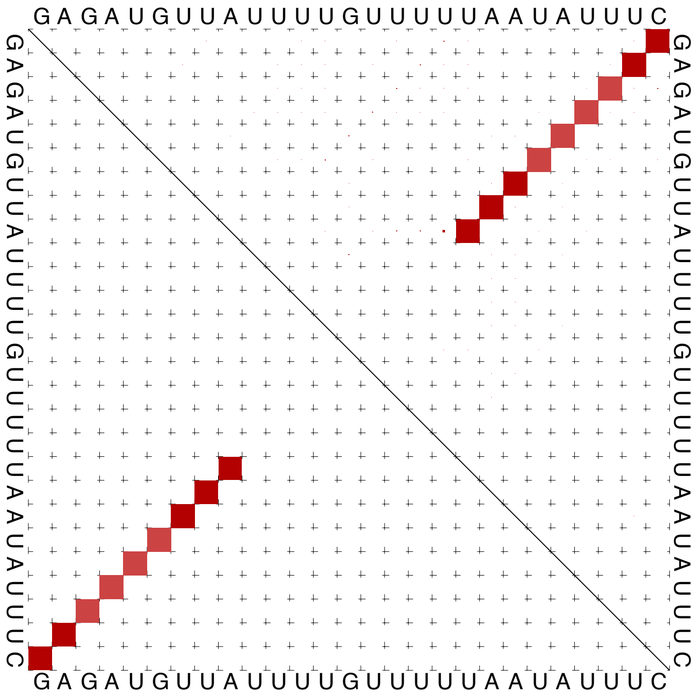
- 110 full length CHIKV 3’UTRs
- Variable length (510nt — 930nt)
- Structural alignments + Covariance models
- Thermodynamic modelling based on ViennaRNA Package

CHIKV 3’UTR: Conserved RNA Structures

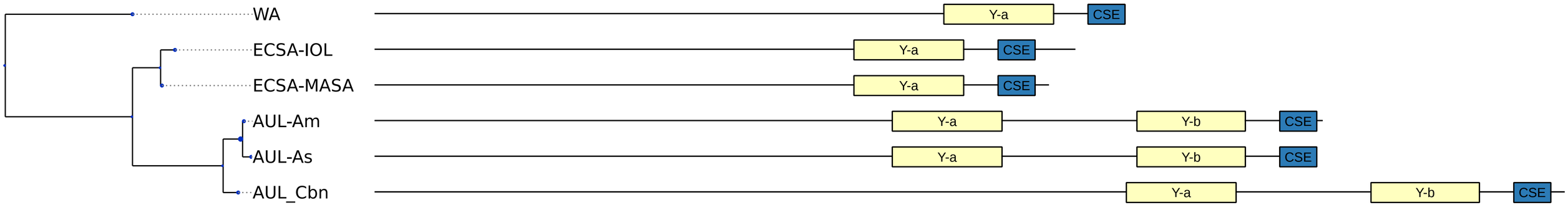
Conserved Sequence Element (CSE)

```
                                 (((((((((.........)))))))))
KF318729.1/688-714   GAGAUGUUAUUUUGUUUUUAAUAUUUC 27
KT327163.2/862-888   GAGAUGUUAUUUUGUUUUUAAUAUUUC 27
JF274082.1/472-498   GAGAUGUUAUUUUGUUUUUAAUAUUUC 27
KY038946.1/472-498   GAGAUGUUAUUUUGUUUUUAAUAUUUC 27
AY726732.1/540-566   GAGAUGUUAUUUUGUUUUUAAUAUUUC 27
MF001517.1/685-707   GAGAUGUUAUUUUGUUUUUAA----UC 23
                     .........10........20......
```

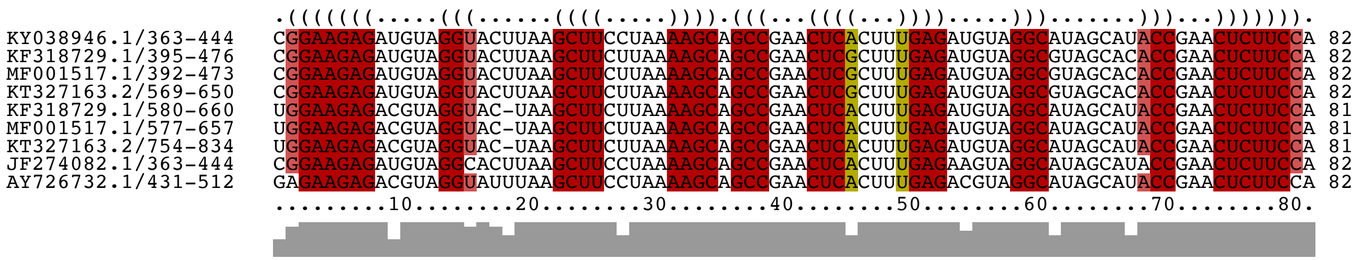


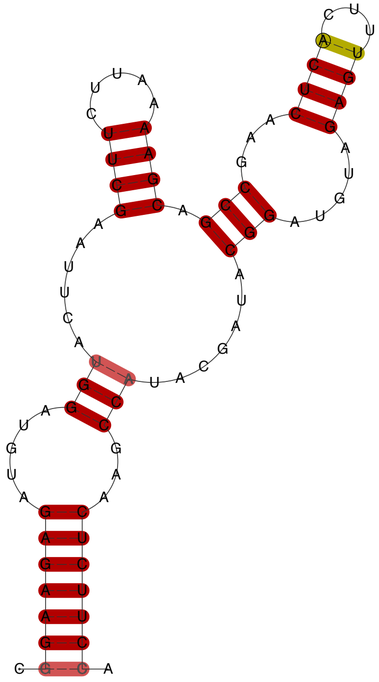


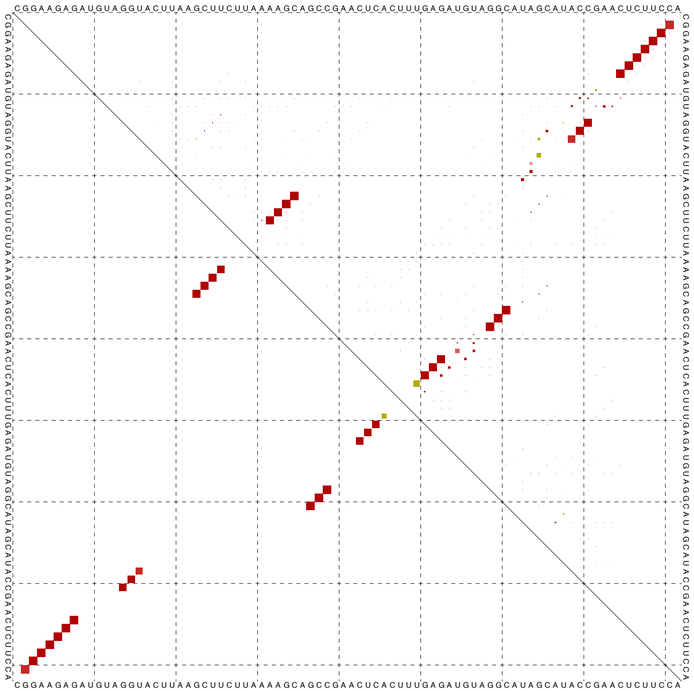
CHIKV 3’UTR: Conserved RNA Structures



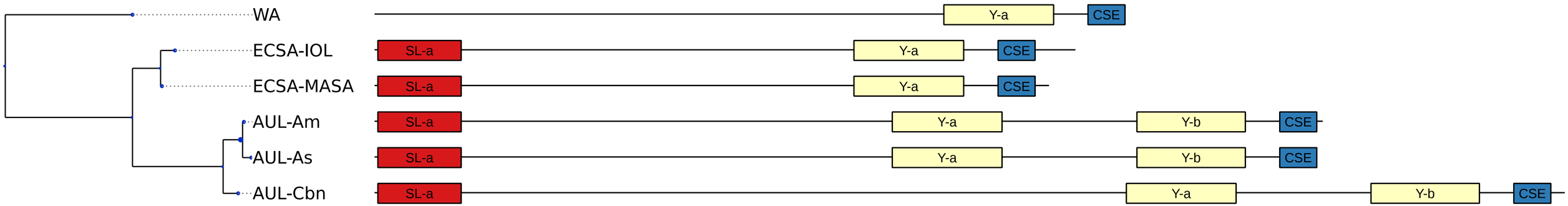
Y-shaped Element (SL-Y)



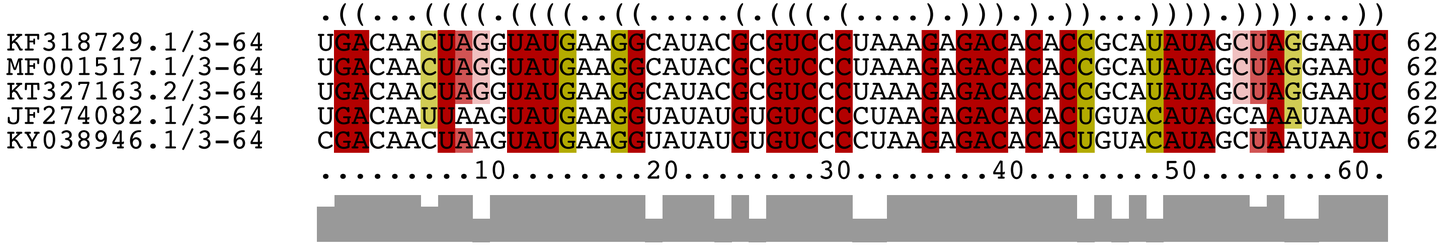


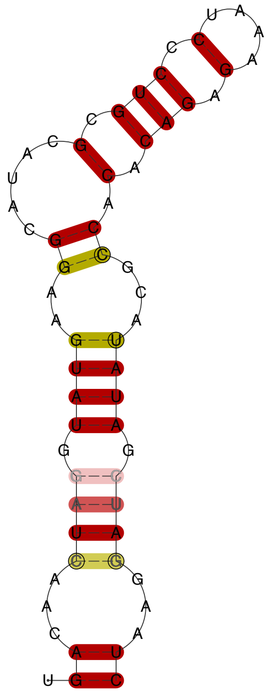


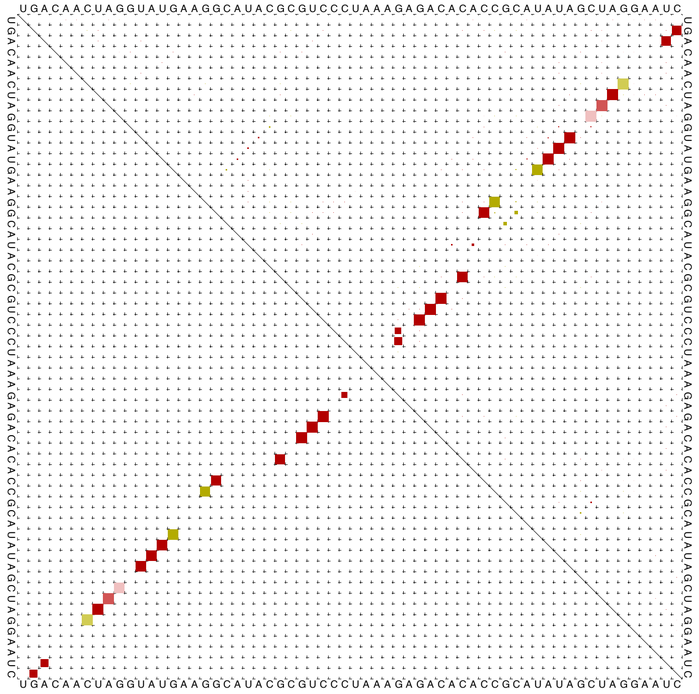
CHIKV 3’UTR: Conserved RNA Structures



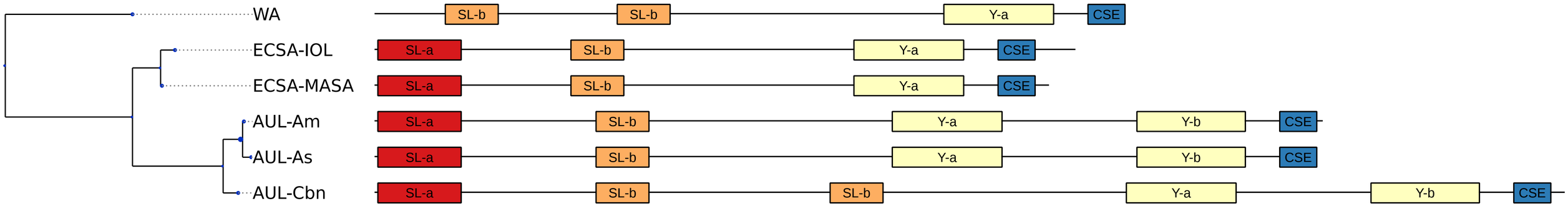
Stem-loop a (SL-a)



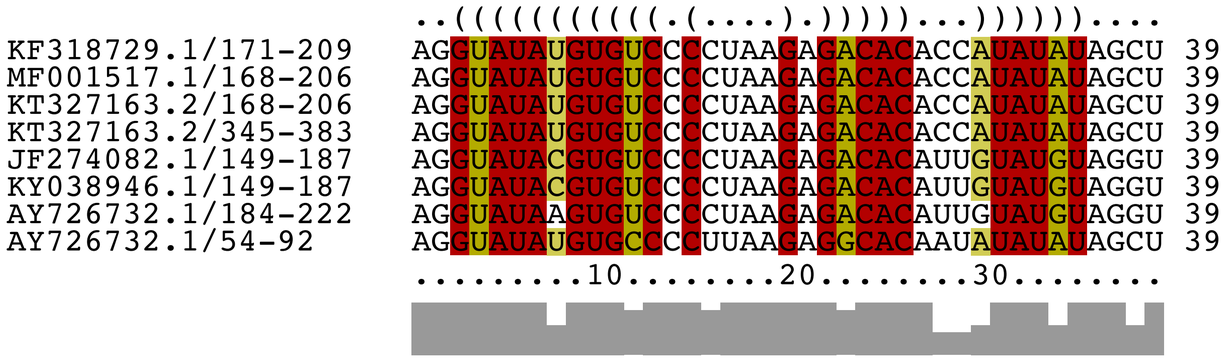


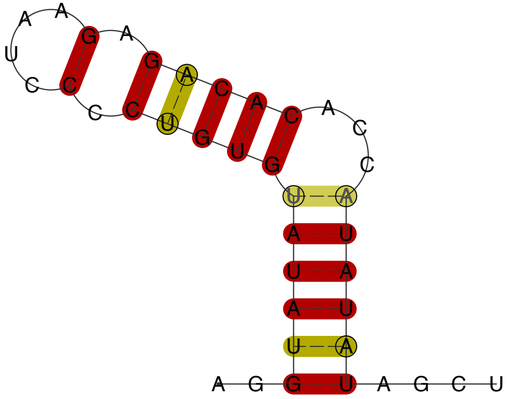


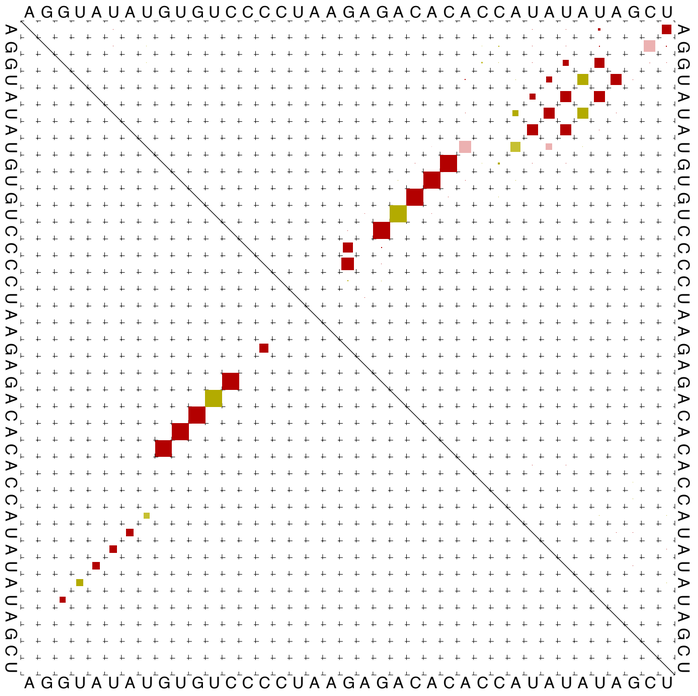
CHIKV 3’UTR: Conserved RNA Structures



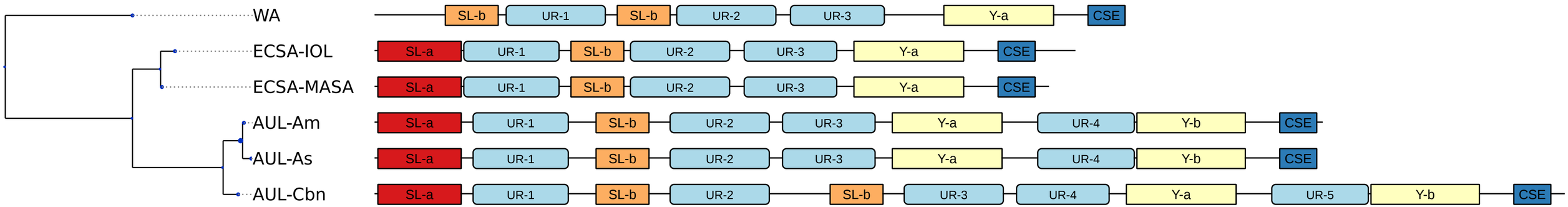
Stem-loop b (SL-b)





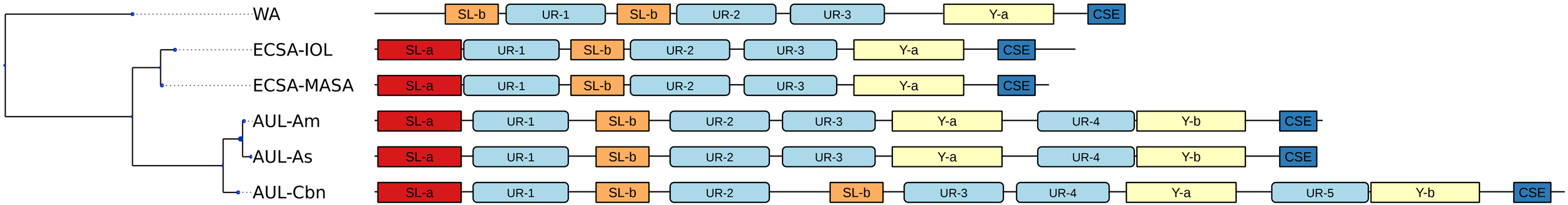


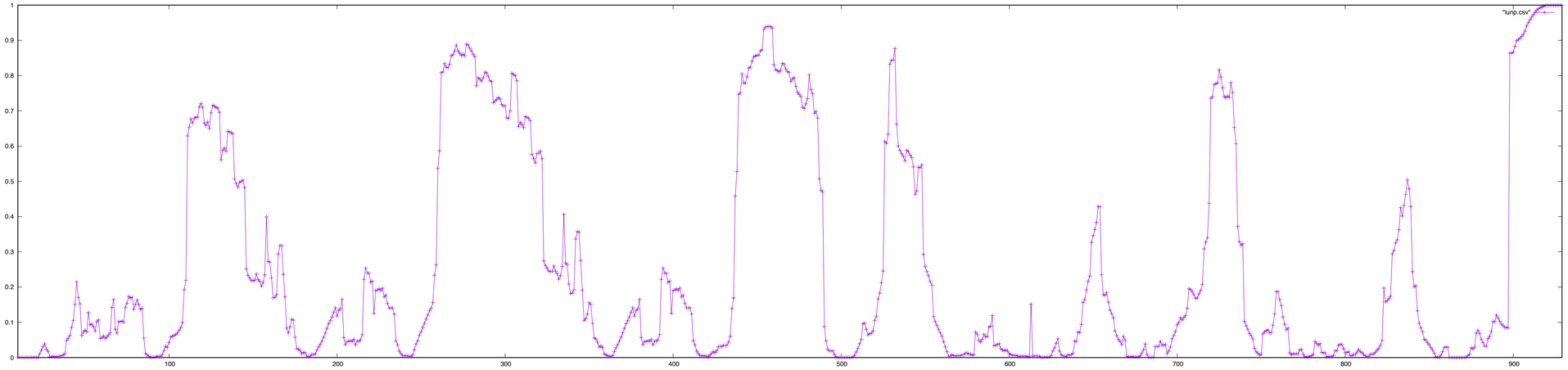
CHIKV 3’UTR: Conserved RNA Elements



Unstructured repeat (UR)

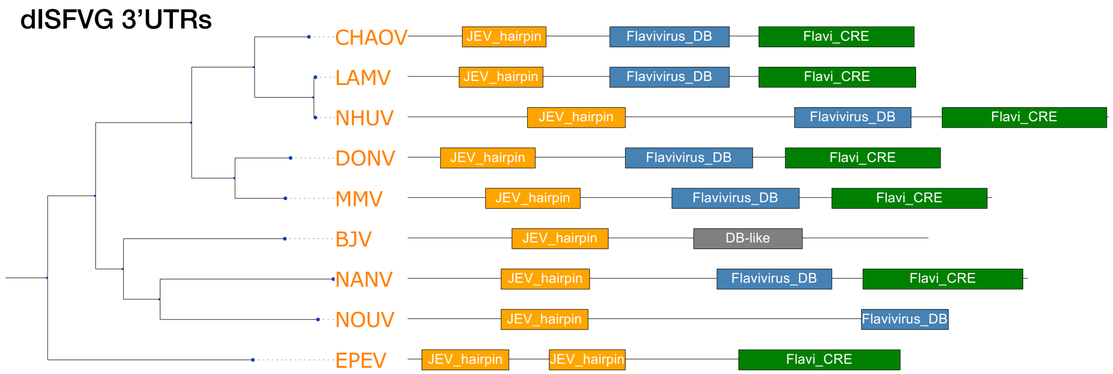
CHIKV-Cbn Accessibility

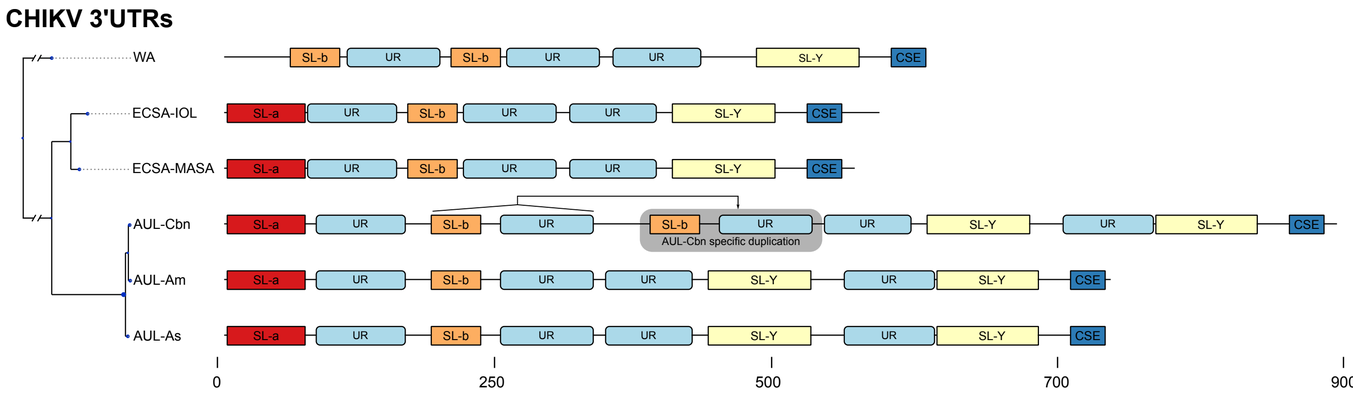




Summary

- Functional RNA structures are evolutionarily conserved
- Structural alignments and covariance models
- Varied architecture of conserved RNAs in flavivirus UTRs
- Lineage-specific non-coding RNA architecture in Alphaviruses
- Tools freely available through ViennaRNA Package





Acknowledgements

TBI Vienna

Roman Ochsenreiter
Ronny Lorenz
Ivo L. Hofacker

Univ. San Diego

Adriano de Bernardi Schneider

Functional RNA Structures in the 3′ UTR of Tick-Borne, Insect-Specific and No-Known-Vector Flaviviruses. *Viruses* 11.3 (2019): 298. **https://doi.org/10.3390/v11030298**

Updated phylogeny of Chikungunya virus suggests lineage-specific RNA architecture. *Viruses* 11.9 (2019): 798. **https://doi.org/10.3390/v11090798**

Musashi binding elements in Zika and related Flavivirus 3′ UTRs: A comparative study in silico. *Scientific reports* 9.1 (2019): 6911. **https://doi.org/10.1038/s41598-019-43390-5**

Twitter: @mtwolfinger