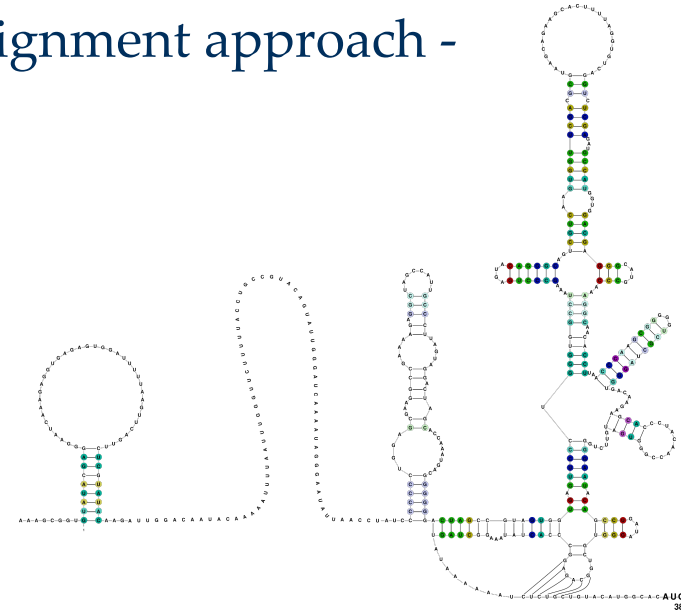


RNA secondary structures prediction in viruses

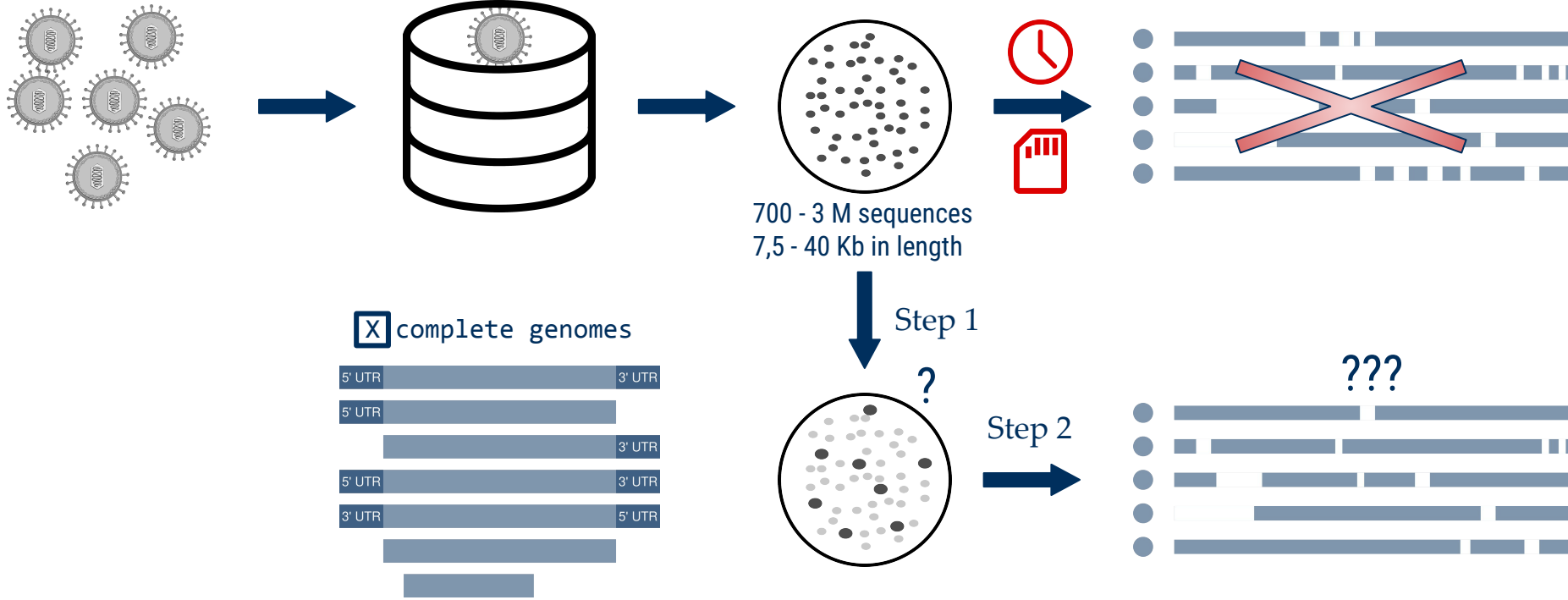
- a full-genome multiple sequence alignment approach -

Sandra Triebel

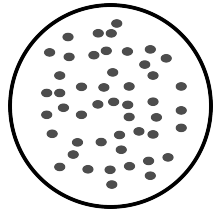
RNA Bioinformatics & High-Throughput Analysis
Friedrich Schiller University Jena, Germany



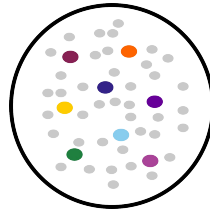
Working with viral genomes...



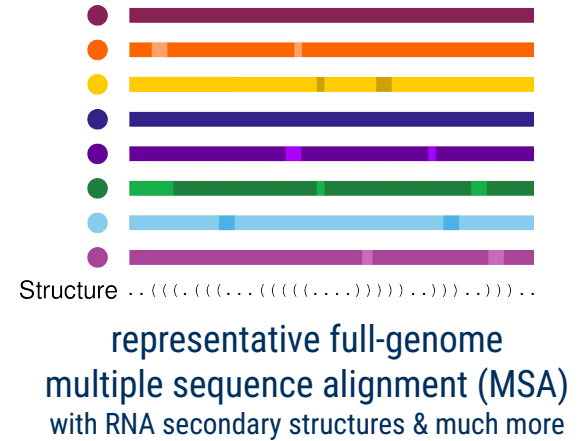
The rough workflow



original
data set



representative
genomes

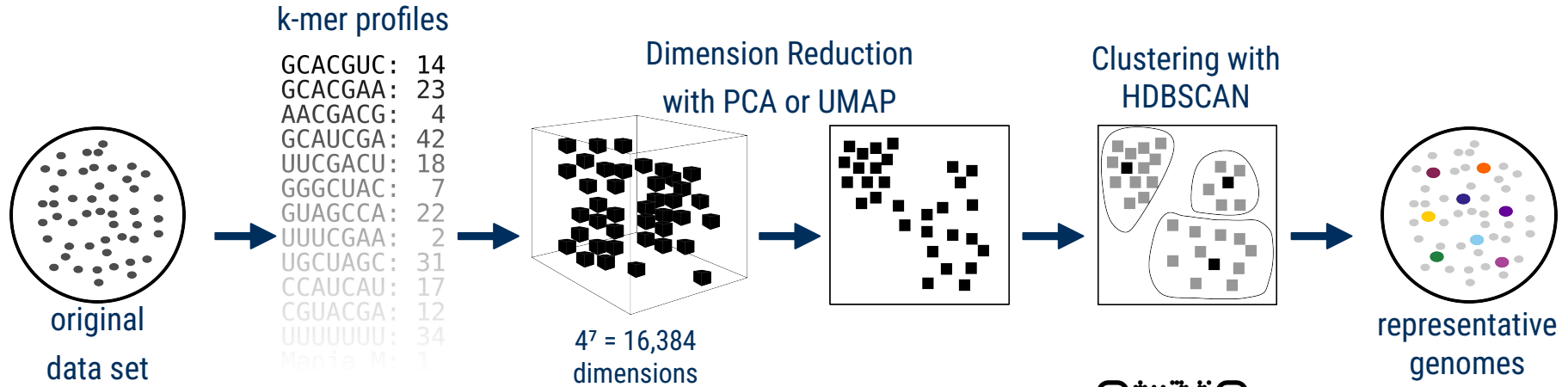


- Step 1 - Representative Genomes

Representative genomes



Kevin Lamkiewicz

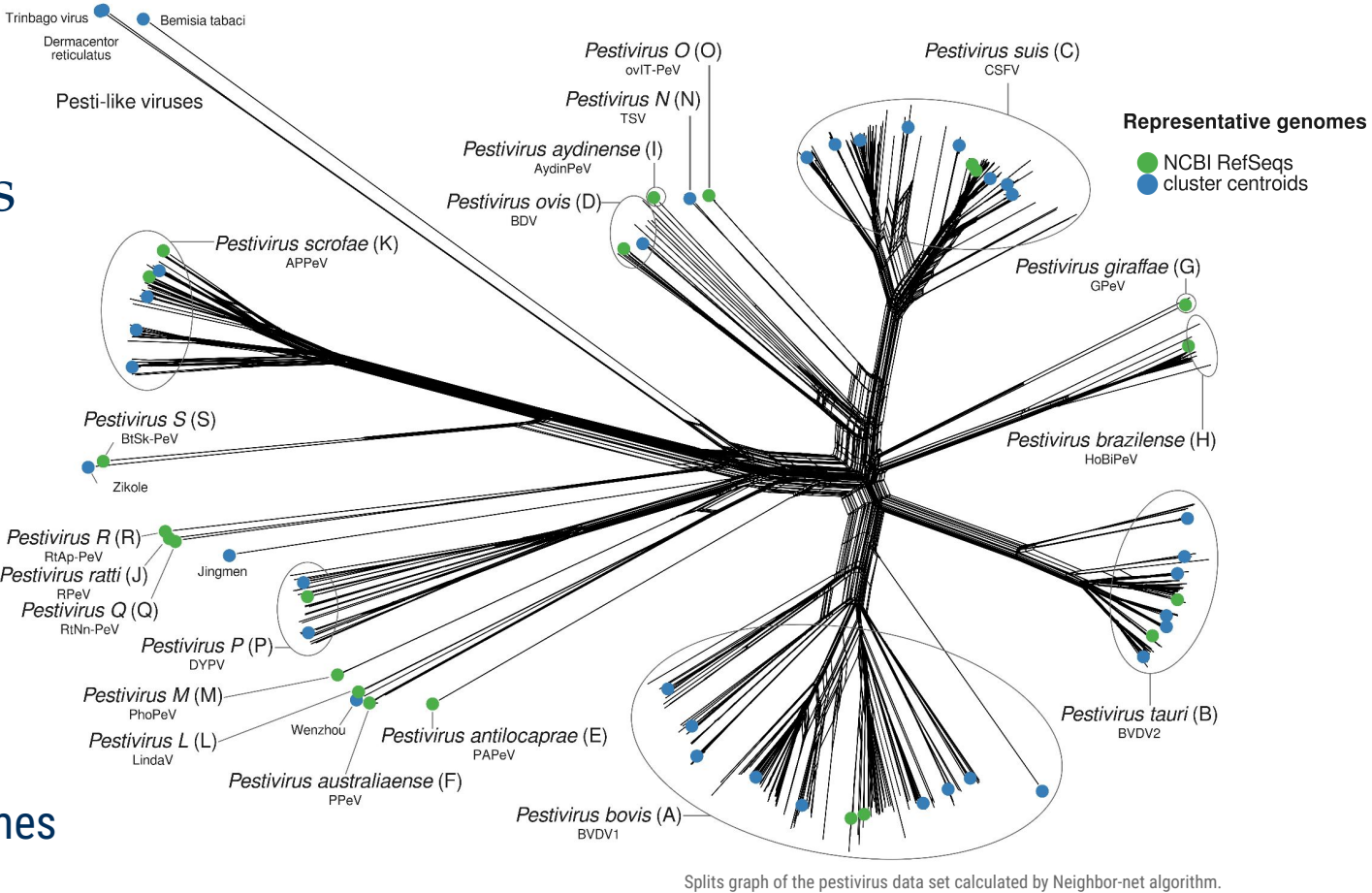


k = 7

Check out our ViralClust pipeline!



Pestivirus representatives



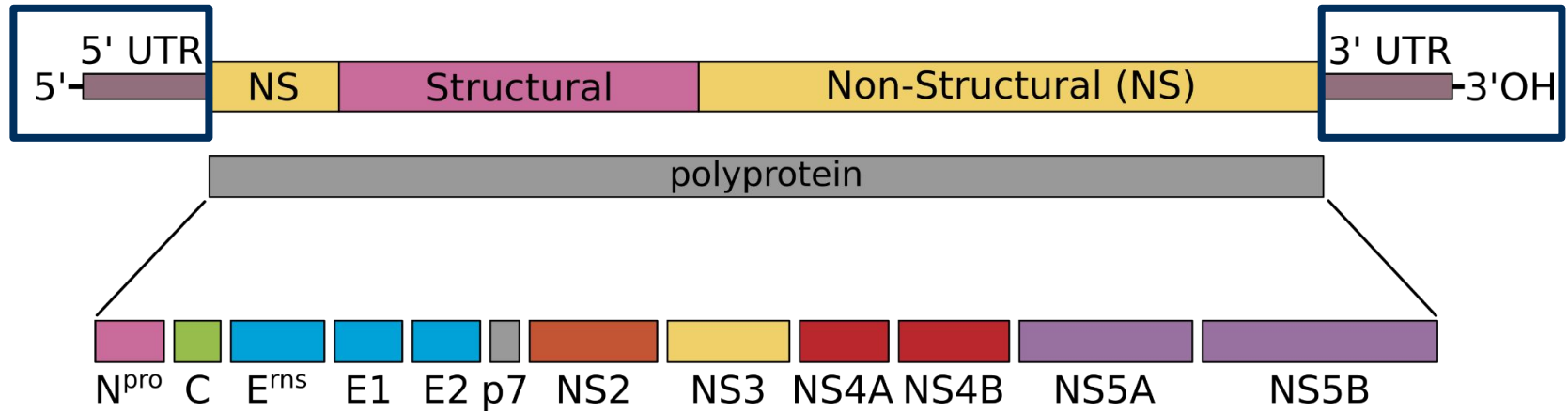
1,684 genomes



59 representative genomes

- Step 2 -
Multiple Sequence Alignment
&
RNA Secondary Structures

Start with the UTRs



Schematic overview of the genome organization of pestiviruses (adapted from ICTV).
UTR = untranslated region.

genome length ~12,000 nt

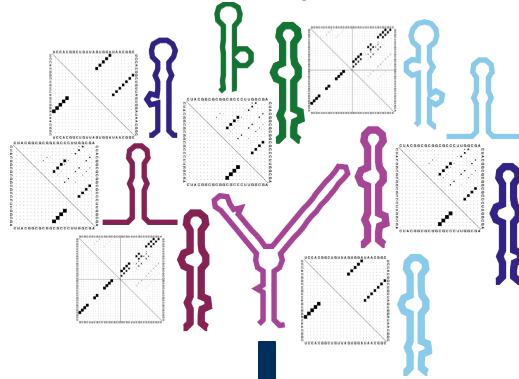
LocARNA - local alignment of RNAs

all possible structures for each sequence

McCaskill's algorithm

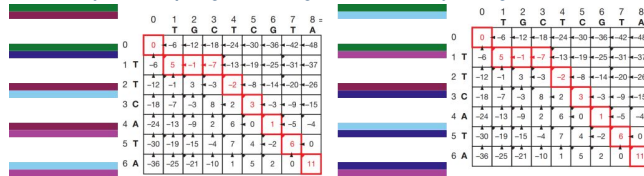
bunch of
RNA sequences

seq1
seq2
seq3
seq4
seq5



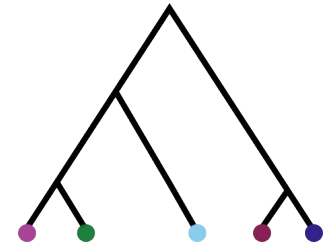
all pairwise alignments

dynamic programming - Sankoff-style algorithm

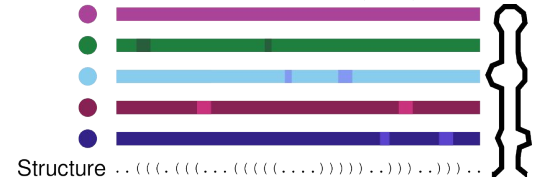


guide tree

WPGMA

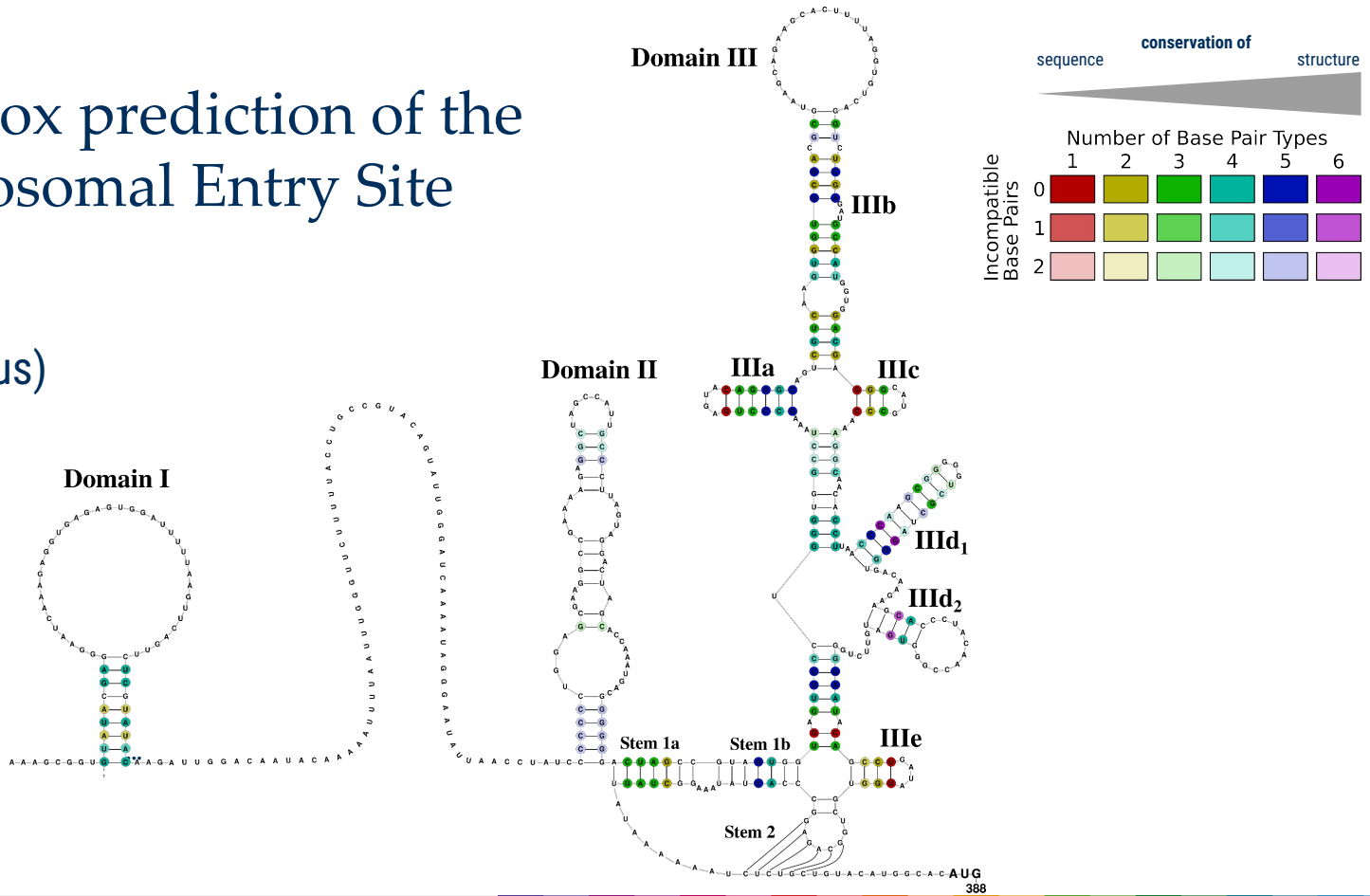


progressive alignment &
consensus secondary structure
minimum free energy (MFE)



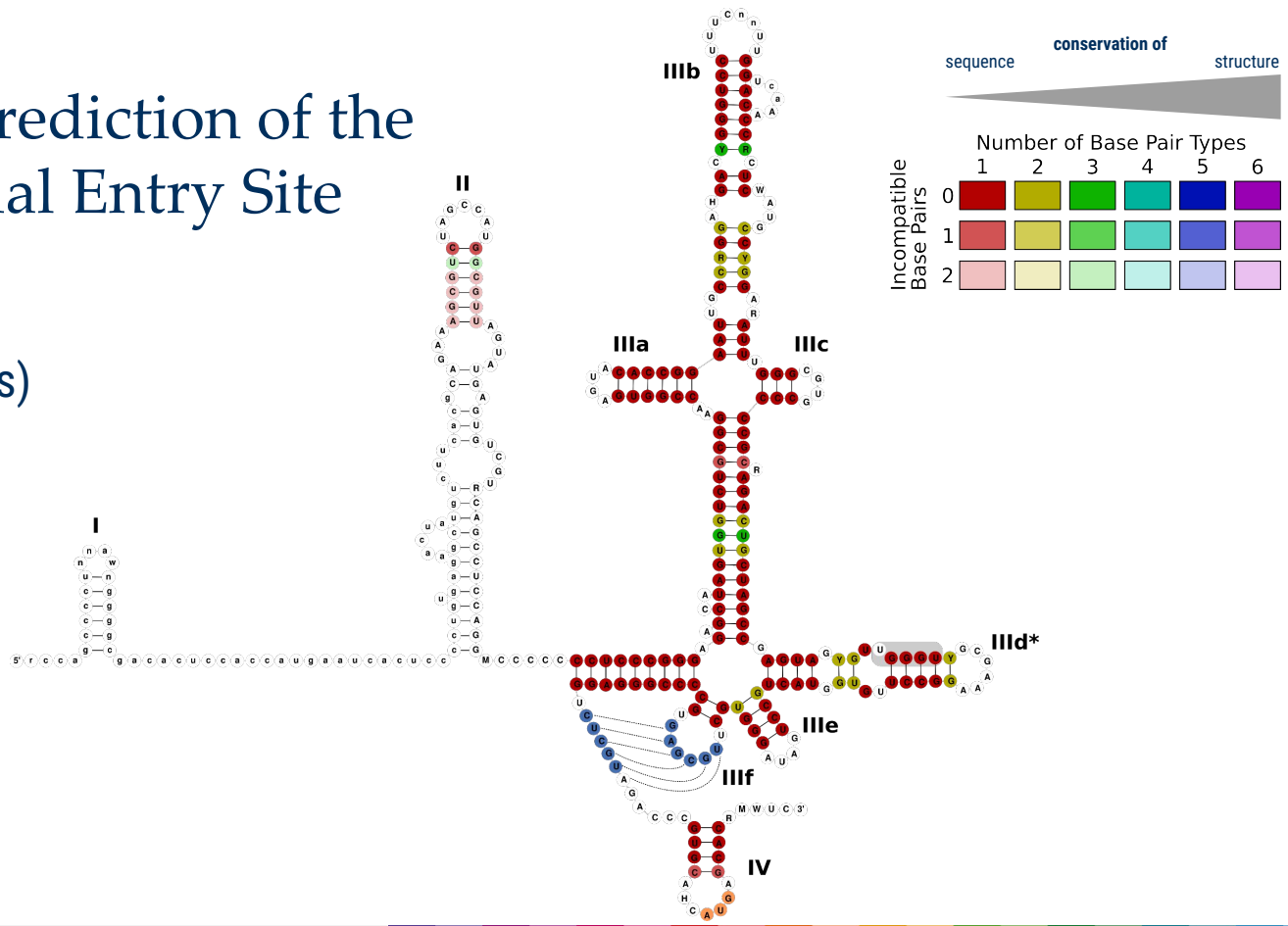
Out-of-the-box prediction of the Internal Ribosomal Entry Site

Pestiviruses (genus)



Out-of-the-box prediction of the Internal Ribosomal Entry Site

Hepatitis C virus (species)



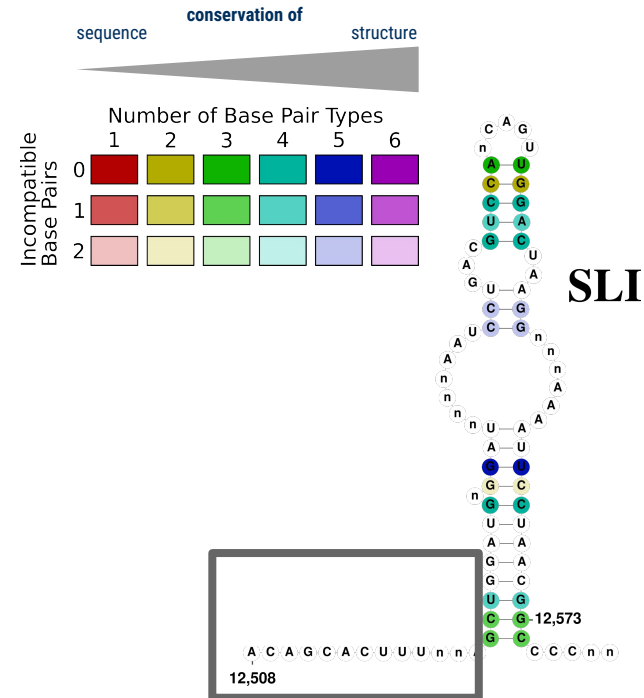
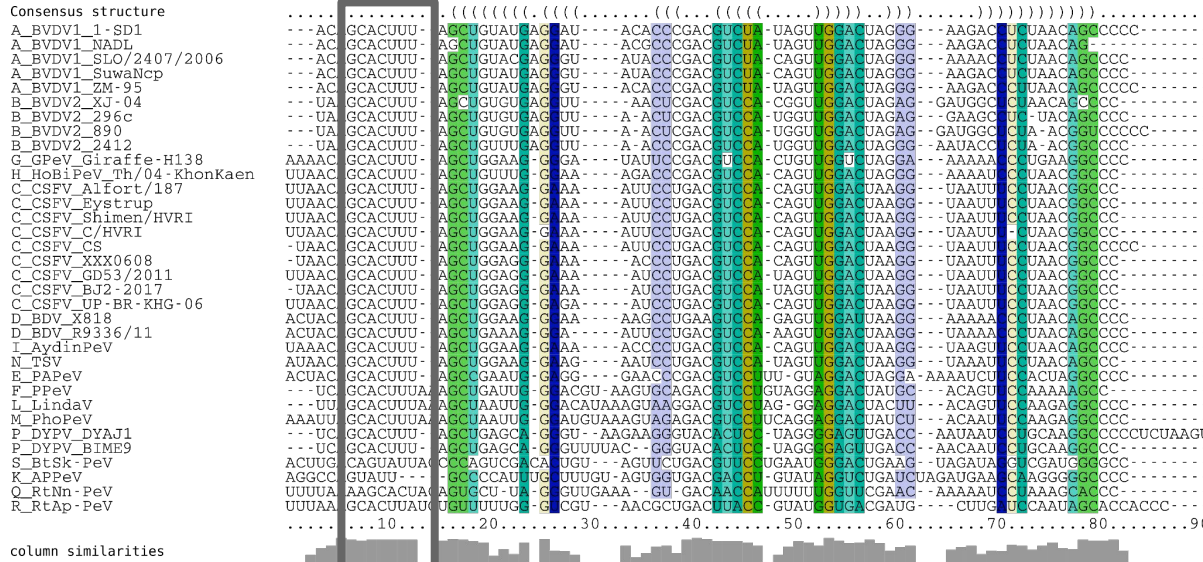
Article | [Open access](#) | Published: 02 July 2024

Comprehensive survey of conserved RNA secondary structures in full-genome alignment of Hepatitis C virus

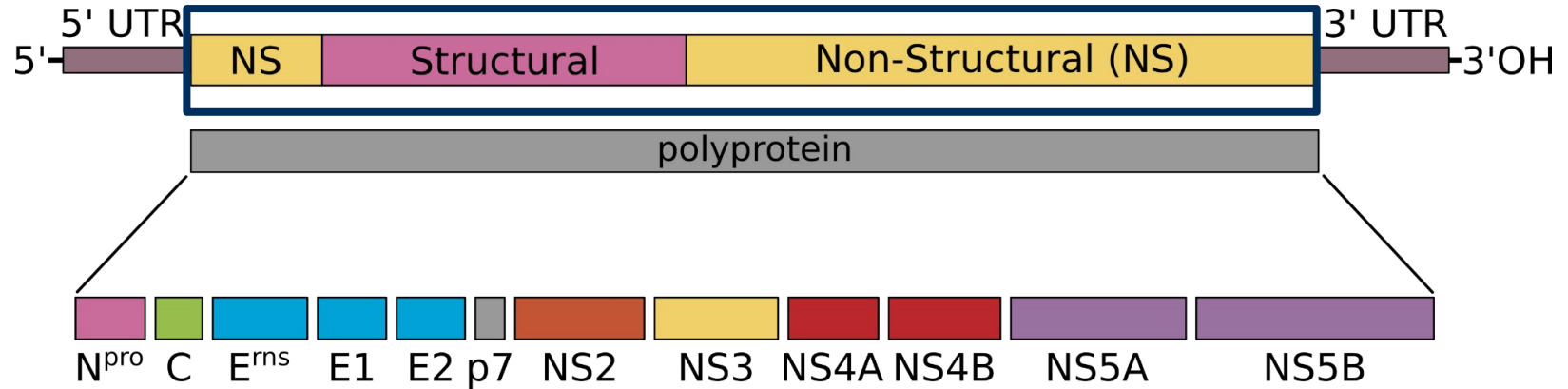
[Sandra Triebel](#), [Kevin Lamkiewicz](#), [Nancy Ontiveros](#), [Blake Sweeney](#), [Peter F. Stadler](#), [Anton I. Petrov](#), [Michael Niepmann](#) & [Manja Marz](#)

Scientific Reports 14, Article number: 15145 (2024) | [Cite this article](#)

3' UTR stem-loop SLI conserved across all pestiviruses & conserved miRNA binding site



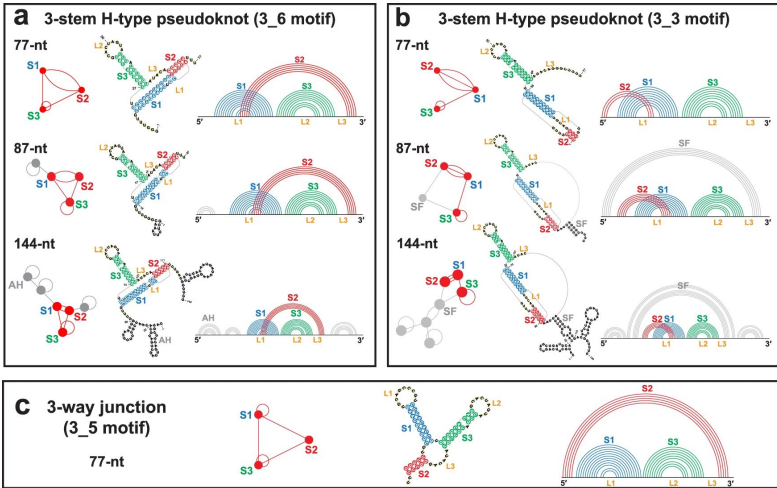
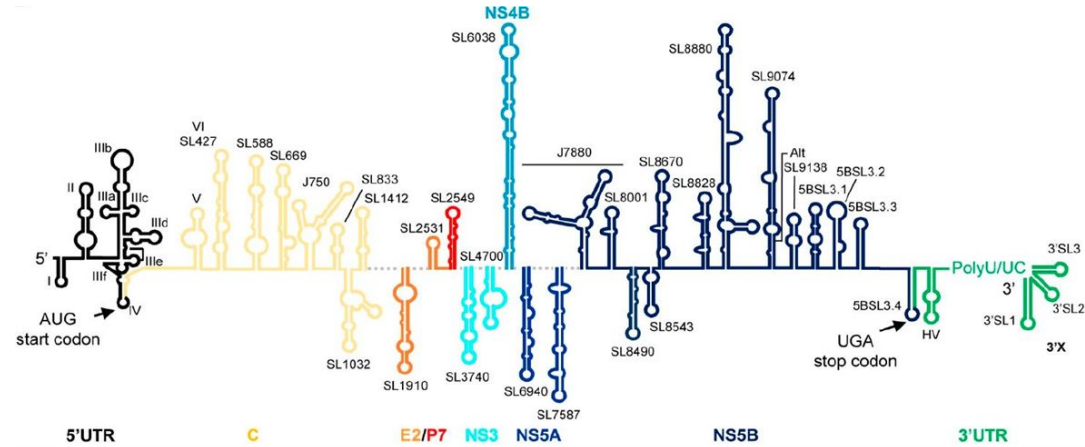
Continue with the (protein-)coding region



Schematic overview of the genome organization of pestiviruses (adapted from ICTV).
UTR = untranslated region.

Why searching for RNA secondary structures in the coding region?

Hepatitis C virus



SARS-CoV-2 frameshifting element

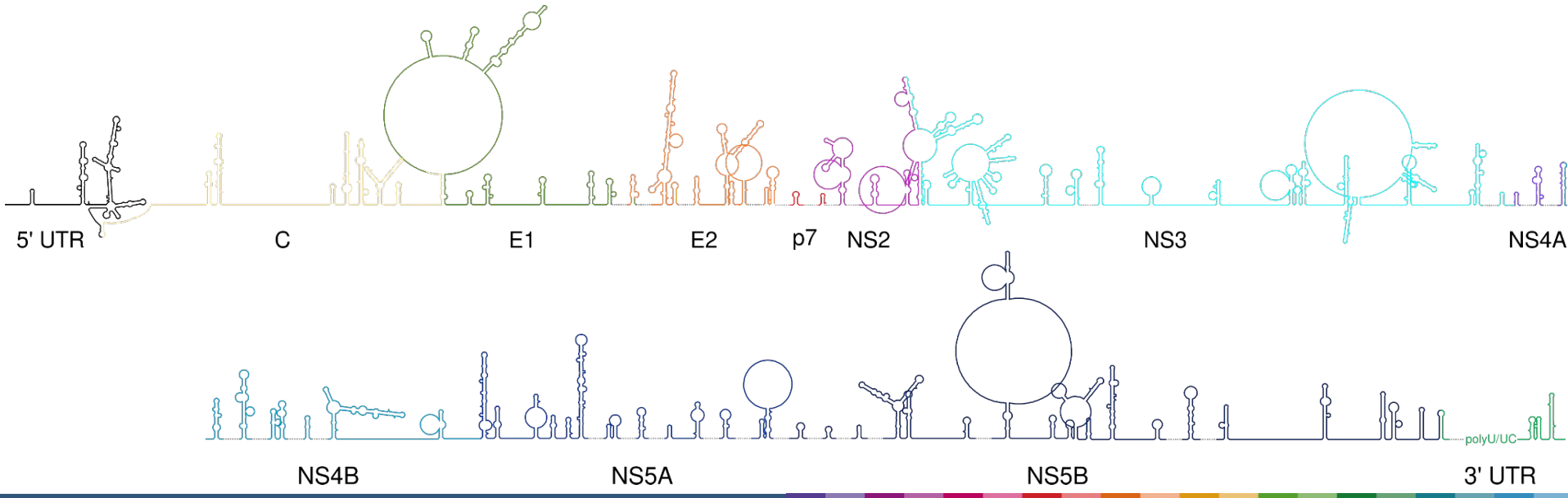
First example out in the world!

Article | [Open access](#) | Published: 02 July 2024

Comprehensive survey of conserved RNA secondary structures in full-genome alignment of Hepatitis C virus

[Sandra Triebel](#), [Kevin Lamkiewicz](#), [Nancy Ontiveros](#), [Blake Sweeney](#), [Peter F. Stadler](#), [Anton I. Petrov](#),
[Michael Niepmann](#) & [Manja Marz](#) 

Scientific Reports **14**, Article number: 15145 (2024) | [Cite this article](#)





Tom Eulenfeld

Multiple sequence alignment of the coding region from amino acids to nucleotides & RNA secondary structures

calculate “anchors” on amino acid level with anchoRNA

```

Genome1  AYEKAVAFS FLLMYSWNPLIRRICLLTSLSEPVTSPSRLTTY YEGDPI SA YKEWPN
Genome2  DSSGS FLLMYSWNPLIRRICLYVLSTVNISSPSRYSTIY FSGDPI AAY REA
Genome3  AYEKAVAFS FLLLYFWNPIIRRICLRVLGETENKYSKVTTYFYCKGDP I GASTGISDPT
Genome4  SC FLLMYAWNPLIRRICLYIISTTTCTDPNPLAPILL QYRGDP I GAFKD
Genome5  RGTAY LLMYPWNPLVRRVALLVLSTTENHYNTLQLVSDEGIS YOGDPI AAY RSRD
Genome6  AYLLMYPWNPLIRRIGLLLLSTTDPMGKEETPCSDEGVK YVGDPI AAY RDDEVK

```

word length ≥ 5



cut into subsequences

```

Genome1  AYEKAVAFS FLLMYSWNPLIRRICLLTSLSEPVTSPSRLTTY YEGDPI SA YKEWPN
Genome2  DSSGS FLLMYSWNPLIRRICLYVLSTVNISSPSRYSTIY FSGDPI AAY REA
Genome3  AYEKAVAFS FLLLYFWNPIIRRICLRVLGETENKYSKVTTYFYCKGDP I GASTGISDPT
Genome4  SC FLLMYAWNPLIRRICLYIISTTTCTDPNPLAPILL QYRGDP I GAFKD
Genome5  RGTAY LLMYPWNPLVRRVALLVLSTTENHYNTLQLVSDEGIS YOGDPI AAY RSRD
Genome6  AYLLMYPWNPLIRRIGLLLLSTTDPMGKEETPCSDEGVK YVGDPI AAY RDDEVK

```



extract nucleotide sequence and align with LocARNA

```

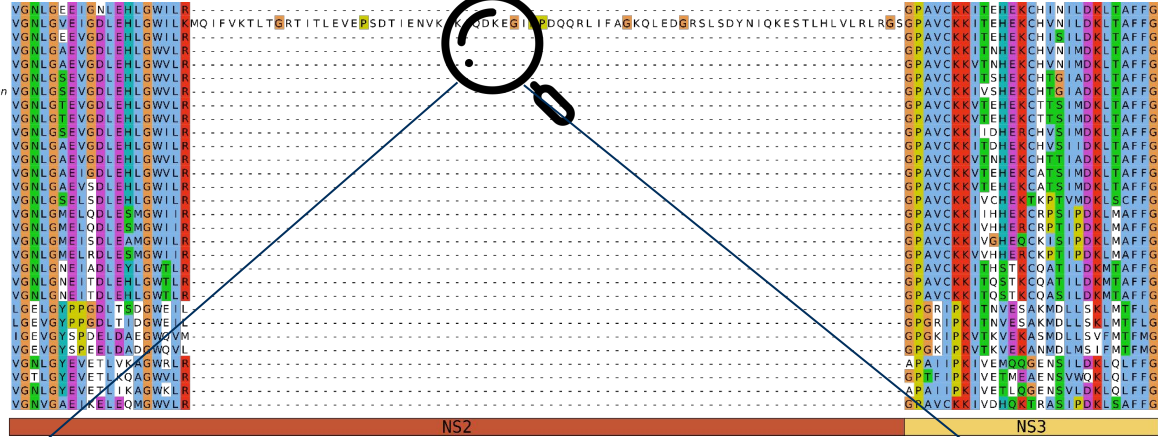
Genome1  UUUUCUCUGAUGUACUC AUGGAACCCUCUGGUAGAAGG AUUUGCCUACUUAUCGUAUCGAGCGAGUUGGGCACCAAACCUAGUAAACGG --- ACCACCUAUUAC ----- UACGAAGGCGAUCCAUAUCUGCCUUA
Genome2  UUCUUAUAUAUGUACAGCUGGAACCCGCUUAUUAGGAGG AUUUGUCUAUAUGUAUUGUCAACAGUAACAAGUUAUCCCAUCCGUAUAGCACC AUUAUACCC ----- UUUUCAGGGGAUCCAUAUCGAGCUUA
Genome3  UUCUUAUAUAUGUACAGCUGGAACCCGCUUAUUAGGAGG AUUUGUCUAUAUGUAUUGUCAACAGUAACAAGUUAUCCCAUCCGUAUAGCACC AUUAUACCC ----- ACCACCUAUUAC ----- UGCAAGGAGAUCCAUAUCGGUGCAUCU
Genome4  UUUUCUCUGAUGUACAGCUGGAACCCGCUUAUUAGGAGG AUUUGUCUAUAUGUAUUGUCAACAGUAACAAGUUAUCCCAUCCGUAUAGCACC AUUAUACCC ----- UUCAAUACAGAGGGGAUCCAUAUCGGGCAUUU
Genome5  UAUCUACUGAUGUACCUUGGAACCCACUGGUCAGGAGA AUUCAGCCUUAUUGCUAUCGACACUGACCCAAUUGGGAAAGAGGAAACCCCAUGCUCGUAUGAGGGGUAAGUAUGUCGGGGACCCAUAGCCGCAUAC
Genome6  UACCUUCUAUUGUACCUUGGAACCCGCUUAUUAGGAGG GUGGCAUUGCUAUGUCUUGUCUACACAGAAACCAUACACACCUUAUUAUUGUUCAGACGAGGUAUUAUGUACAGGAGACCCAUUAGCAGCCUUA
SS_ cons .(((.((((.(.....(((.....))).....)))))).))).....(((.((((.(.....(((.....))).....)))))).))).....(((.((((.(.....(((.....))).....)))))).))).....

```

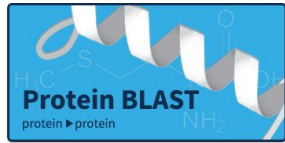


Cool things for virologists → Insertion of Ubiquitin

A_BVDV1_NADL
 A_BVDV1_Ostloss
 A_BVDV1_LO151/09
 B_BVDV2_296c
 B_BVDV2_P199
 G_PeV_Giraffe-H138
 H_HoBiPeV_Th04-KhonKaen
 C_CSFV_Peru-L6
 C_CSFV_Bj2-2017
 D_BDV_X818
 D_BDV_R9336/11
 I_AydinPeV
 N_TSV
 O_ov1-PeV
 E_PaPeV
 F_PPeV
 L_LindaV
 M_PhPeV
 Wenzhou
 P_DYPV_DYA1
 P_DYPV_BIME1
 P_DYPV_BIMES9
 K_APPeV_Bavaria-SS/9
 K_APPeV_AH-GL-2017.04
 S_Btsk-PeV
 Zikole
 Q_RtNn-PeV
 R_RtAp-PeV
 J_rPeV
 Jingmen

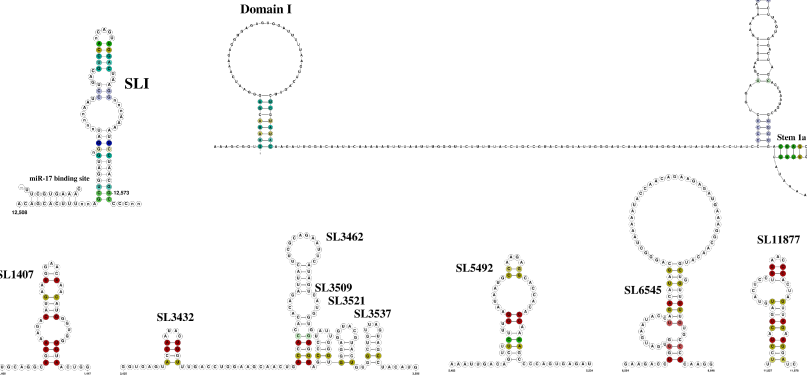


MQIFVKTLTGRTITLEVEPSDTIENVKAKIQDKEGIPPDQQRIFAGKQLEDGRSLSDYNIQESTLHLVLRRLRGS



Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len
hypothetical protein LDENG_00141360 [Lucifuga dentata]	Lucifuga dentata	153	153	98%	7e-46	98.67%	128
ubiquitin-40S ribosomal protein S27a [Microtus ochrogaster]	Microtus ochrogaster	152	152	98%	9e-46	97.33%	110
Chain A, UBIQUITIN [Homo sapiens]	Homo sapiens	151	151	100%	9e-46	97.37%	76
ubiquitin-40S ribosomal protein S27a [Clarias magur]	Clarias magur	152	152	98%	1e-45	97.33%	107
ubiquitin [Galleria mellonella]	Galleria mellonella	152	152	98%	1e-45	97.33%	108

Integration into databases



<i>A_BVDV1_NADL</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>A_BVDV1_Ovines</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>A_BVDV1_LD232.09</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>B_BVDV2_29C</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>B_BVDV2_P99</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>G_GrpeV_Griffin11.30</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>H_BSDrPeV_Ty04.KhonKaen</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>C_CSPV_Peru-8</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>C_CSPV_B2-2037</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>D_BdV_XS18</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>D_BdV_P832021</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>J_AydonPeV</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>L25V</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>O_oxfPeV</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>F_BdV</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>F_PeV</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>L_LindAV</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>H_PraePeV</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>Wenshou</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>D_DrPeV_DNA1</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>D_DrPeV_BdM2</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>L_DrPeV_BdM59</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>K_ApPeV_Bavaria 559</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>K_ApPeV_Ah-QZ-2017.04</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>O_BdV-PeV</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>Zikou</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>O_BdV-PeV</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>B_BdV-PeV</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>J_SpPeV</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG
<i>Jingmen</i>	VGNLQVLEGLNPLRWLIL	WQIFVKVTLTTRTITLLEVI	SDTENVKAKIQDKF	LDQDRLIFAKOLEDD	NSLSDYNIQKESTLMLVRLR	GAVCXKIVGICSHNLDKIAFFG

Acknowledgements



Manja Marz



Tom Eulenfeld



Kevin Lamkiewicz



RNA Bioinformatics & High-Throughput Analysis Jena

- Manja Marz
- Tom Eulenfeld
- Kevin Lamkiewicz

Rfam

- Nancy Ontiveros
- Blake Sweeney

Ribosome Ltd.

- Anton I. Petrov

University Leipzig

- Peter F. Stadler

University Giessen

- Michael Niepmann

University Lübeck

- Norbert Tautz



NFDI4MICROBIOTA



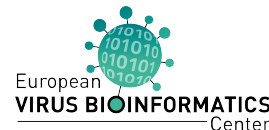
Ver: 0.1



UNIVERSITÄT
LEIPZIG



UNIVERSITÄT ZU LÜBECK



Deutsche
Forschungen
German Research Fc