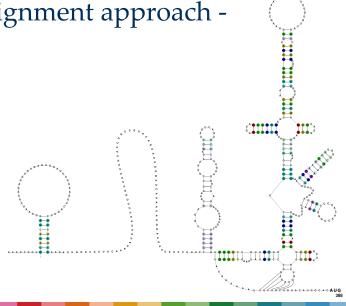


RNA secondary structures prediction in viruses

- a full-genome multiple sequence alignment approach -

Sandra Triebel

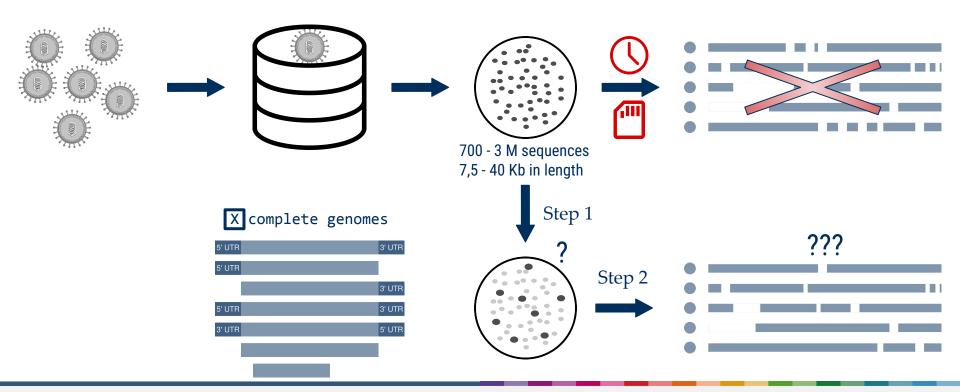
RNA Bioinformatics & High-Throughput Analysis Friedrich Schiller University Jena, Germany







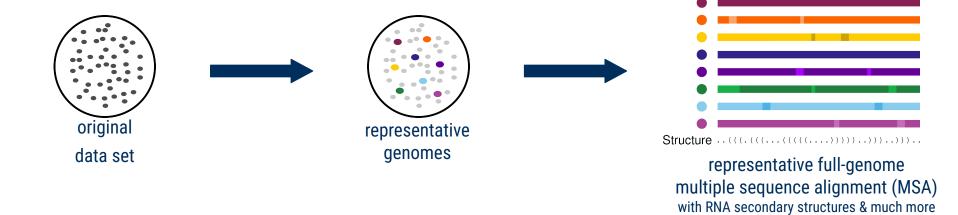
Working with viral genomes...







The rough workflow







- Step 1 -Representative Genomes

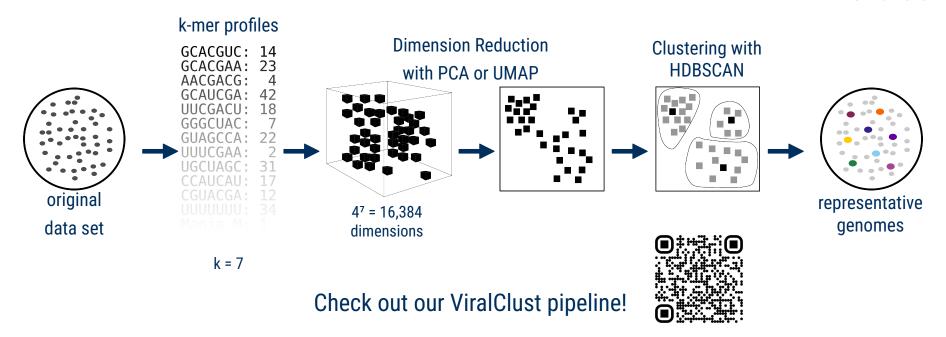




Representative genomes



Kevin Lamkiewicz





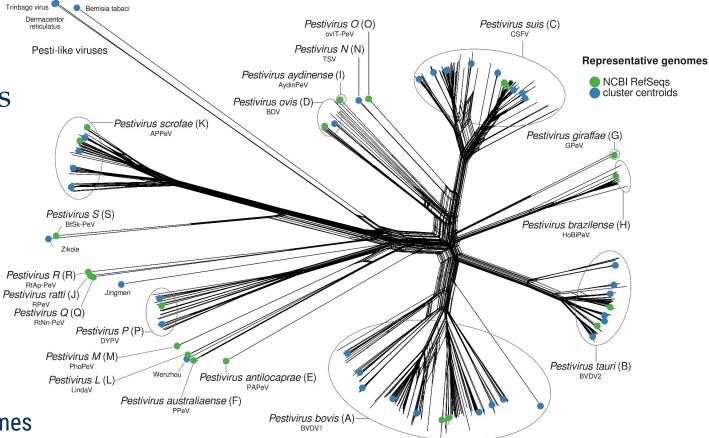


Pestivirus representatives

NCBI Virus

**-BV-BRC

1,684 genomes





Splits graph of the pestivirus data set calculated by Neighbor-net algorithm.



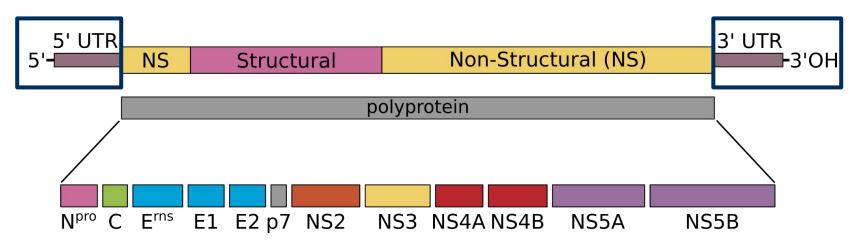


- Step 2 Multiple Sequence Alignment &
RNA Secondary Structures





Start with the UTRs



Schematic overview of the genome organization of pestiviruses (adapted from ICTV).

UTR = untranslated region.

genome length ~12,000 nt



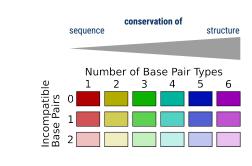


LocARNA - local alignment of RNAs all possible structures for each sequence quide tree McCaskill's algorithm **WPGMA** bunch of RNA sequences seq1 seq2 sea3 seq4 seq5 progressive alignment & consensus secondary structure all pairwise alignments minimum free energy (MFE) dynamic programming - Sankoff-style algorithm

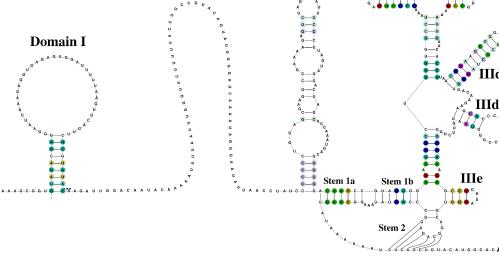




Out-of-the-box prediction of the Internal Ribosomal Entry Site



Pestiviruses (genus)







Domain II

Domain III

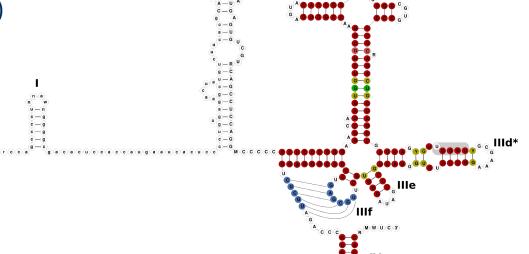
IIIa

IIIb

IIIc

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

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





conservation of

Number of Base Pair Types

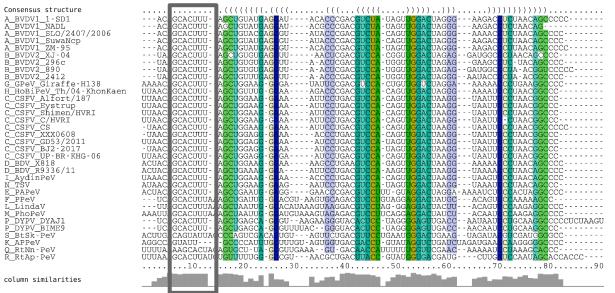
Incompatible Base Pairs

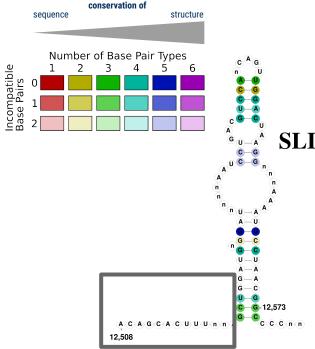
IIIc

structure

3' UTR stem-loop SLI conserved across all pestiviruses

& conserved miRNA binding site

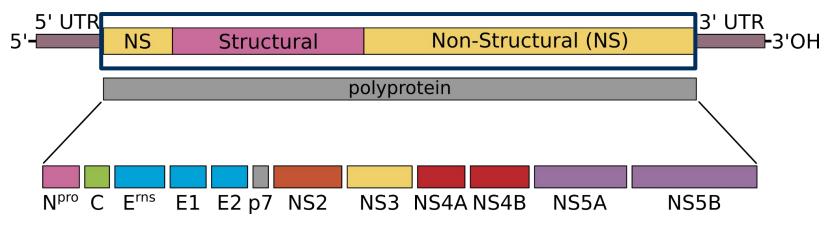








Continue with the (protein-)coding region



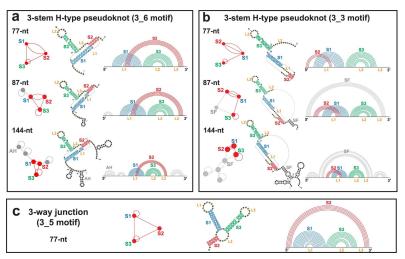
Schematic overview of the genome organization of pestiviruses (adapted from ICTV).

UTR = untranslated region.

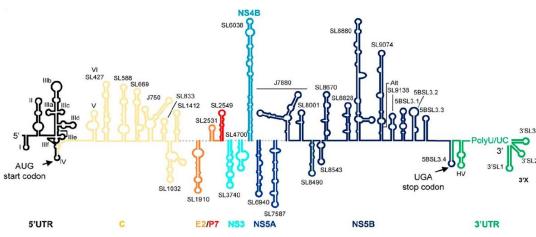




Why searching for RNA secondary structures in the coding region?



Hepatitis C virus



SARS-CoV-2 frameshifting element





First example out in the world!

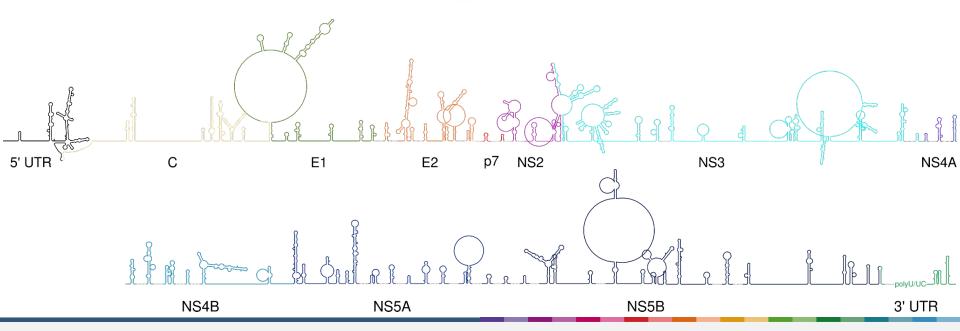
Article Open access | Published: 02 July 2024

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

 $\underline{Sandra\ Triebel}, \underline{Kevin\ Lamkiewicz}, \underline{Nancy\ Ontiveros}, \underline{Blake\ Sweeney}, \underline{Peter\ F.\ Stadler}, \underline{Anton\ I.\ Petrov}, \underline{Michael\ Niepmann\ \&\ Manja\ Marz\ } \\ \underline{ }$



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







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



Tom Eulenfeld

calculate "anchors" on amino acid level with anchoRNA

Genome1 Genome2 Genome3 Genome4 Genome5 Genome6 AYEKAVAFSFLLMYSWNPLIRRICLLTLSSEPVTSPSRLTTYYYEGDPISAYKEWPN
DSSGSFLLMYSWNPLIRRICLYVLSTVNISSPSRYSTIYTFSGDPIAAYREA
AYEKAVAFSFLLLYFWNPIIRRICLRVLGETENKYSKVTTFYCKGDPIGASTGISDPT
SCFLLMYAWNPLIRRICLYIISTTTCTDPNPLAPILLQYFGDPIGAFKORGTAYLLMYPWNPLVRRVALLVLSTTENHYNTLQLVSDEGISYQGDPIAAYRSRD
AYLLMYPWNPLIRRIGLLLLSTTDPMGKEETPCSDEGVKYVGDPIAAYRDDEVK

word length ≥ 5



cut into subsequences

AYEKAVAFSFLLMYSWNPLIRRICLLTLSSEPVTSPSRLTTYYYEGDPISAYKEWPN
DSSGSFLLMYSWNPLIRRICLYVLSTVNISSPSRYSTIYTFSGDPIAAYREA
AYEKAVAFSFLLLYFWNPIIRRICLRVLGETENKYSKVTTFYCKGDPIGASTGISDPT
SCFLLMYAWNPLIRRICLYIISTTTCTDPNPLAPILLQYKGDPIGAFKD
RGTAYLLMYPWNPLVRRVALLVLSTTENHYNTLQLVSDEGISYQGDPIAAYRSRD
AYLLMYPWNPLIRRIGLLLLSTTDPMGKEETPCSDEGVKYVGDPIAAYRDDEVK



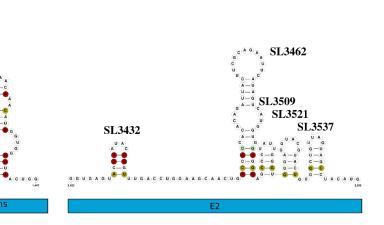
extract nucleotide sequence and align with LocARNA

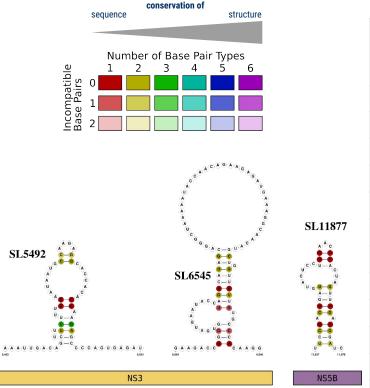
Genome1 Genome2 Genome3 Genome4 Genome5 Genome6 SS_cons 





New structure candidates in the coding region







Gene	Number of potential structures
Npro	9
С	4
Erns	14
E1	8
E2	13
p7	4
NS2	13
NS3	30
NS4A	3
NS4B	16
NS5A	10
NS5B	28
total	152



SL1407



New structure candidates in the coding region

SL6545

A BVDV1 NADL A BVDV1 Osloss A BVDV1 L0/151/09 B BVDV2 296c B BVDV2 PI99 G GPeV Giraffe-H138 H HoBiPeV Th/04-KhonKaen C CSFV Peru-L-8 C CSFV BJ2-2017 D BDV X818 D BDV R9336/11 I_AydinPeV N TSV 0 ovIT-PeV E PAPeV F PPeV L LindaV M PhoPeV Wenzhou P DYPV DYAJ1 P_DYPV_BIME1 P DYPV BIME9 O RtNn-PeV R RtAp-PeV J RPeV Jingmen

#=GC SS cons

GAGGAUCUUGGUAGUCAGUUCCUUGAUAUAGCAGGGUUAAAAAUACCAGUGGAUGAGAGGCAAUAUGUUGGUUUUUGUACCAACGAGA GAAGACCUUGGAAGCCAGUUCCUUGACAUAGCGGGGCUAAAAAUCCCGGUUGAGGAGAUGAAGGGUAACAUGCUGGUCUUCGUACCCACAAGA GAAGACCUUGGUAGCCAAUUCCUCGAUAUAGCAGGAUUGAAGAUCCCCGUAGAGGAGAUGAAGGGUAACAUGCUGGUUUUUGUGCCGACAAGG GAAGAUCUAGGAAGUGAAUACUUGGAUAUUGCAGGGCUGAAGAUACCAACUGAAGAGAUAAAAGGCAACAUGCUCGUUUUUUGUCCCAACCAGA GAGGAUCUAGGUAGUGAAUACUUAGACAUUGCAGGGUUGAAGAUACCAACUGAAGAGAUGAAAGGCAACAUGCUUGUGUUUGUGCCAACCAGA GAAGAACUAGGUUCUGAAUUCAUAGAUAUAGCUGGACUCAAGAUCCCAACAGAGGAGAGAGGCAACAUGCUAGUCUUCGUCCCCACUAGA GAAGACCUGGGAUCAAACUUCCUGGAUAUAGCUGGACUUAAGAUACCAACAGAAGAGAUGAAGGGAAAUAUGUUGGUGUUCGUACCCACUAGG GAAGACUUGGGCUCAGAGUACUUGGACAUUGCUGGGCUAAAGAUACCAGCAGAGGAGAUGAAGAGUAACAUGCUGGUUUUCGUGCCUACUAGG GAAGACUUAGGUUCGGAGUACUUAGACAUAGCCGGGCUAAAGAUACCAGUAGAAGAUGAAGAACAAUAUGCUAGUUUUUUGUACCAACUAGG GAAGAUCUUGGCUCUGAGUUCCUGGAGAUUGCAGGCCUGAAAAUACCCACUGAGGAAAUGAAAGGUAACAUGCUGGUAUUCGUGCCCACCAGG GAGGACUUGGGUUCCGACUUUCUAGAAAUAGCAGGACUGAAAAUACCAACUGAAGAGAUGAAGGGCAACAUGCUGGUUUUUGUACCCACCAGA GAAGACCUGGGCUCAGAGUUCUUAGACAUAGCGGGGUUAAAAAUACCAACAGAUGAAAUGAAAGGCAACAUGCUGGUGUUUGUGCCCACCAGG GAGGACUUGGGAUCUGAGUACUUAGACAUAGCUGGGUUGAAGAUACCAGUGGAAGAGAUGAAAGGUAAUAUGCUGGUGUUUGUACCCACGAGG GAAGACUUGGGCUCAGAAUUUCUAGACAUAGCCGGAUUGAAAAUACCGGUGGAUGAGAUGAAAGGCAAUAUGCUAGUGUUCGUGCCCACGAGG GAGGACCUUGGAGAAAACUUCUUGGAUAUAGCUGGUCUAAAAGUGCCUAGGGAUGAAGAAGAACAACGUGCUCGUAUUCGUGCCGACCAGG GAGGACCUUGCUGAUGAUUACAUCGAAAUAGCAGGGUUGAAGGUGCCAAAGAAGAGUUAGAGGGUAACGUACUGACUUUUGUGCCUACAAGG GAAGACCUAGGCGAGGAUUAUUUGGACAUAGCCGGACUAAAAAUACCAAAAUCGGAACUACAAGGGAAUGUCUUAACGUUUGUUCCGACAAAA GAAGACCUUACUGAGCAACACAUAGAAAUAGCUGGAUUAAAGAUUCUAAAAAAGGAGCUAGAUAAUAAUGUAUUAGUUUUUGUUUCUACCAAA GAAGAUCUUGGAGACGAAUUCGUGGACAUAGCAGGCCUAAAGAUUCCUAGAGCUGAACUAAAGGCGAACAUACUAGCUUUCGUCCCCACAAAG GAAGACUUGGGAGAAGAAUAUAUAAACAUAGCAGACCUAAAAGUCCCAACAAUGGAGCUAAAAAACAACUGCCUGGUCUUUGUACCCACAAGA GAAAAUCUGGGUGAUGAGUACAUAAAUAUAGCUGACCUAAAAGUCCCUAUUUCAGAACUAAAGAAUAACUGCCUAAUAUUUGUACCUACUAGG GAAAAUCUGGGGGAUUCCUAUCAUAAUGUUGCAGGCCUAAAGGUAGGCAAGAACAUAUUAAAAGAUAACACCCUUGUAUUUGUGGCUACUAGG GAAAAUCUAGGUGAAAAUUACAUUGAGUUGGCUGGGUUGAAAGUAAAAGAACCAUACUCAAAGAAAAUACACUUAUGUUUGUCCCUACAAGA GAGAAUCUUGGAGAAAAUUACCUCAAUGUAGCUGGCUUAAAAGUUGCAAGAACUAUUCUAAAAGAGAAUACAUUAGUCUUUGUCCCAACUAGA GAGAACCUAGGGAAUGACUUCUUGAAUAUUGCUGGAUUGAAGAUAGCCAAAGUAGAAUUGCAACAGAACACGCUCAUAUUUGUACCAACAAGG

SL3432 - SL3537

A BVDV1 NADL A BVDV1 Osloss A BVDV1 L0/151/09 B BVDV2 296c B BVDV2 PI99 G GPeV Giraffe-H138 H HoBiPeV Th/04-KhonKaen C CSFV Peru-L-8 C CSFV BJ2-2017 **D BDV X818** D BDV R9336/11 I AydinPeV N TSV O ovIT-PeV E PAPeV F PPeV L LindaV M PhoPeV Wenzhou P DYPV DYAJ1 P DYPV BIME1 P DYPV BIME9

GGAGAGUACCAAUACUGGUUCGACCUAGAAGUUACAGACCACCACAAAGACUACUUCGCCGAGUUUAUCCUCGUGAUAGUGGUGGCCCCUCCUGGGUGGAAGGUACGUGCUCUGGCUGCUAGUCACUUAUAUAUG GGGAAGUGGCAAUAUUGGUUUGACCUGGAUUCUAUAGACCACCACAAAGACUACUUUUCAGAGUUCAUAGUCAUAGCAGUGGUAGCCUUGCUAGGUGGUAGUAUGUACUGUGGCUCUUAGUAACAUAUAUG GGGGGGUGGCAAUAUUGGUUUGACCUGGAUUCUGUGGACCAUCACAAAGAUUACUUUUCAGAGUUCAUAGUCAUAGCAGUGGUAGCUUUGUUGGGUGGUAGUAUUGUAGUAGCAUCG GGAAAAUACCAAUAC<mark>UGGU</mark>UUGACCUCAAAGCUACUGACAGGAAGGACUACUUUGCCGAGUUCCUAGUGAUAGCAGUGGUCG<mark>CU</mark>UAUU<mark>AGGGGGG</mark>AGGUACG<mark>UACU</mark>GUGGUUGUUGGUGACCUAUUUC GGAGAGUAUCAAUACUGGUUUGACCUUGAUAUAACAGACCAUCGUACUGAUUACUUUGCUGAAUUCAUAGUGUUGGCAGUAGUAGCCCUGUUGGGCGGGUAGAUAUG<mark>UACU</mark>AUAGGCUGAUGGUAGUAUAUA UCAGGGUAUCAAUAUUGGUUUGAUCUCACAGCAAAGGAUCAUGUGGCAGACUGGAUCACAAAAUACUUUCCAAUAAUAAUAGUGGCCUUGUUAGGGGGCAGAGGCACCUUGUGGGGUGUUGAUAGGCUUAUGAG GGUGGCUAUCAGUAU<mark>UGGU</mark>UUGACCUGACAGCAAAAGAUCACGUGAUGGAUUGGGUAACACGAUAACUUCCCCAUUAUAGUUGUAGCAUUACUGGGG<mark>GGGUA</mark>GAGCAG<mark>UGCU</mark>GUGGAUCCUAAUUGCGUACGAG GAUGAAUACCAAUAUUGGUUUGACUUAAAAGCAGGGGACCAUGCAGUUGAUUUAGUGUCUAAAUACUUACCUAUAUUGGUAGCUUUAUUGGGGGGAAAAGCUG<mark>UGUU</mark>GUGGAUUUUAGUAGCAUAUACU GAUGGCUACCAAUACUGGUUUGACCUGGAAGCUGGUGACCAUAGGAGGAGUAUUUUAUCAGAAUGUAUAGUAAUUGUCGUGGUAGCUUUACUAGGUGGCAAAAUUUCU<mark>UU</mark>AUGGAUUCUAAUAGCUUUCCAU GAUGGGUACCAAUAUUGGUUUGACCUAGAAGCUGGUGACCAUAGAAGAAGCAUCUUAGCAGAAUAUAUAGGAUGUAGCACUAUUGUGAUUGUAGCACUAUAGGUGGUGGAUCCUAGUAGCUUACCAU GAUGGAUACCAGUAC<mark>UGGU</mark>UUGAUUUAUCAGCCAUGGACCAUCGAAGGAGUGUGGUUGCAGAAUUUAUAGUAAUCAUUGUUGUUGCACUCUUAGGA<mark>GG</mark>GAGGGUGACA<mark>CU</mark>CUGGCUGUUAGGUAGCUUACCAU GAUGAAUACCAGUACUGGUUCGACCUGAAAAGCAGUGACCACUAUAUAGGGGGUGUCGUAAAAUACCUCCCCUUGGUGAUGGUCGUCCUUCUUGGUGGCAAACUUGCAGCAUGGUUGUUGACUGCCUAUUAC AACGGUUGGCAGUACUGGUUUGACCUGGAGGCUUCUGACCAUAUGGGGGCACAAAUCAGUAAAUAUCUCCCAAUAGUGGUCGUGG<mark>CUU</mark>UAUUGGGGGGGUAGGUAUGUACUGUGGCCUUAGUCACAUAUUUA



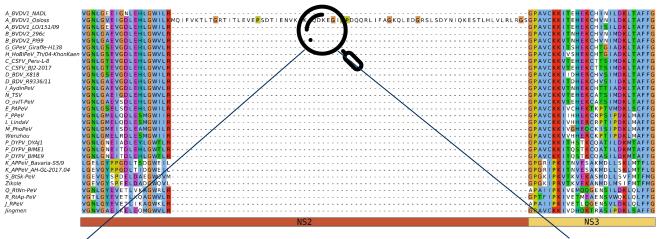
Q_RtNn-PeV R RtAp-PeV

J RPeV

Jingmen #=GC SS cons



Cool things for virologists → Insertion of Ubiquitin



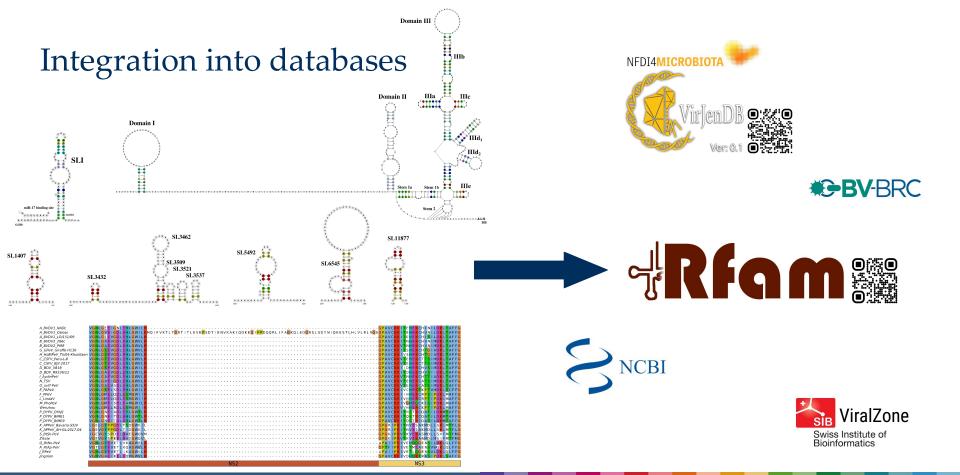




Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len
hypothetical protein LDENG_00141360 [Lucifuga dentata]	Lucifuga dentata	153	153	98%	7e-46	98.67%	128
ubiquitin-40S ribosomal protein S27a [Microtus ochrogaster]	Microtus ochrogaster	152	152	98%	9e-46	97.33%	110
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











Acknowledgements















RNA Bioinformatics & High-Throughput Analysis Jena

- Manja Marz
- Tom Eulenfeld
- Kevin Lamkiewicz

Rfam

- Nancy Ontiveros
- Blake Sweeney

Riboscope Ltd.

Anton I. Petrov

University Leipzig

Peter F. Stadler

University Giessen

Michael Niepmann

University Lübeck

Norbert Tautz



















