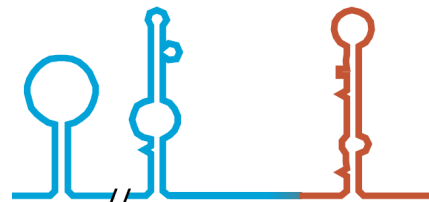
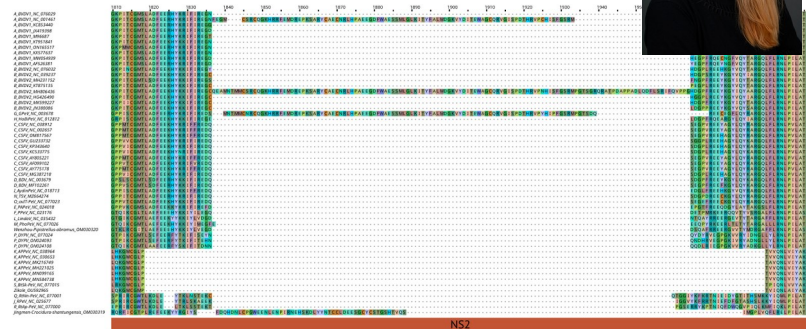
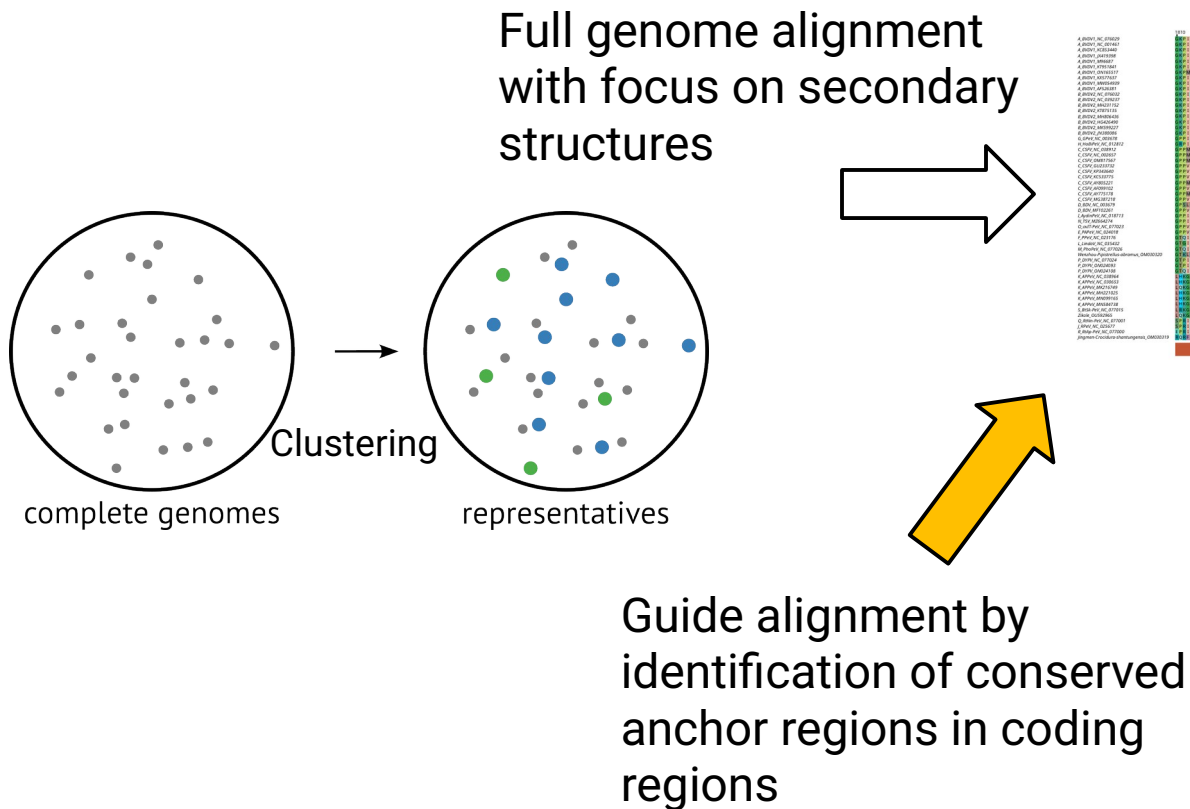


# AnchoRNA

Multiple sequence alignments for secondary structure prediction using anchors in coding regions

Tom Eulenfeld  
RNA Bioinformatics & High-Throughput Analysis  
Friedrich Schiller University Jena / Germany

# Motivation: Secondary structures from virus full genome alignment



Triebel et al. 2024 &  
Triebel et al. 2026

# Motivation: Secondary structures from virus full genome alignment

difficulty

conservation  
difficulty

Hepaciviruses  
(HCV)  
high  
easy

Pestiviruses  
(BVDV)  
medium  
medium

Filoviridae  
(Ebola virus)  
low  
hard

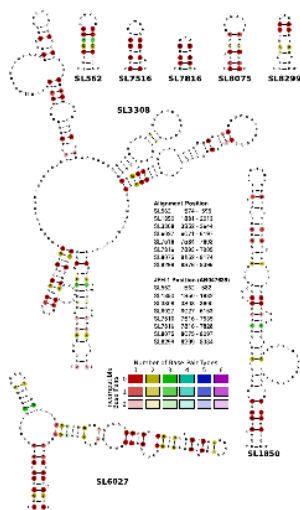
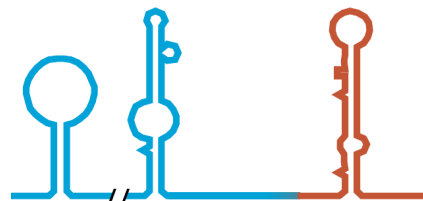


Figure 3. Eight novel selected conserved RNA secondary structure candidates



Triebel et al. 2026

What virologists do:

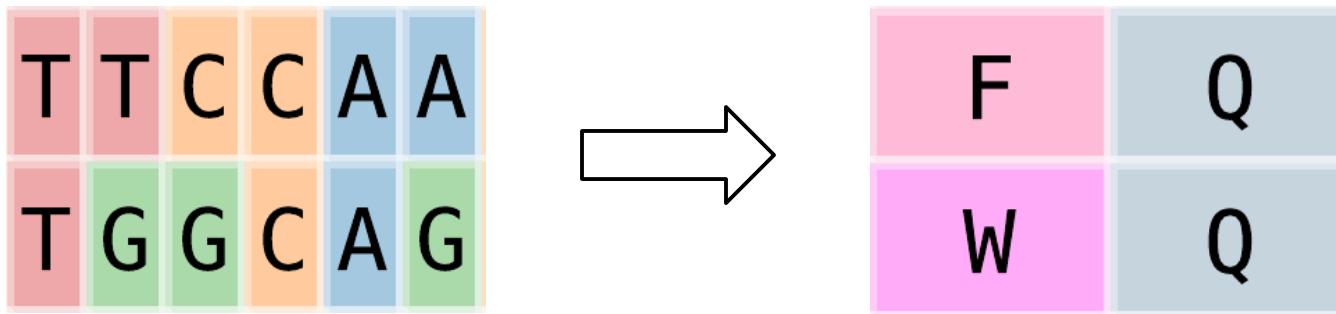
translate

→ align

→ “backtranslate”

Triebel et al. 2024

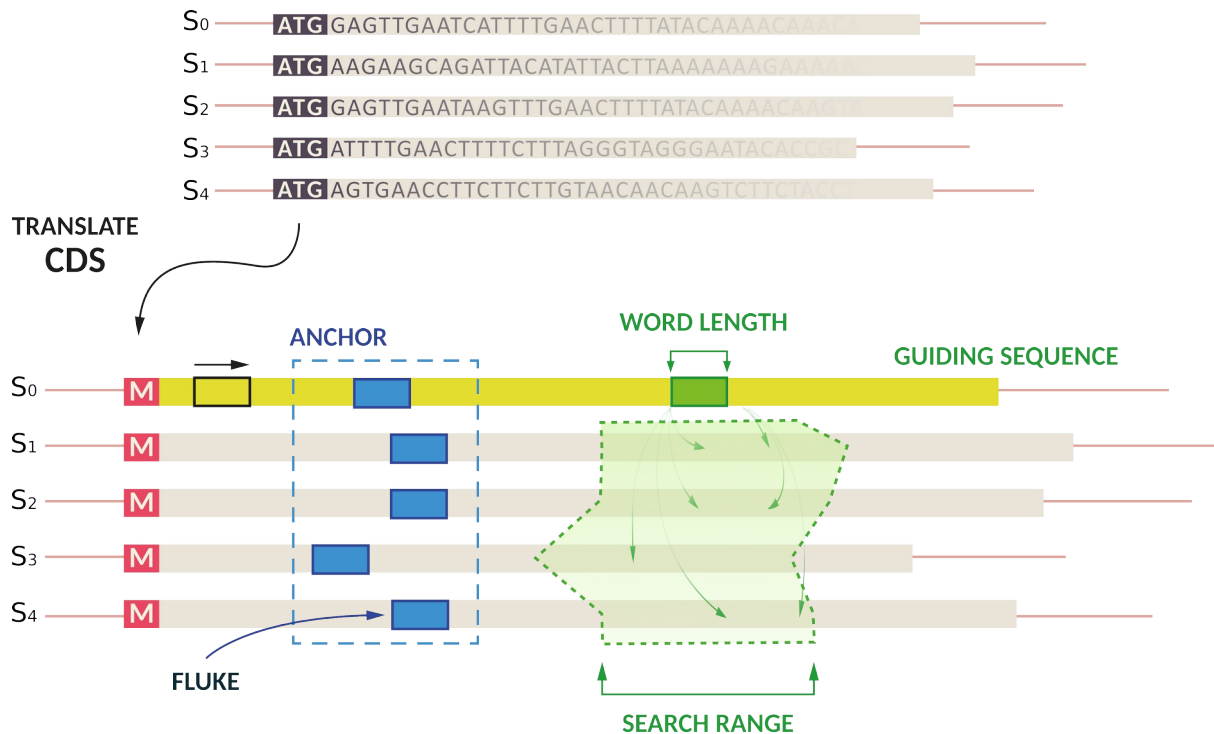
## Use translated CDS to find anchors



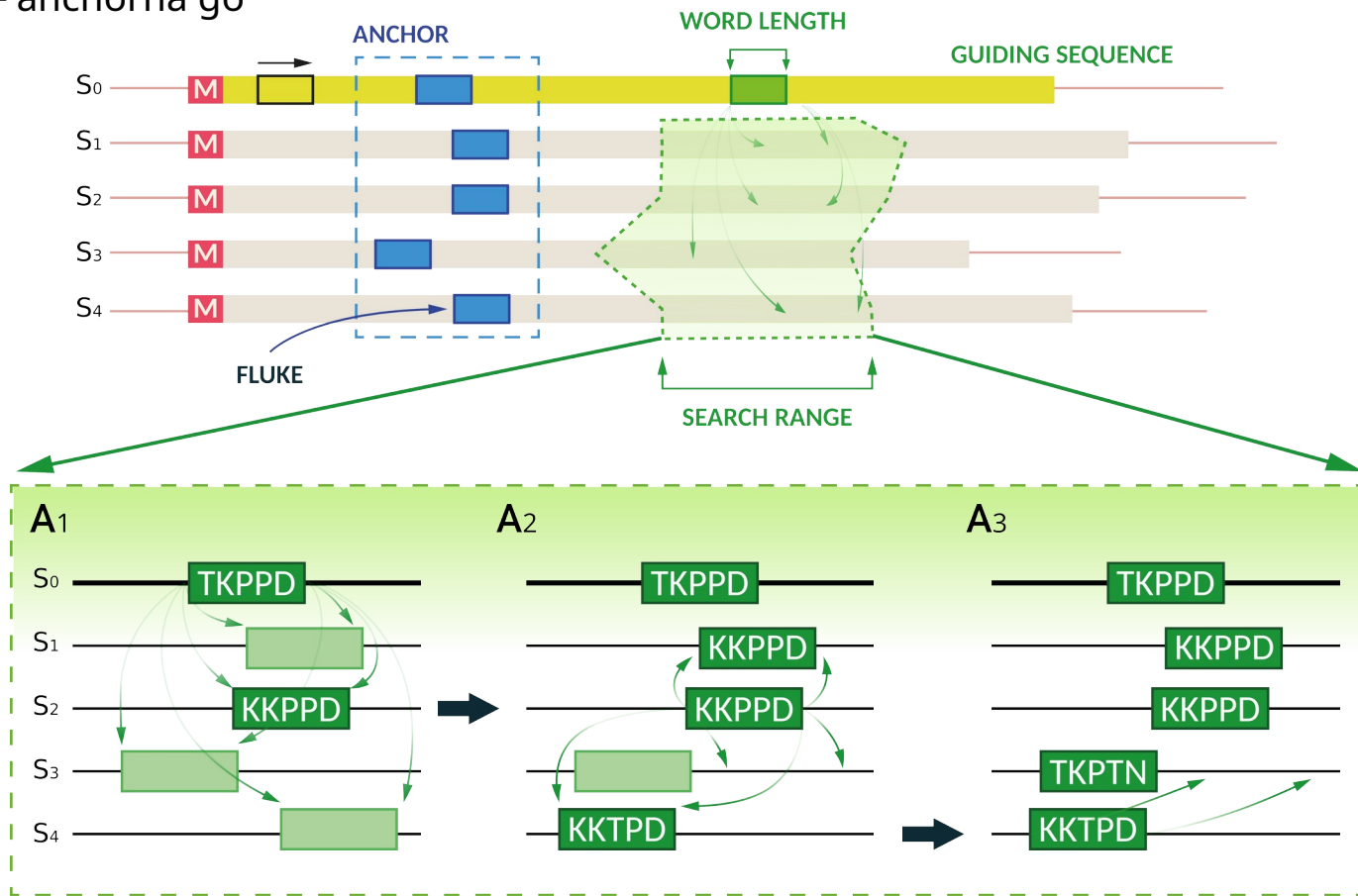
NUC score +5 equal nt, -4 unequal nt  
→ +3 (score for identical: +30)  
Expect equal or better score each 35 nts

BLOSSUM62 score  
→ +6  
(score for identical: +8, common aas)

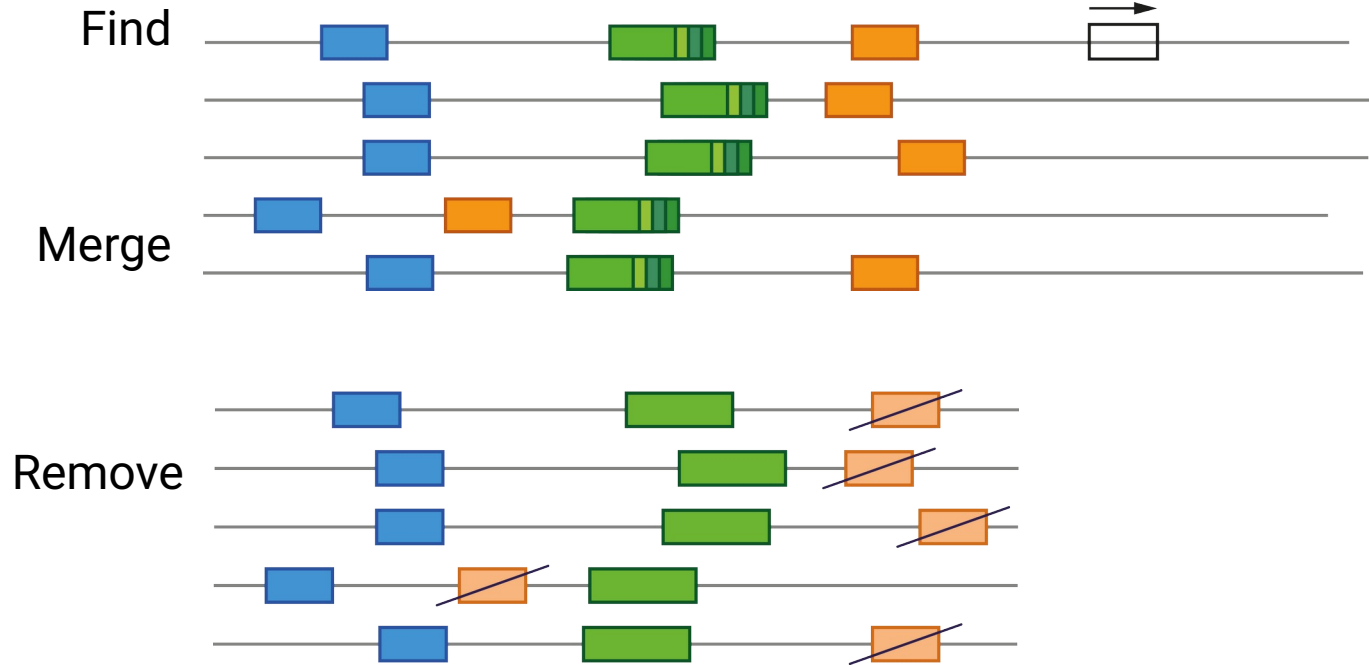
# Algorithm - anchorna go



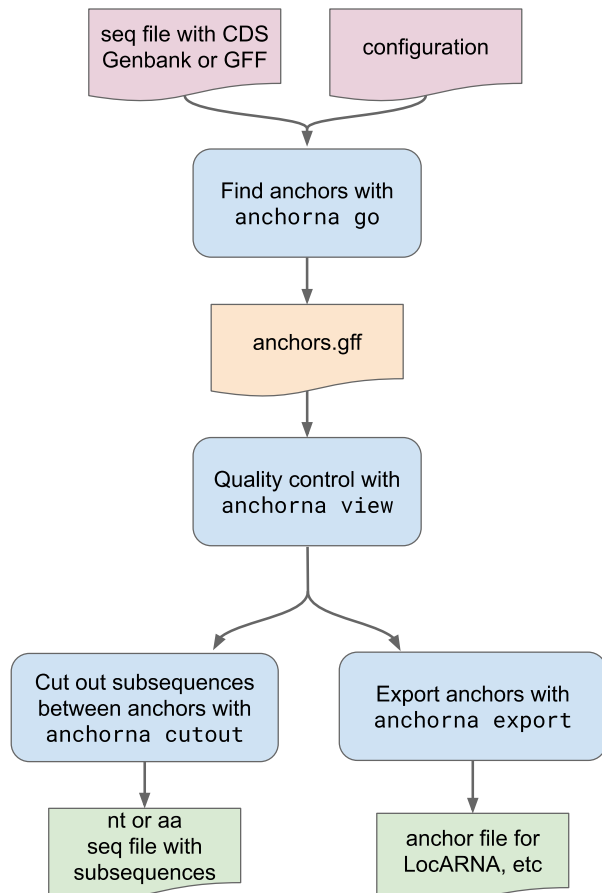
# Algorithm - anchorna go



## Algorithm - anchorna go



# Workflow



anchorna go anchors.gff --njobs 4

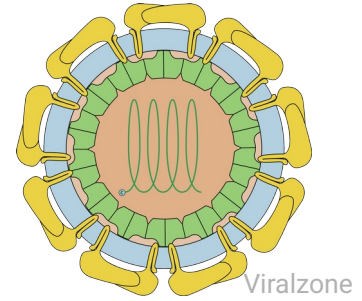


anchorna view anchors.gff --align A1



anchorna cutout anchors.gff A0 A1+6



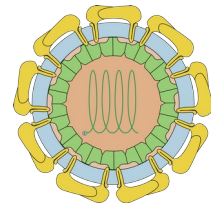


Search in 55 representative *Pestivirus* genomes

56 anchors



Diagram illustrating the organization of the HCV genome. The genome is shown as a linear sequence of proteins: Npro, C, E, E1, E2, p7, NS2-3, NS4A, NS4B, NS5A, and RdRp. The genome is flanked by 5' and 3' ends. The diagram also shows the positions of various open reading frames (ORFs) and specific sites marked by arrows.



BLOSSUM62 36 / 10

5'

- 3'-OH

BLOSSUM62 36 / 10

51 / 21

26 / 20

36 / 9

*A*

[illegible]

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| R | H | G | K | P | I | T | C | G | M | S | L | A | D | F | E | I |
| R | F | G | P | P | I | S | C | G | M | S | L | A | D | F | E | I |
| R | Q | G | K | P | I | N | C | G | M | T | L | A | D | F | E | I |
| R | P | G | K | P | I | T | C | G | M | T | L | A | D | F | E | I |
| R | S | G | K | P | I | S | C | G | M | T | L | A | D | F | E | I |
| R | F | G | P | P | M | T | C | G | M | T | L | A | D | F | E | I |
| R | F | G | P | S | L | S | C | G | M | T | L | S | D | F | E | I |
| R | F | G | P | P | I | S | C | G | M | T | L | A | D | F | E | I |
| R | F | G | P | P | I | T | C | G | M | T | L | A | D | F | E | I |
| R | P | G | K | P | I | S | C | G | M | T | L | A | D | F | E | I |
| G | L | G | P | P | V | K | C | G | M | S | L | A | D | F | E | I |
| S | R | G | T | Q | I | K | C | G | L | T | L | A | E | F | E | I |
| N | R | G | T | G | I | K | C | G | M | T | L | A | E | F | E | I |
| S | K | G | T | Q | I | K | C | G | M | T | L | A | E | F | E | I |
| P | S | R | T | K | L | R | C | G | T | T | L | A | E | F | E | I |
| E | E | L | H | K | G | M | C | G | L | P | T | V | V | Q | N | L |
| S | D | L | R | K | G | M | C | G | L | P | T | P | I | Q | N | L |
| S | D | L | Q | K | G | M | C | G | M | P | T | V | I | Q | N | L |
| A | E | S | P | I | R | C | G | P | W | T | L | K | D | L | E | V |
| A | E | S | P | I | R | C | G | W | T | L | L | K | D | L | E | V |
| V | E | R | O | R | F | I | C | G | T | P | L | R | R | E | F | E |

NUC

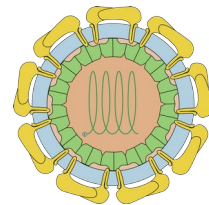
[illegible]

75 / -6

75 / -6

```
CCTCATGGGAAAGCCATCAAGCTGTGGGAATGACCTGGCAAGATTTTGAAGA
CCGCATGGAGAGAGAGGACCTGCTGTGGGATGTGCTGATACGATTTTGAAGA
GAGACGGGGAGAAACCTATCAATTTGGGAATGAACTTGGTGACATTTGAGGA
AGATGGGAAATTAATGAGGAGGAGGAGGAGGAGGAGTGTGCGAGATTTGAGA
GAGAGGCGGCGCCATGCTGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
CTGCTGGCAGGGCCATCATCTGCGGGAATGACCTTGGTGACATTTGAGGA
CGTTTGGTGCGACCAATAGCTGCTGTGGGATGACCTTGGTGACATTTGAGGA
AAGTTTGGGCTGCAACATCTGCTGTGGGATGACCTTGGTGACATTTGAGGA
CGATTGTGGCGCCGACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
CTGTGGGCGCGGCTGTAAAGTGTGGCATGAGTGTGGGCGCATTTGAGGA
AATGAGGAGCACTCAATAAAGTGTGGGCGTACCTGCGCGAGTTTGGAGGA
ACGAGGAGACAGGAATAGAGCTGGGATGACCTTGGTGACATTTGAGGA
ACGACAGGAGACCAATTAAGTGTGGGATGACCTTGGTGACATTTGAGGA
ACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
TGTGACCTTGAAGAAAGATGTGTGGGACCTTCAACCCCTATTACAAACTT
TGTGCGCTCAAAAAGAGGATGTGTGGCATCTGCACAGATGTGTCAAAACCT
GTGGAGGAGCTGAGGATAGATGTGGGCGGAGGAGGAGGAGGAGGAGGAGGAG
CCGATGACCTGAGGAGATGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
CTGCGAGGAGCAGAGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
GTGGAGGAGACAGAGGTTGATGTGTGGGACCCCTACGAGGAGTTTGAAGA
```

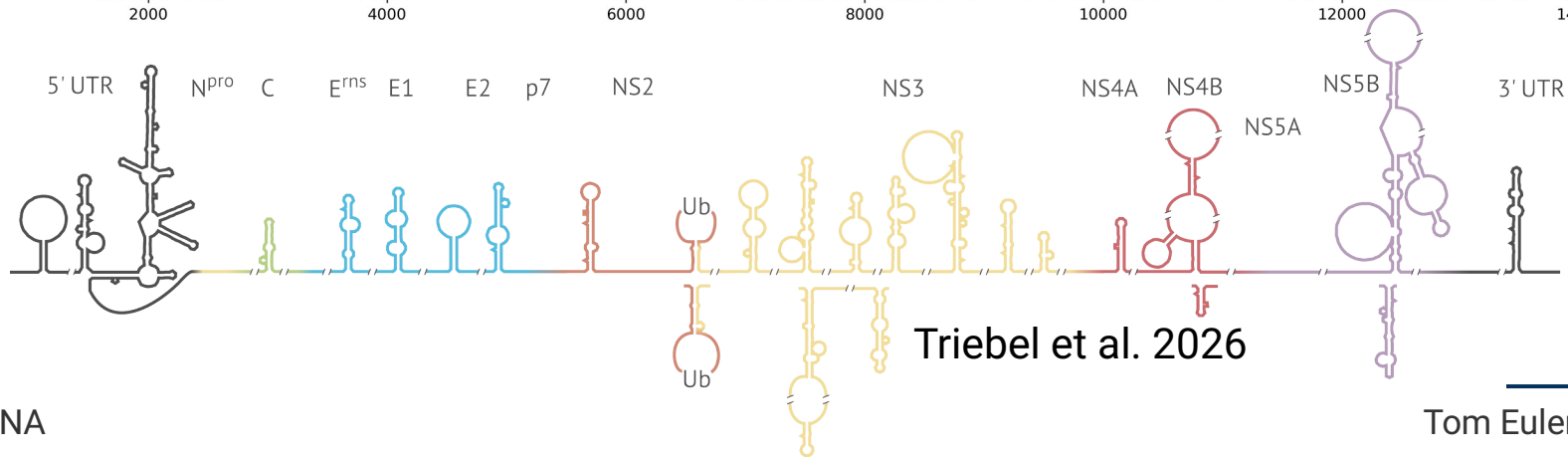
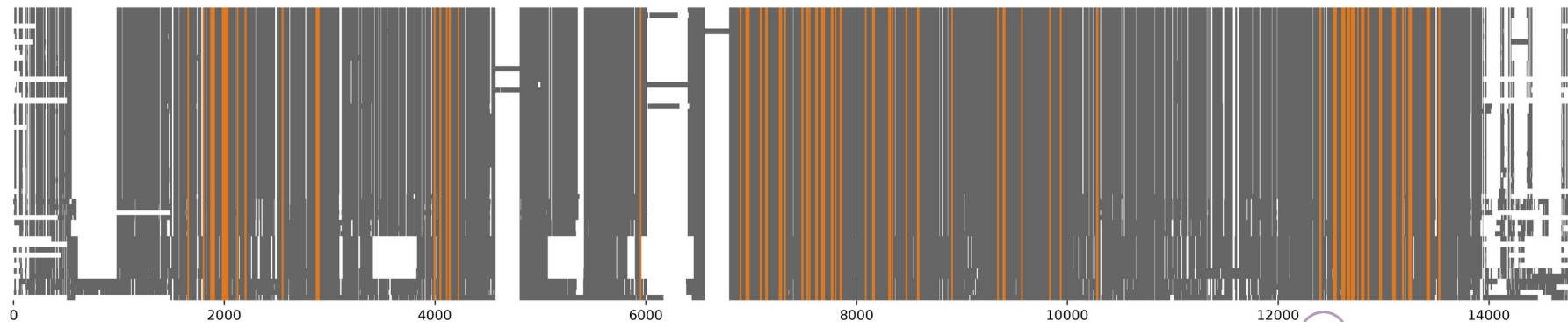
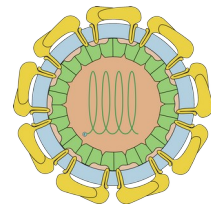
## Performance



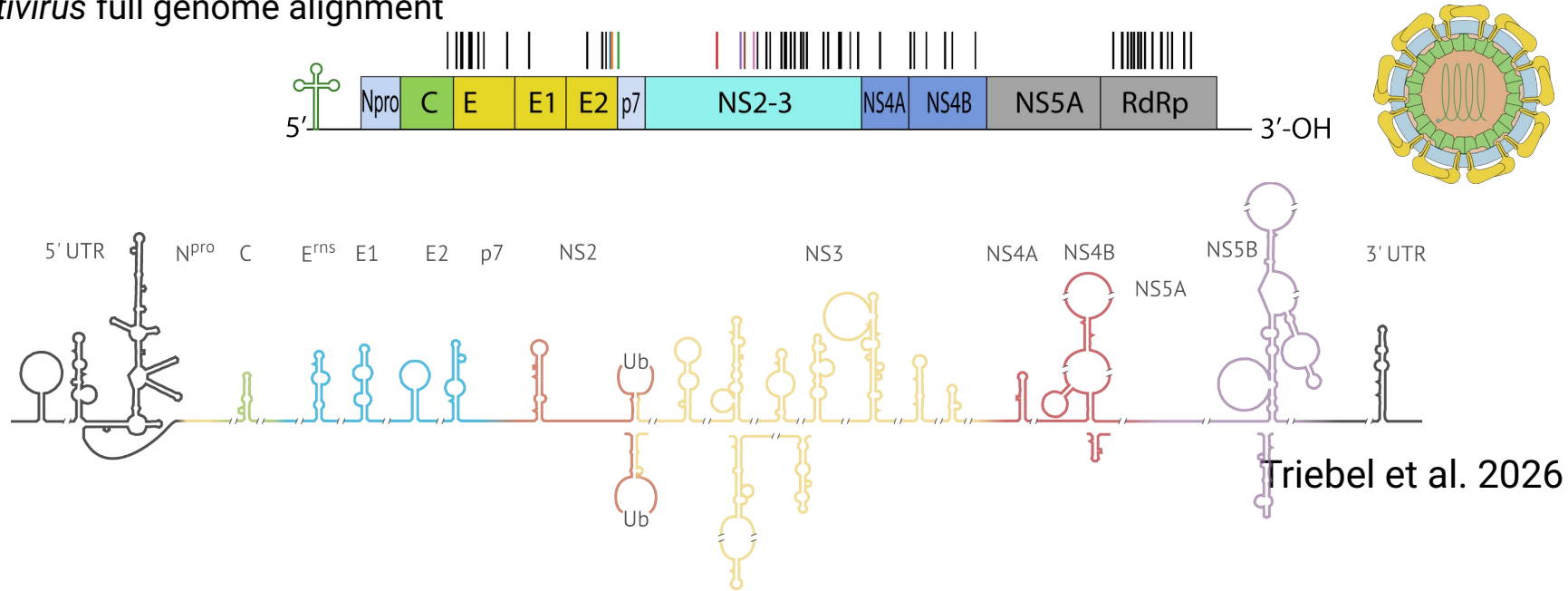
- Full alignment of 50 *Pestivirus* genomes
- Mean alignment score based on NUC score, -11/-1 gap penalty, perfect score is 5.0
- Alignment length

| vanilla      |             |             |                 |       |
|--------------|-------------|-------------|-----------------|-------|
| MAFFT        | 0.41/16100  | 0.53/15500  | anchorna cutout |       |
| Dialign      | -1.00/24900 | -0.62/22300 | anchorna export |       |
| LocARNA      | 0.67/14300  | 0.68/14300  | anchorna export |       |
|              |             | 0.66/14300  | anchorna cutout |       |
| hand-curated | 0.63/14800  |             |                 |       |
| runtime      | ≈ 2 min     | ≈ 7 h       | ≈ 12 h          | ≈ 6 d |

# Pestivirus full genome alignment



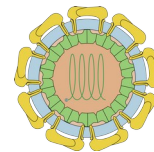
## Pestivirus full genome alignment



- Secondary structure for full virus genome
- Secondary structure present in coding regions

## Summary

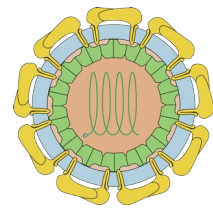
- Identify conserved anchor regions in coding sequences
- Ready usable tool AnchoRNA <https://github.com/rnajena/anchorna>
- Applied to representative set of genomes for *Pestivirus*



Thanks to great RNAJena lab!

Eulenfeld T, Triebel S, Stadler PF and Marz M (2025), AnchoRNA: Full virus genome alignments through conserved anchor regions, *bioRxiv*, doi:10.1101/2025.01.30.635689.

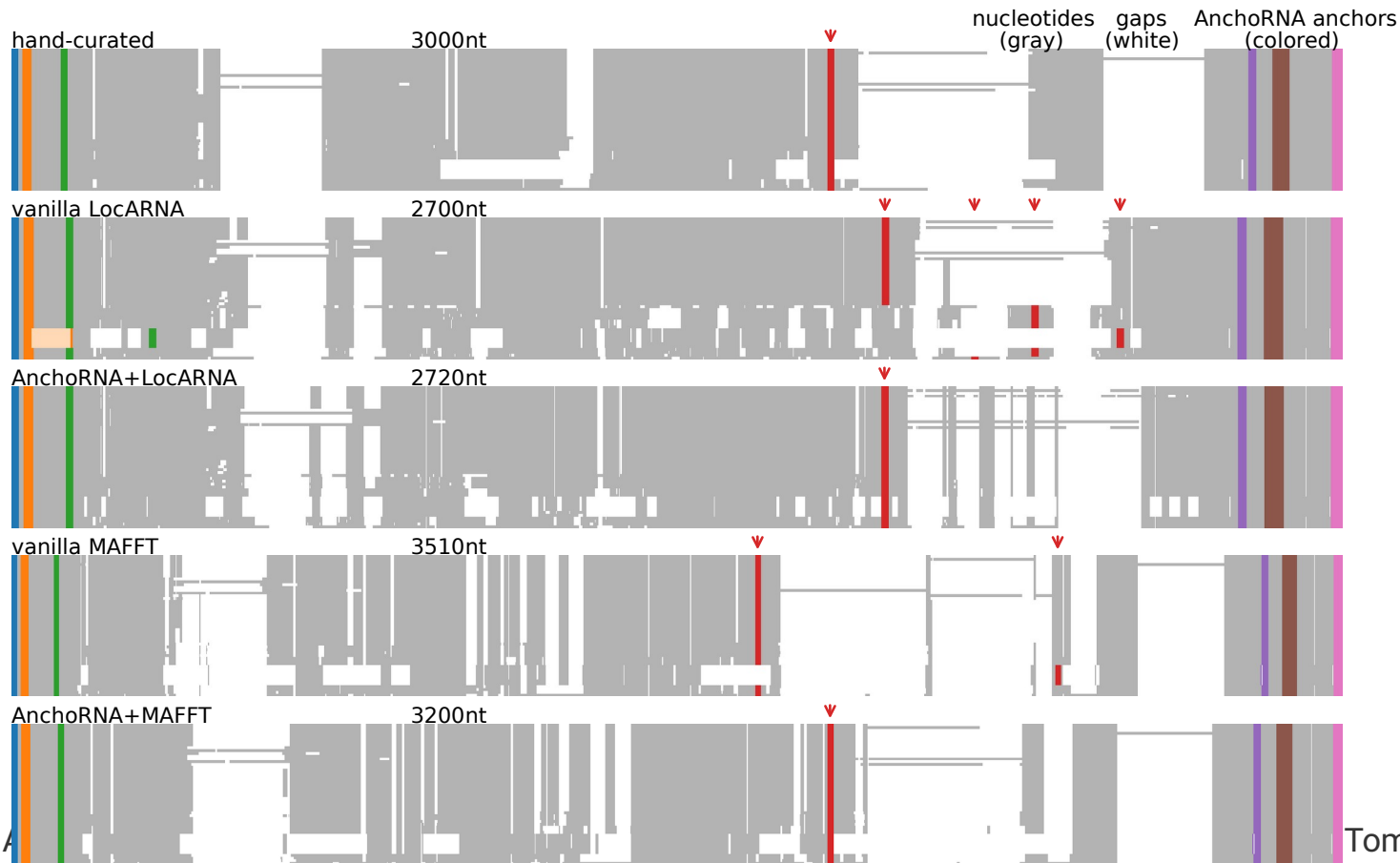
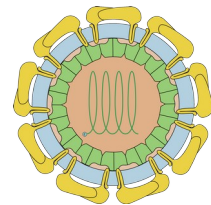
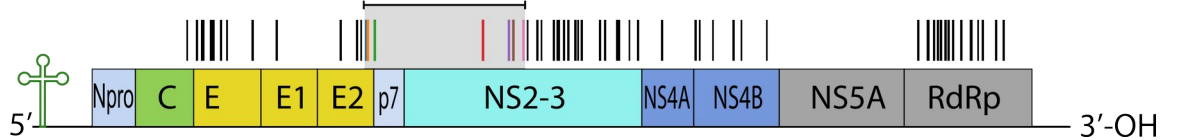
Triebel S, Eulenfeld T, Ontiveros N, Sweeney B, Tautz N, and Marz M (2026), First full-genome alignment representative for the genus *Pestivirus*, *RNA*, doi:10.1261/rna.080732.125.



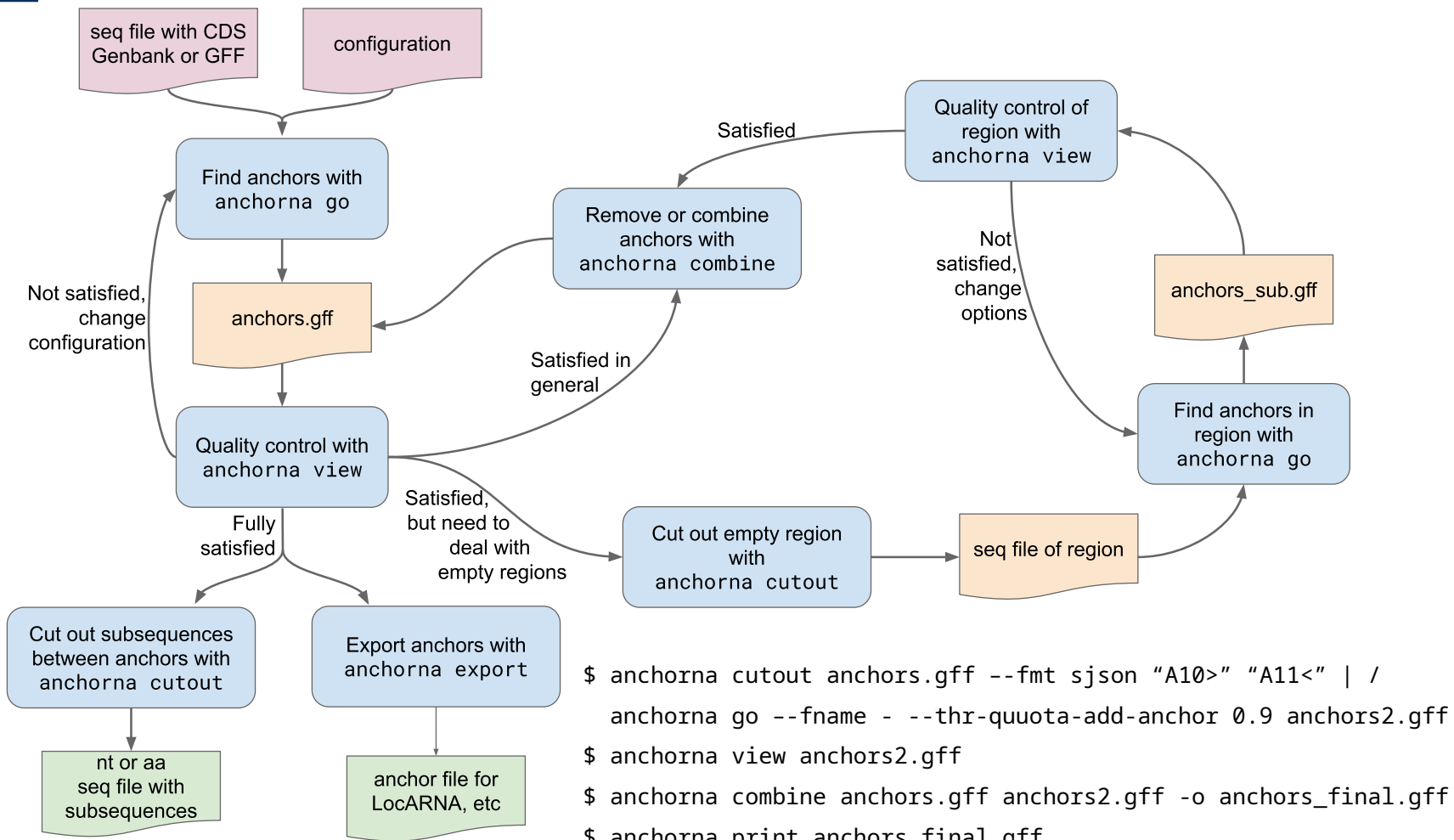
Triebel S, Lamkiewicz K, Ontiveros N, Sweeney B, Stadler PF, Petrov AI, Niepmann M, and Marz M (2024), Comprehensive survey of conserved RNA secondary structures in full-genome alignment of Hepatitis C virus, *Scientific Reports*, 14(1), doi:10.1038/s41598-024-62897-0.

ViralZone: a knowledge resource to understand virus diversity, doi:10.1093/nar/gkq901.

# Performance



## Extended workflow



```
$ anchorna cutout anchors.gff --fmt sjson "A10>" "A11<" | /
  anchorna go --fname - --thr-quuota-add-anchor 0.9 anchors2.gff
$ anchorna view anchors2.gff
$ anchorna combine anchors.gff anchors2.gff -o anchors_final.gff
$ anchorna print anchors_final.gff
```



## A Python framework for bioinformatics



The sugar project aims to provide a lightweight framework to facilitate rapid application development for bioinformatics.

```
pip install rnajena-sugar
```

