RNA secondary structures in RNA viruses

Manja Marz

University of Jena, Germany
Chair for HTS Data Analysis

Director of European Virus Bioinformatics Center

Institute of Data Driven Science: MSCJ
Leibniz Institute for Age Research: FLI
Head of ITN "Viroinf"

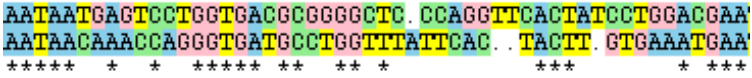
ECT* Structure and topology of RNA in living systems
31.01.2023

I. Evolution and RNA secondary structures

Non-coding RNAs: sequence vs. structure

Non-coding RNAs: sequence vs. structure

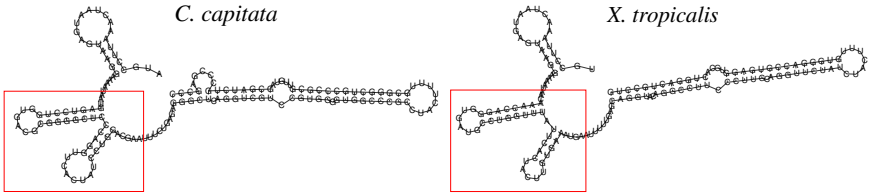
Example: U12 snRNA of *C. capitata* and *X. tropicalis* (nt 25-78)



Non-coding RNAs: sequence vs. structure

Example: U12 snRNA of *C. capitata* and *X. tropicalis* (nt 25-78)

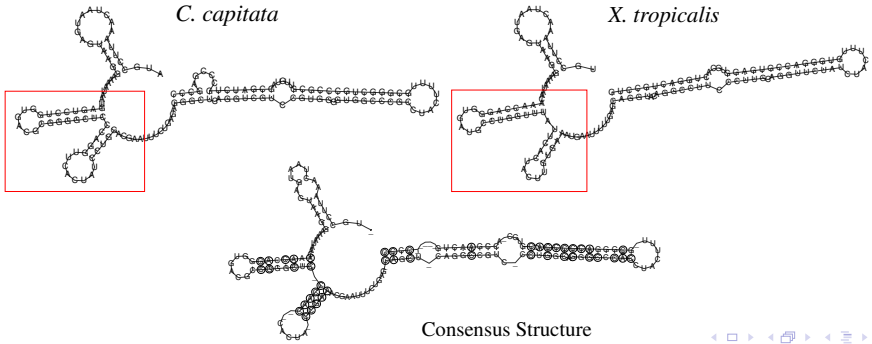
```
AATAATGAGTCCTGGTGACGCGGGGCTC.CCAGGTTCACTATCCTGGACGAA
AATAACAAACCAGGGTGATGCCTGGTTTATTCAC..TACTT.GTGAAATGAA
*****  *  *  ***** **  ** *        ***       * ***
```



Non-coding RNAs: sequence vs. structure

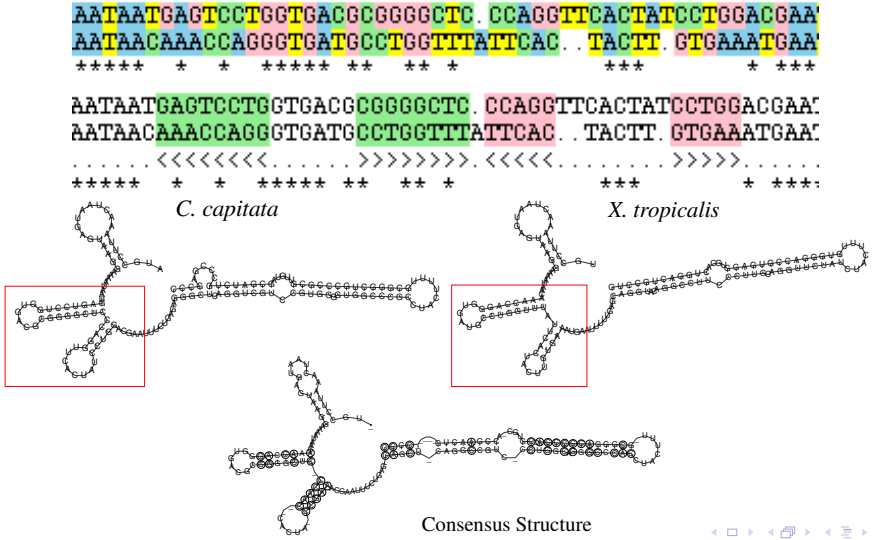
Example: U12 snRNA of *C. capitata* and *X. tropicalis* (nt 25-78)

```
AATAATGAGTCCTGGTGACGCGGGGCTC.CCAGGTTCACTATCCTGGACGAA
AATAACAAACCAGGGTGATGCCTGGTTTATTCAC..TACTT.GTGAAATGAA
*****  *  *  ***** **  ** *        ***       * ***
```



Non-coding RNAs: sequence vs. structure

Example: U12 snRNA of *C. capitata* and *X. tropicalis* (nt 25-78)



take home message

take home message

- RNAfold >=200 nt (!)

take home message

- RNAfold >=200 nt (!) nature of transcription/processing/translation

take home message

- RNAfold >=200 nt (!) nature of transcription/processing/translation
- compensatory mutation information

take home message

- RNAfold >=200 nt (!) nature of transcription/processing/translation
- compensatory mutation information
- clustalw/RNAalifold :(

take home message

- RNAfold >=200 nt (!) nature of transcription/processing/translation
- compensatory mutation information
- clustalw/RNAalifold :(vs. locarna (Sankoff) :)

take home message

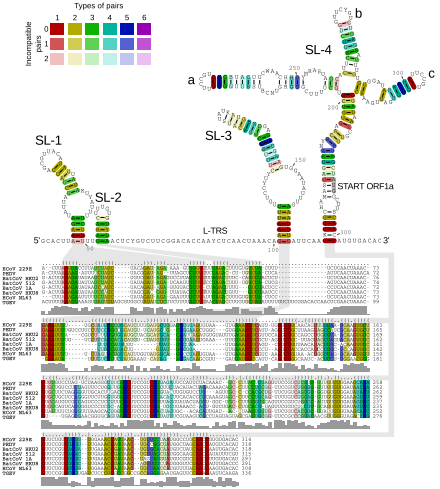
- RNAfold >=200 nt (!) nature of transcription/processing/translation
- compensatory mutation information
- clustalw/RNAalifold :(vs. locarna (Sankoff) :)
- proof by

take home message

- RNAfold >=200 nt (!) nature of transcription/processing/translation
- compensatory mutation information
- clustalw/RNAalifold :(vs. locarna (Sankoff) :)
- proof by di(!)nucleotide-shuffling

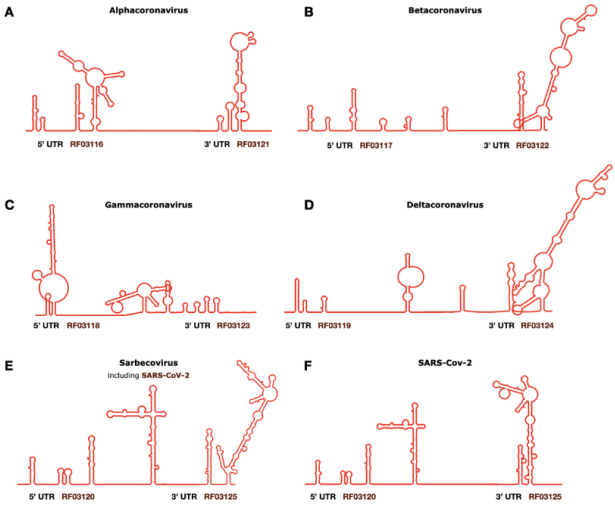
Secondary structures in RNA viruses: Coronaviruses

Secondary structures in RNA viruses: Coronaviruses



Madhugiri *et al.*, Virology, 2018; Madhugiri *et al.*, Adv Virus Res, 2016
Madhugiri *et al.*, Virus Res, 2014

SARS-CoV2 UTRs and full genome alignments



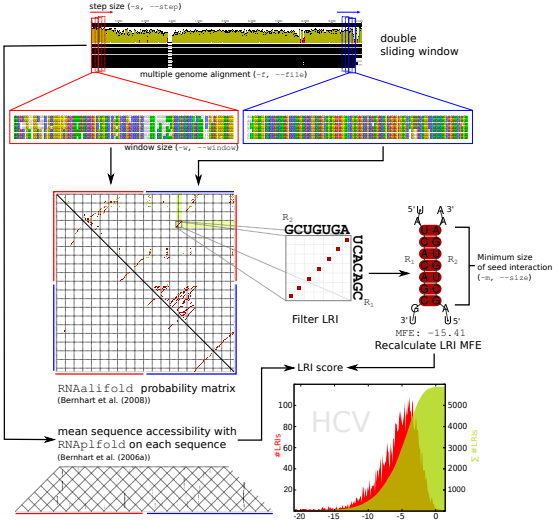
Kalvari *et al.*, Nucleic Acids Res, 2021

Secondary structures in mRNAs?

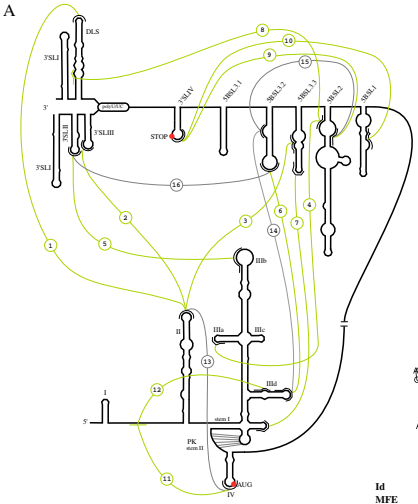
Secondary structures in mRNAs?

- Coronaviruses
- HCV
- Filoviruses
- Flaviviruses
- Pestiviruses
- ...

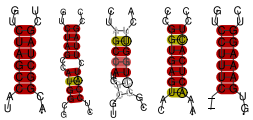
Secondary structures in RNA viruses: Long-range interactions



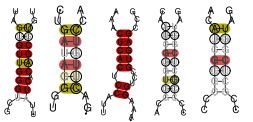
Impact of secondary structures in RNA viruses: Long-range interactions in HCV



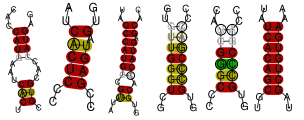
B



Id	1	2	3	4	5
MFE	-13.21	-6.64	-7.47	-11.83	-9.51
#Isolates	19	19	100	106	19

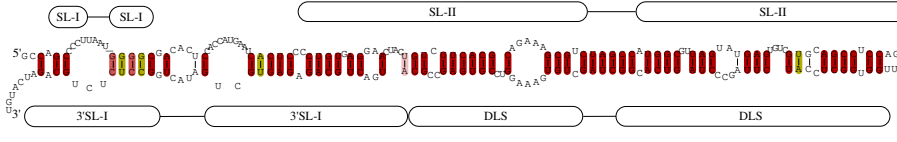


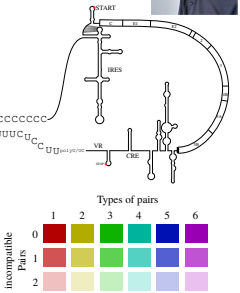
Id	6	7	8	9	10
MFE	-8.29	-6.73	-11.62	-12.75	-14.40
#Isolates	104	106	19	102	101



Id	11	12	13	14	15	16
MFE	-9.53	-5.55	-16.24	-13.31	-18.59	-15.41
#Isolates	82	84	102	104	104	19

Secondary structures in RNA viruses: Circularization of HCV





Fricke *et al.*, RNA, 2015

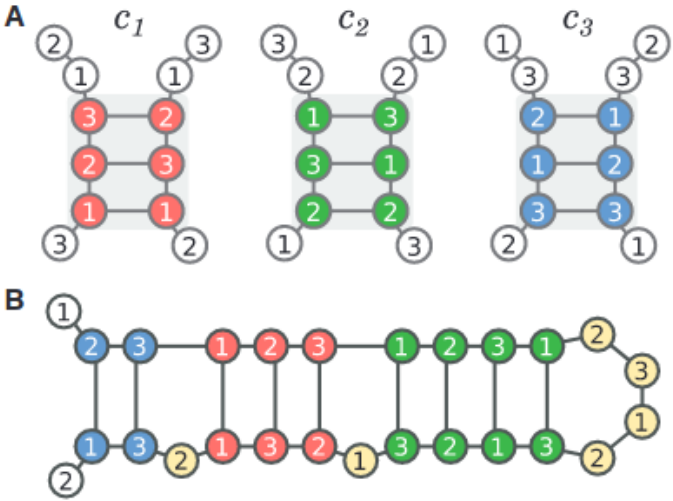
RNA virus packaging – Influenza A Virus

RNA virus packaging – Influenza A Virus

Ib. RNA secondary structures in protein-mRNA

RNA secondary structures in proteins

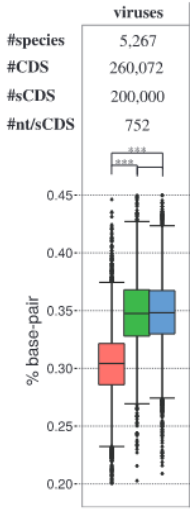
RNA secondary structures in proteins



Fricke *et al.*, 2019

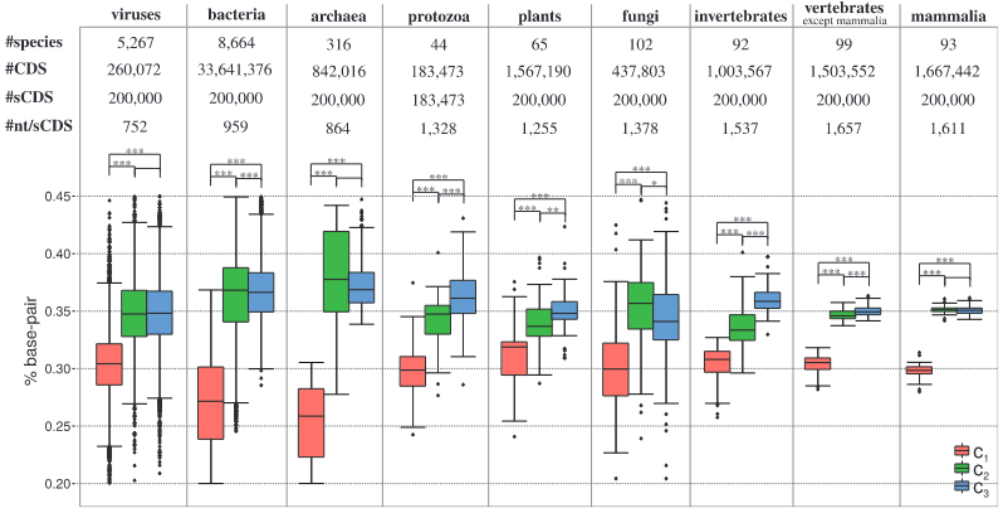
What is going on in viruses?

What is going on in viruses?



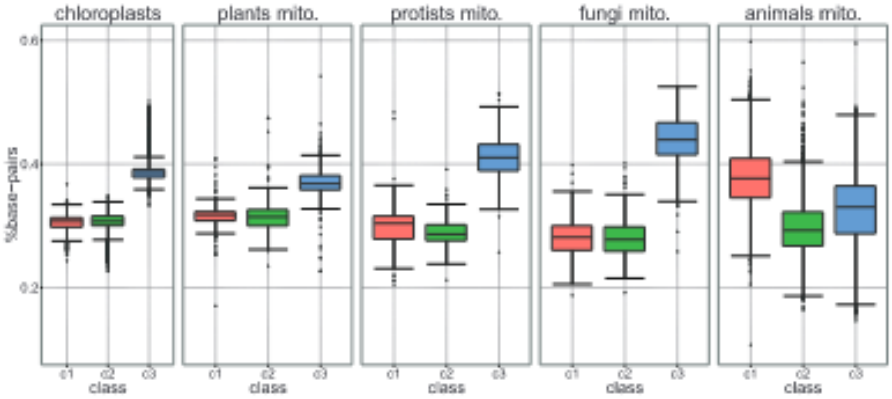
Fricke *et al.*, 2019

Oh .. it appears in all organisms



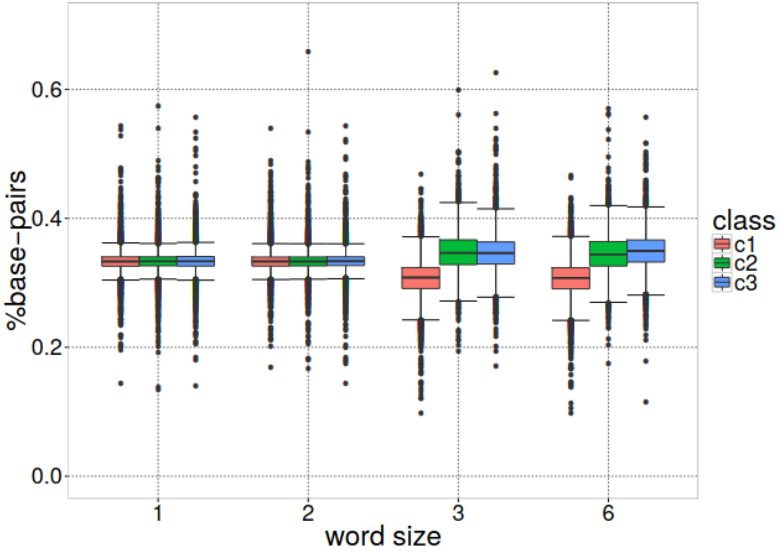
Fricke *et al.*, 2019

Chloroplasts and mitochondria

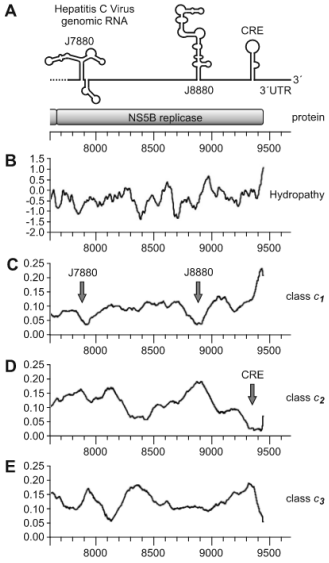


Fricke *et al.*, 2019

Shuffling



Proof of principle

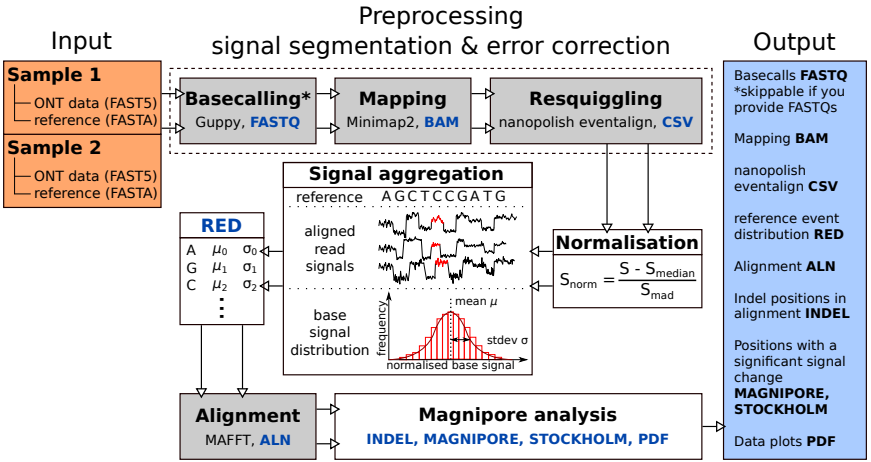


II. ONT and RNA modifications

New generation of sequencing methods: Direct RNA Seq

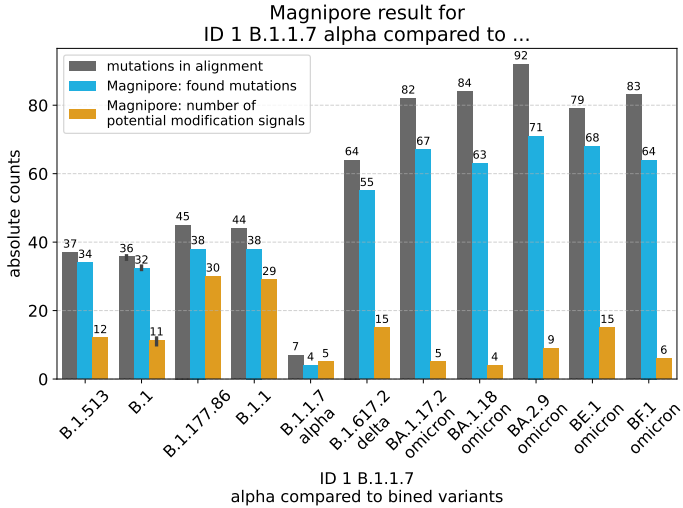
Minion

Magnipore: Differential single nucleotide detection



submitted, NAR, 2023

Magnipore: Mutation verification



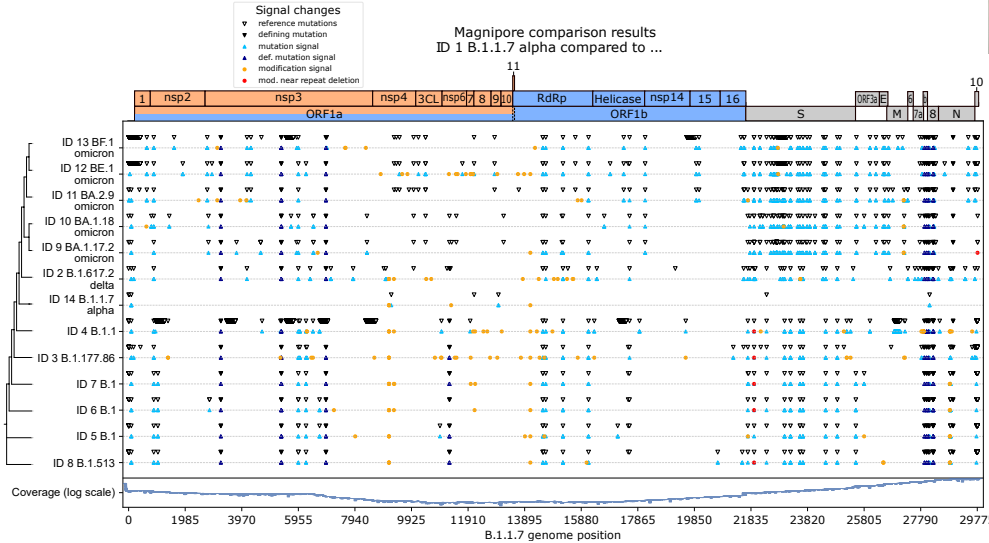
submitted, NAR, 2023

Magnipore: Biological interpretation

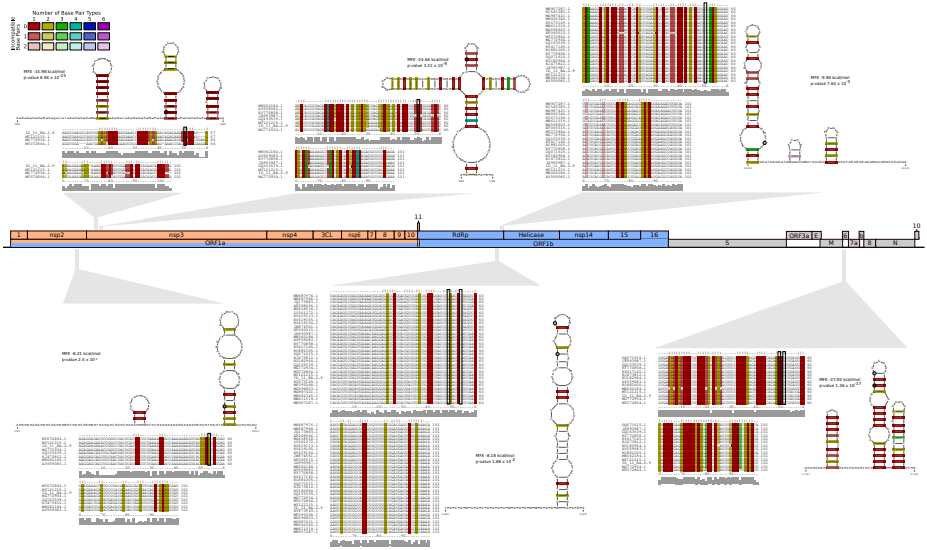
Reference alignment	Magnipore signal diff.	Designation	Symbol
mismatch	✓	base & signal mismatch mutation	▲
GGCCCUA GGCACUA	✗	undetected mutation or error in reference	should not occur
mismatch with gap	✓	indel mutation	▲
GGCACUACUA GGCA---UUA	✗	undetected mutation or error in reference	should not occur
match	✓	potential modification	●
GGCACUA GGCACUA	✗	bases & signals match	standard
match with gap	✓	potential modification near repeat deletion	●
GGCACUACUA GGCA---CUA	✗	bases & signals match	standard

submitted, NAR, 2023

Modification vs. secondary structures



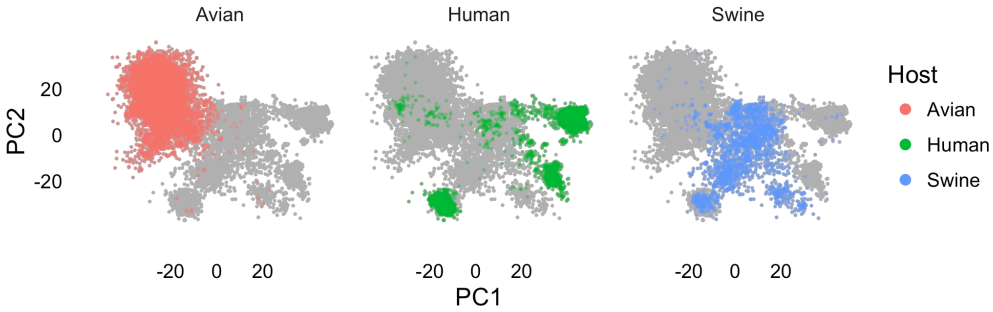
Modification vs. secondary structures



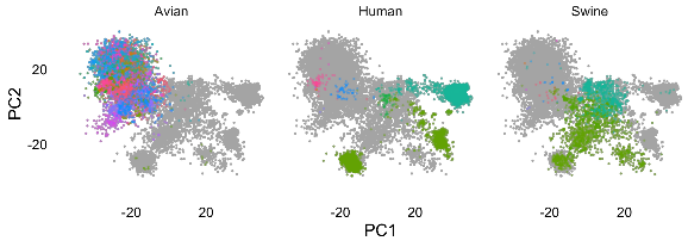
III. RNA based virus host prediction

Give me your virus and I tell you the host

Give me your virus and I tell you the host



Give me your virus and I tell you the host



Subtype

H10N1	H1N1	H3N8	H6N3
H10N2	H1N2	H3N9	H6N5
H10N3	H1N3	H4N1	H6N6
H10N4	H1N6	H4N2	H6N8
H10N5	H1N8	H4N3	H7N1
H10N6	H1N9	H4N4	H7N2
H10N7	H2N1	H4N6	H7N3
H10N8	H2N2	H4N8	H7N4
H10N9	H2N3	H4N9	H7N6
H11N1	H2N5	H5N1	H7N7
H11N2	H2N7	H5N2	H7N8
H11N3	H2N8	H5N3	H7N9
H11N8	H2N9	H5N5	H8N4
H11N9	H3N1	H5N6	H9N1
H12N5	H3N2	H5N8	H9N2
H13N6	H3N3	H5N9	H9N5
H13N8	H3N5	H6N1	
H16N3	H3N6	H6N2	

Give me your virus and I tell you the host

$y \backslash \hat{y}$	Avian	Human	Swine	All
Avian	3207	49	13	3269
Human	6	4470	82	4558
Swine	9	10	849	868
All	3222	4529	944	8695

metric \ host	Avian	Human	Swine	All
accuracy				0.95
recall	0.99	0.95	0.94	
precision	0.88	1.00	0.89	
F1-score	0.94	0.97	0.93	

Codon usage

$y \backslash \hat{y}$	Avian	Human	Swine	All
Avian	5524	227	576	6327
Human	168	11314	1251	12733
Swine	25	800	995	1820
All	5717	12341	2822	20880

Dinucleotides

$y \backslash \hat{y}$	Avian	Human	Swine	All
Avian	4294	1048	985	6327
Human	173	8376	4184	12733
Swine	17	630	6342	1820
All	4484	10054	6342	20880

VIDHOP: VIrus Deep learning HOst Prediction

VIDHOP: VIrus Deep learning HOst Prediction

Examples: Influenza Virus A, Rabies Lyssavirus, Rotavirus

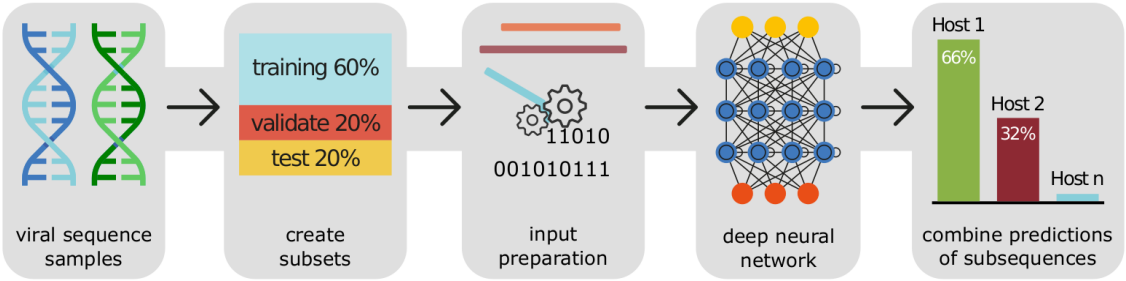
- G1 select, preprocess and condense viral sequences with little information loss
- G2 handle highly unbalanced data sets
- G3 present the output userfriendly

VIDHOP: VIrus Deep learning HOst Prediction

Examples: Influenza Virus A, Rabies Lyssavirus, Rotavirus

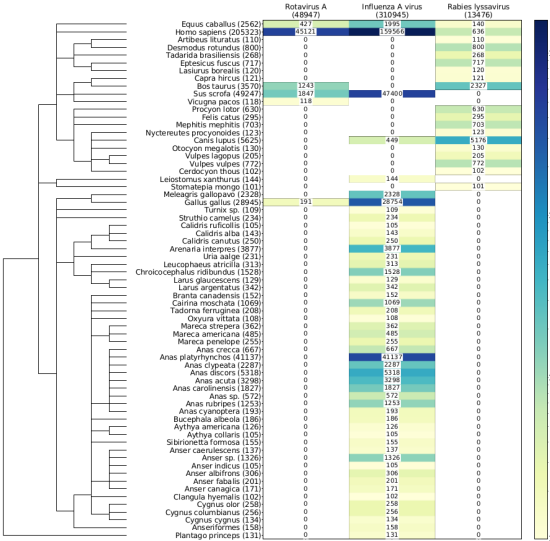
- G1 select, preprocess and condense viral sequences with little information loss
- G2 handle highly unbalanced data sets
- G3 present the output userfriendly

Workflow:

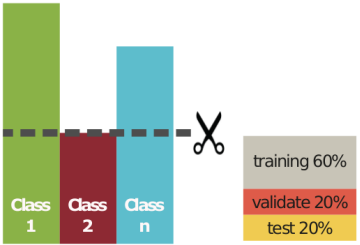


Mock *et al.*, 2020

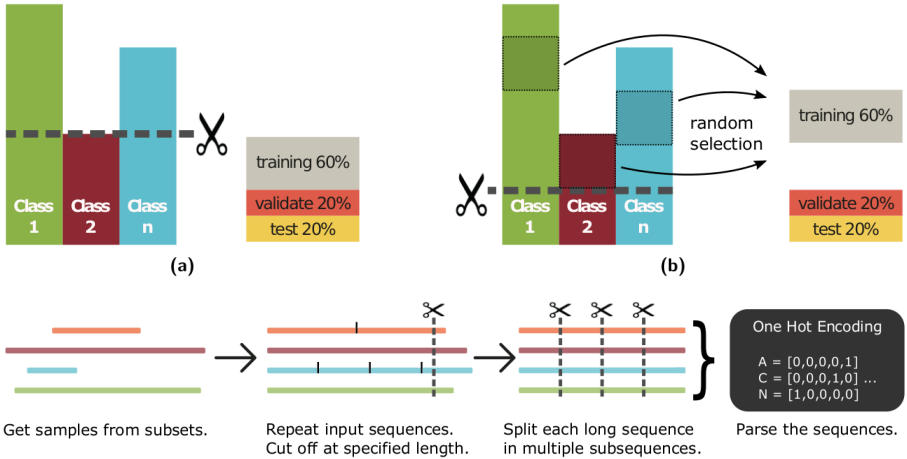
VIDHOP Input Data



VIDHOP Divide and Preparation Data



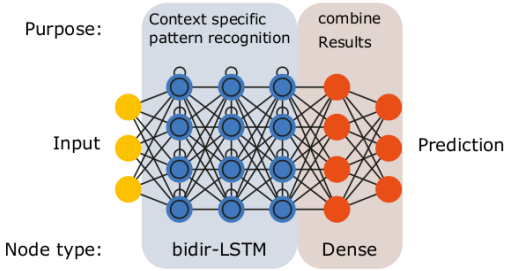
VIDHOP Divide and Preparation Data



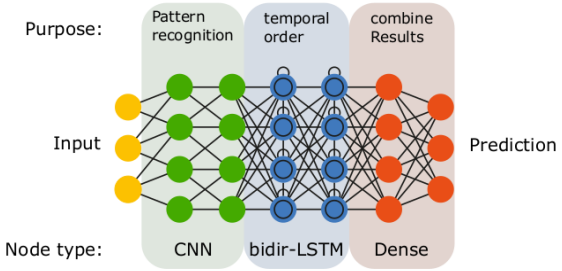
Mock *et al.*, 2020

VIDHOP Deep learning

use information vs. overfitting



(a)



(b)

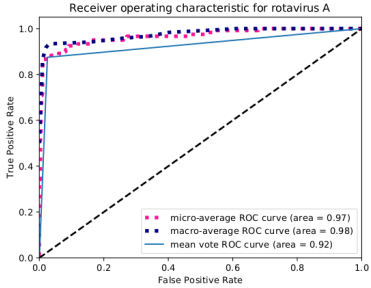
Bidirectional long short-term memory $\in$ artificial recurrent neural network

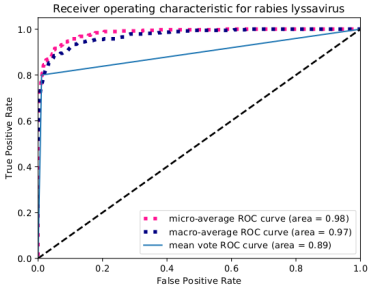
Convolutional neural networks, 150 nodes
500 epochs

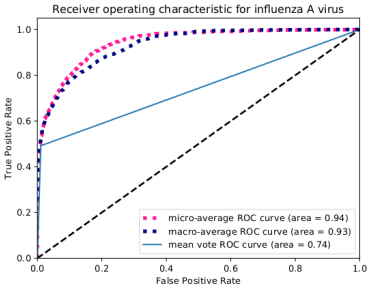
Mock *et al.*, 2020

VIDHOP Results

combination method vs training setup	Standard	Voting	Mean	Std-div	ACC by chance	AUC
LSTM *rotavirus A*, normal repeat gaps	85.83	87.50	86.67	87.50	16.67	98%
CNN+LSTM *rotavirus A*, random repeat	81.30	85.00	85.00	85.83	(6)	
LSTM *rabies lyssavirus*, random repeat	74.02	79.21	80.00	80.00	5.26	98%
CNN+LSTM *rabies lyssavirus*, random repeat	71.98	77.11	77.63	77.63	(19)	
LSTM *influenza A*, random repeat	44.20	47.85	49.18	49.29	2.04	94%
CNN+LSTM *influenza A*, normal repeat gaps	43.53	47.35	49.39	49.39	(49)	







Mock *et al.*, 2020

Thank you!

European VIRUS BIOINFORMATICS Center

Understanding virus-host interactions

DFG

Bundesministerium für Bildung und Forschung