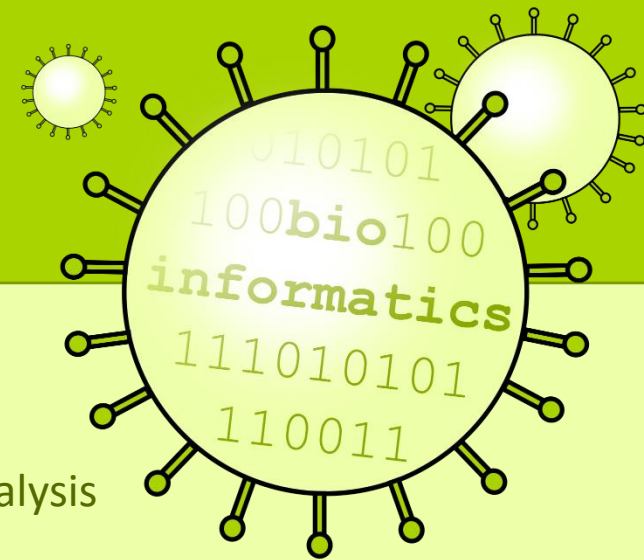


RNA structure analysis and conserved long-range RNA-RNA interaction prediction of full viral RNA genomes

31st TBI Winterseminar

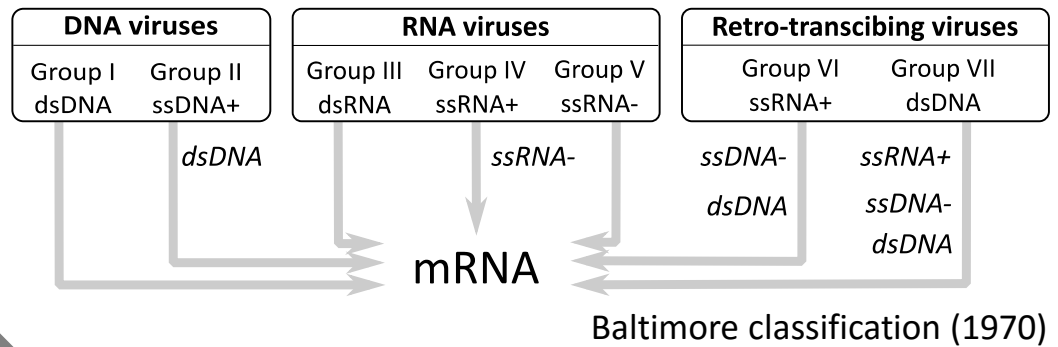
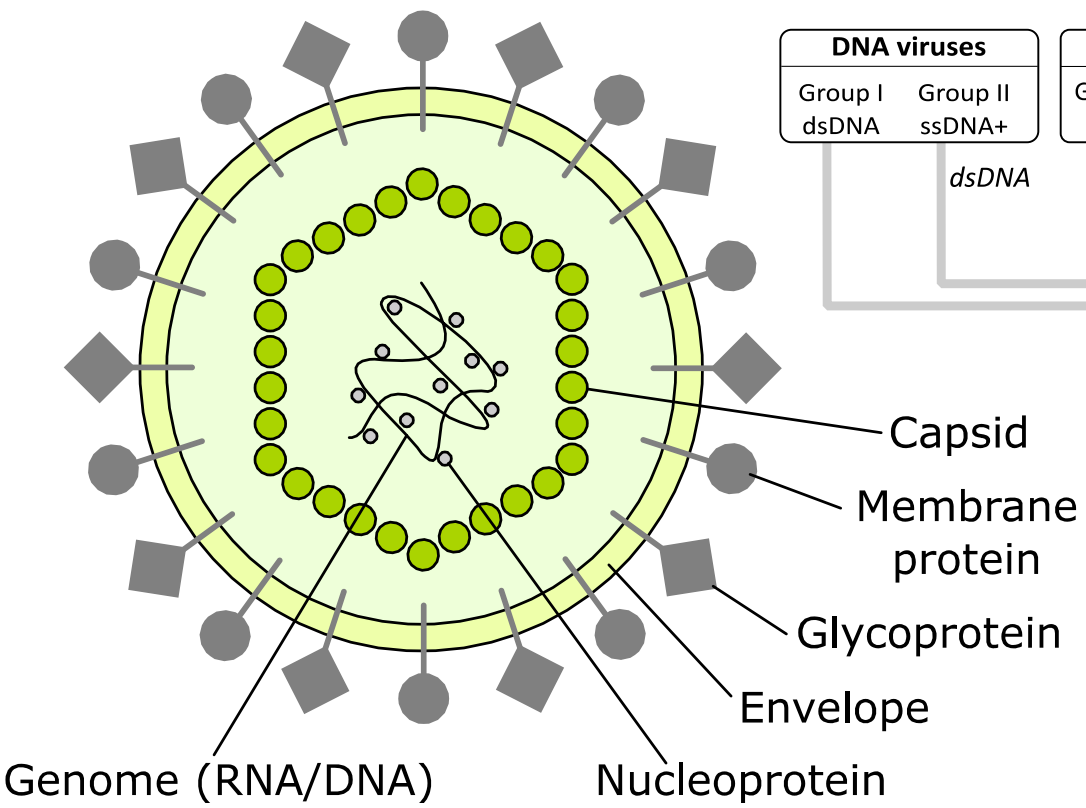
Bled



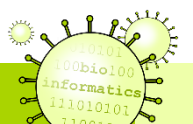
Dipl. Bioinf. **Markus Fricke**
RNA Bioinformatics and High Throughput Analysis
Friedrich Schiller University Jena

What is a virus?

- discovered at the end of the 19. century by *Louis Pasteur*
- virus - Latin *poison* or *slime*
- only 3186 viral species are known today (ICTV 2014)
- at least in mammalian hosts 320,000 viruses are unknown (Anthony2013)



- ICTV classification**
- Order (-virales)
 - Family (-viridae)
 - Subfamily (-virinae)
 - Genus (-virus)
 - Species



Species:

“a monophyletic group of viruses whose properties can be distinguished from those of other species by multiple **criteria**”

Criteria:

“criteria may include, but are not limited to, natural and experimental host range, cell and tissue tropism, pathogenicity, vector specificity, antigenicity, and the degree of relatedness of their genomes or genes. The criteria used should be published in the relevant section of the ICTV Report and reviewed periodically by the Study Group”



International Committee on Taxonomy of Viruses

VIROLOGY DIVISION - IUMS



Hepatitis C virus and Coronaviruses

	Hepatitis C virus (HCV)	Coronaviruses (CoV)
Classification	<i>Order</i> Unassigned <i>Family</i> Flaviviridae	<i>Order</i> Nidovirales <i>Family</i> Coronaviridae <i>Subfamily</i> Coronavirinae
Genus	Hepatitis C virus	Alpha-, Beta-, Gamma- and Deltacoronaviruses
Genome	plus-strand RNA (9.6 kb)	plus-strand RNA (30 kb)
Transmission	Sexual contact, blood	Respiratory or fecal-oral in humans
Diseases	acute or chronic liver diseases liver cirrhosis hepatocellular carcinoma	Mainly respiratory diseases (pneumonia) and gastroenteritis





More than **1 MILLION PEOPLE DIE** each year
from disease caused by **hepatitis B & C**.

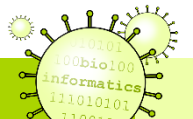


Most of those infected don't know they have it,
INCREASING THE RISK of developing **severe**
liver disease and transmitting the virus to others.

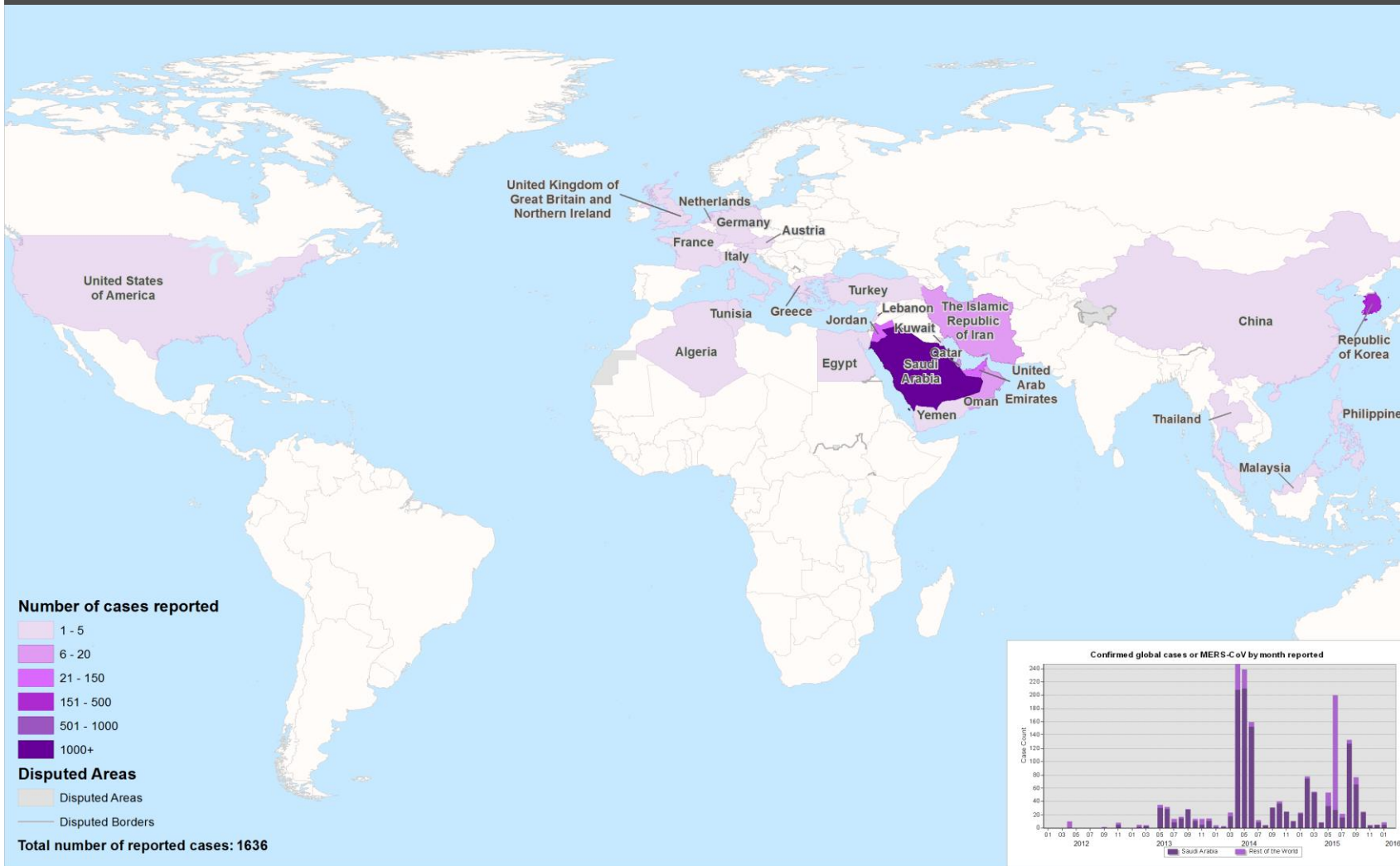


**World Health
Organization**

© World Health Organization 2014



CONFIRMED GLOBAL CASES OF MERS-COV 2012 - 2016



Map Scale (A3): 1:1,109,175,783
 1 cm = 11,092 km
 Coordinate System: GCS WGS 1984
 Datum: WGS 1984
 Units: Degree

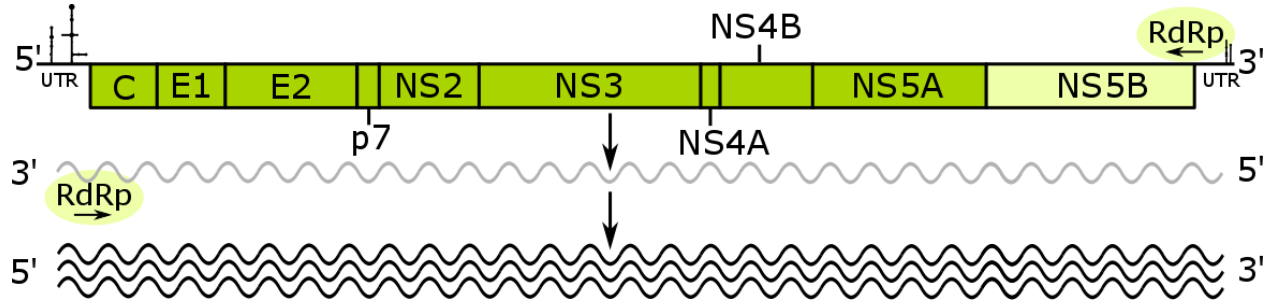
The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

Data Source: World Health Organization
 © WHO 2014. All rights reserved.
 Map date: 05/02/2016

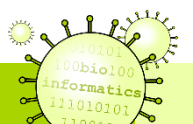
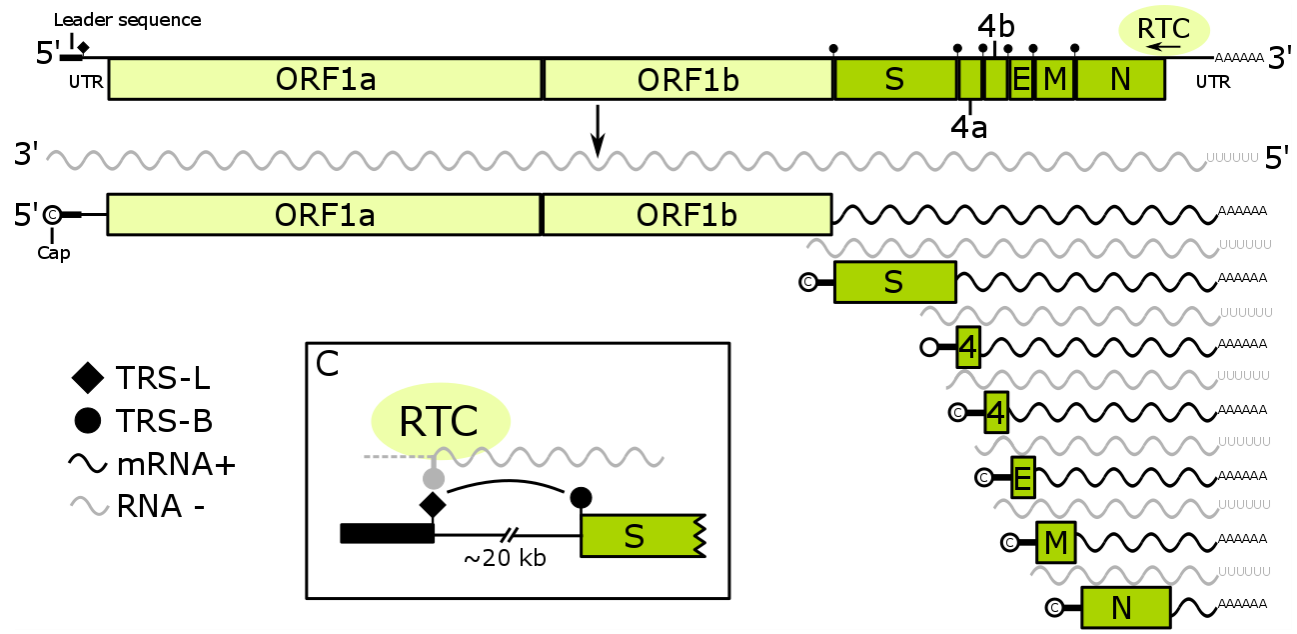


Replication of HCV and CoV (Group IV)

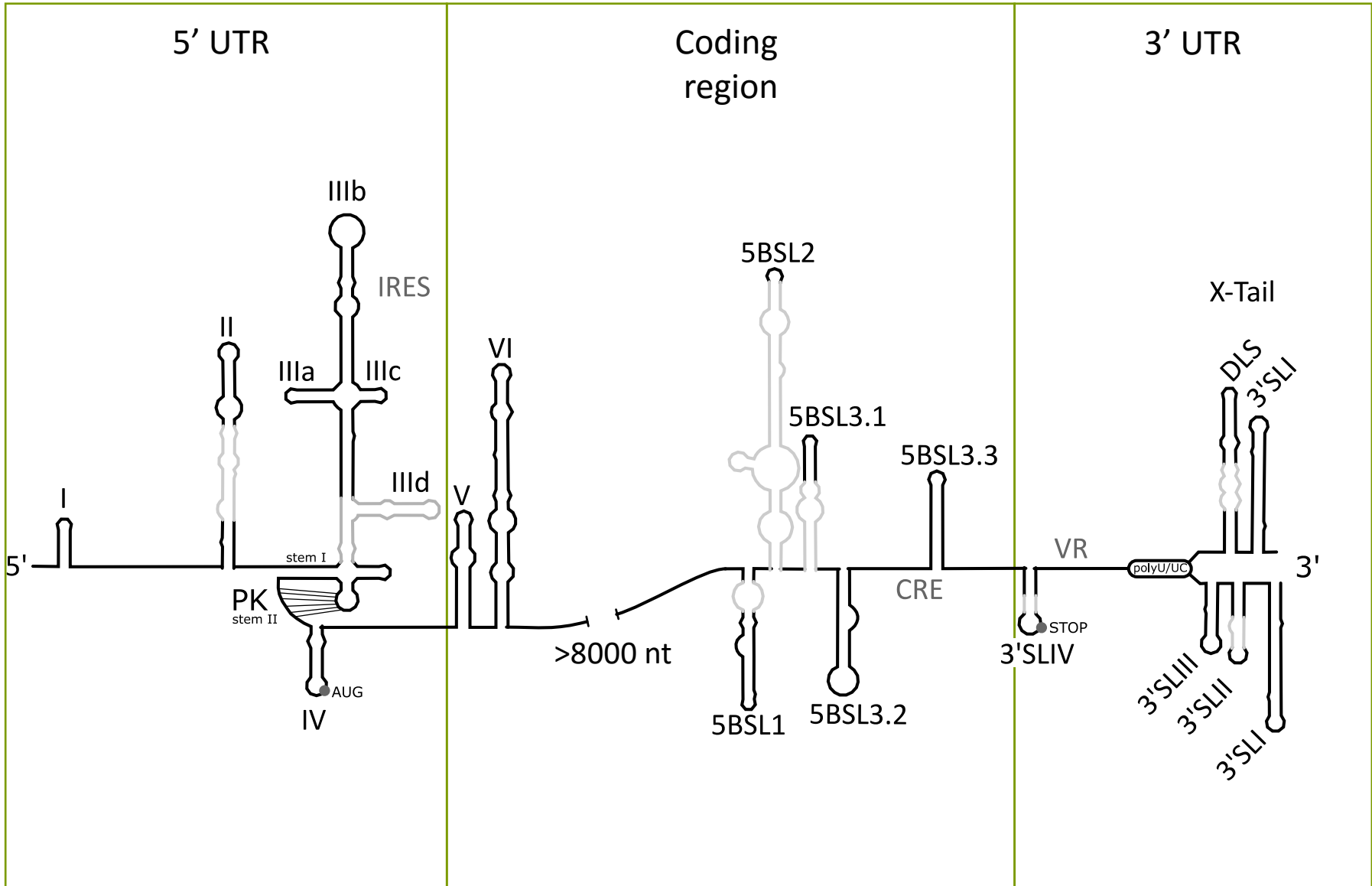
HCV



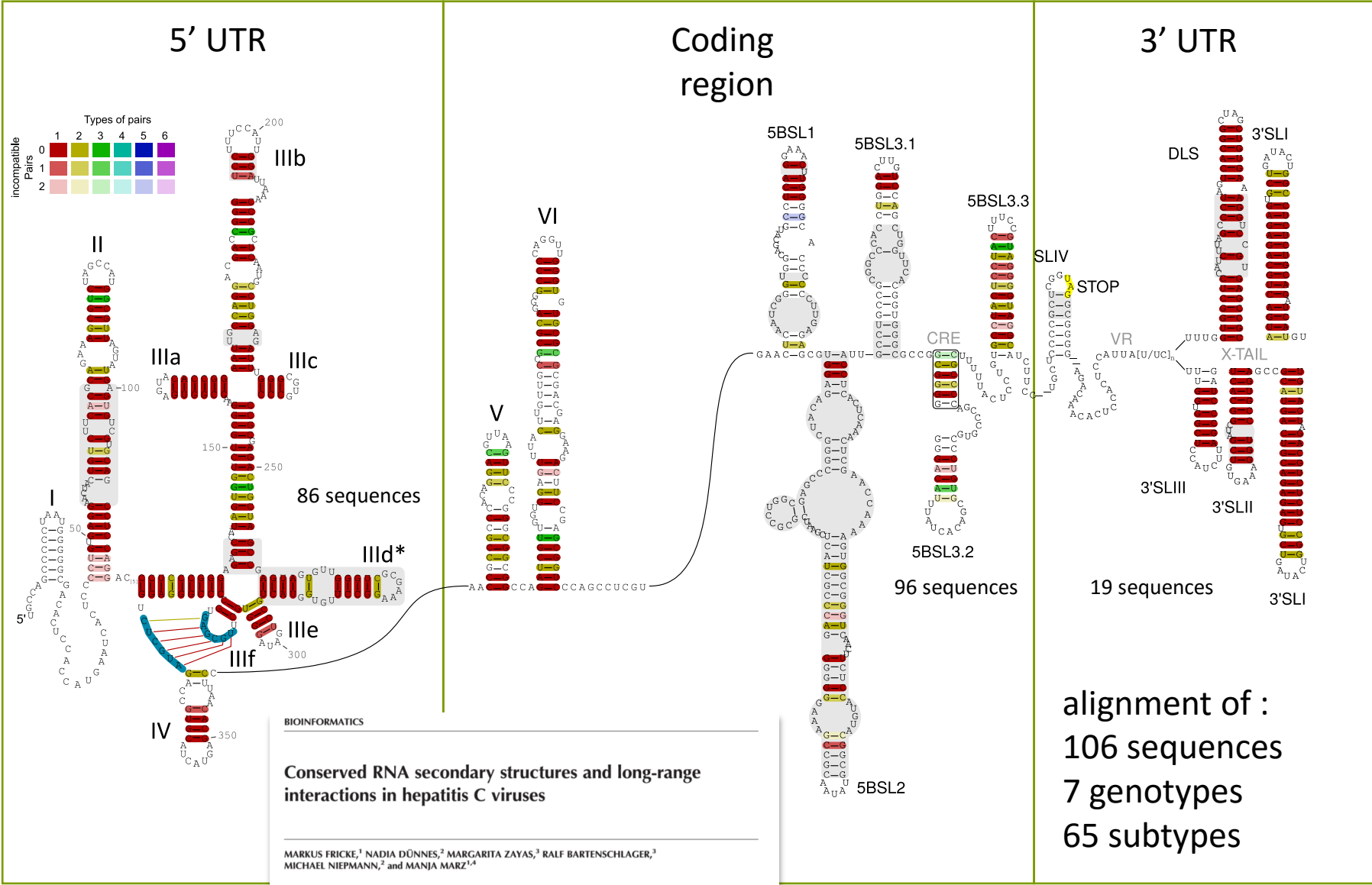
CoV



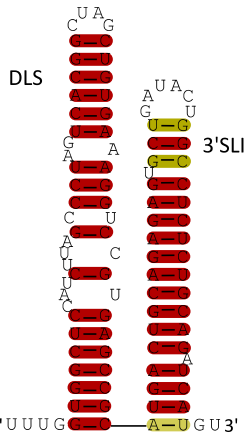
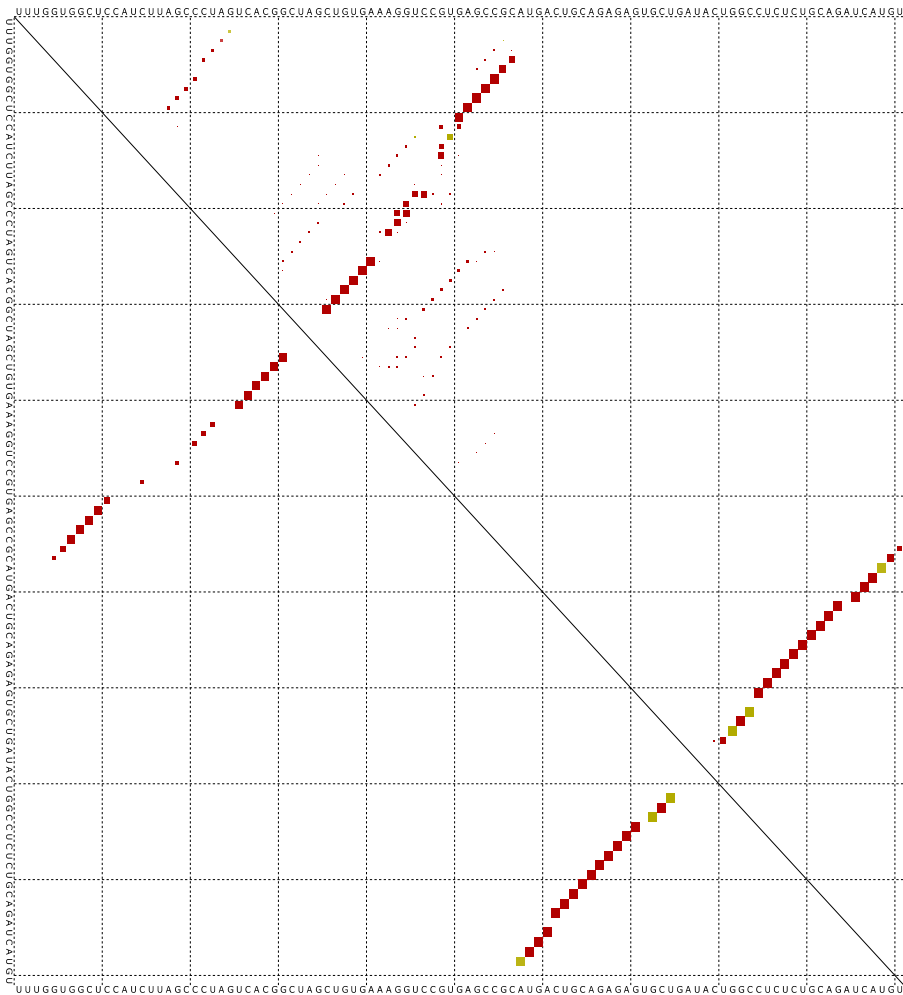
Known RNA secondary in HCV



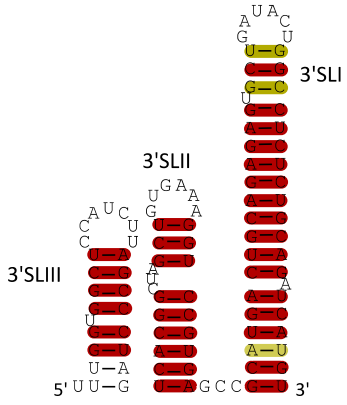
Conserved RNA secondary in HCV



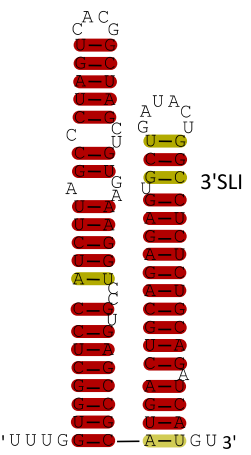
Conserved RNA secondary in HCV



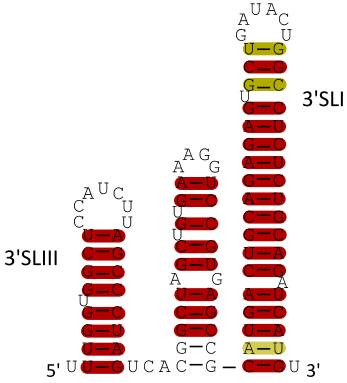
MFE = -66.43 = -43.52 + -22.91 (kcal/mol)



MFE = -65.19 = -44.13 + -21.06 (kcal/mol)



MFE = -65.13 = -43.38 + -21.75 (kcal/mol)

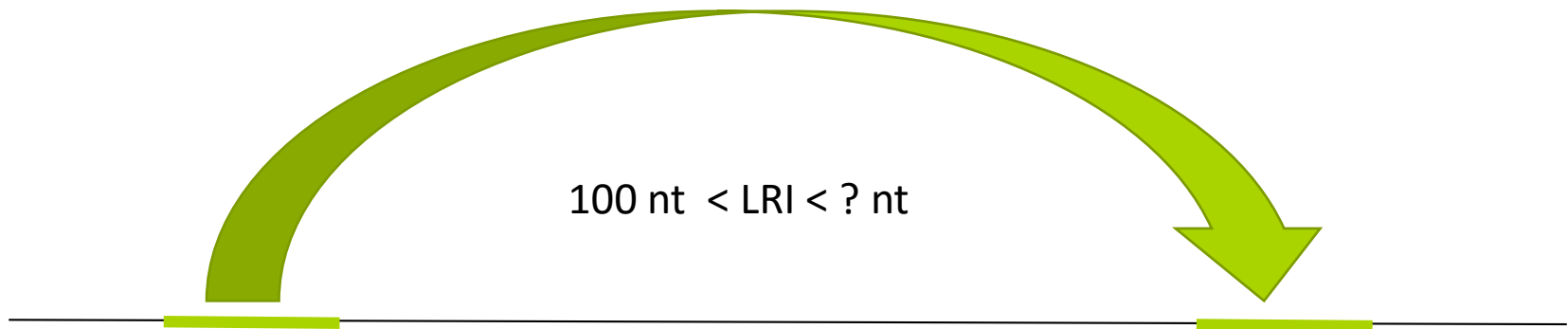


MFE = -65.25 = -42.53 + -22.72 (kcal/mol)



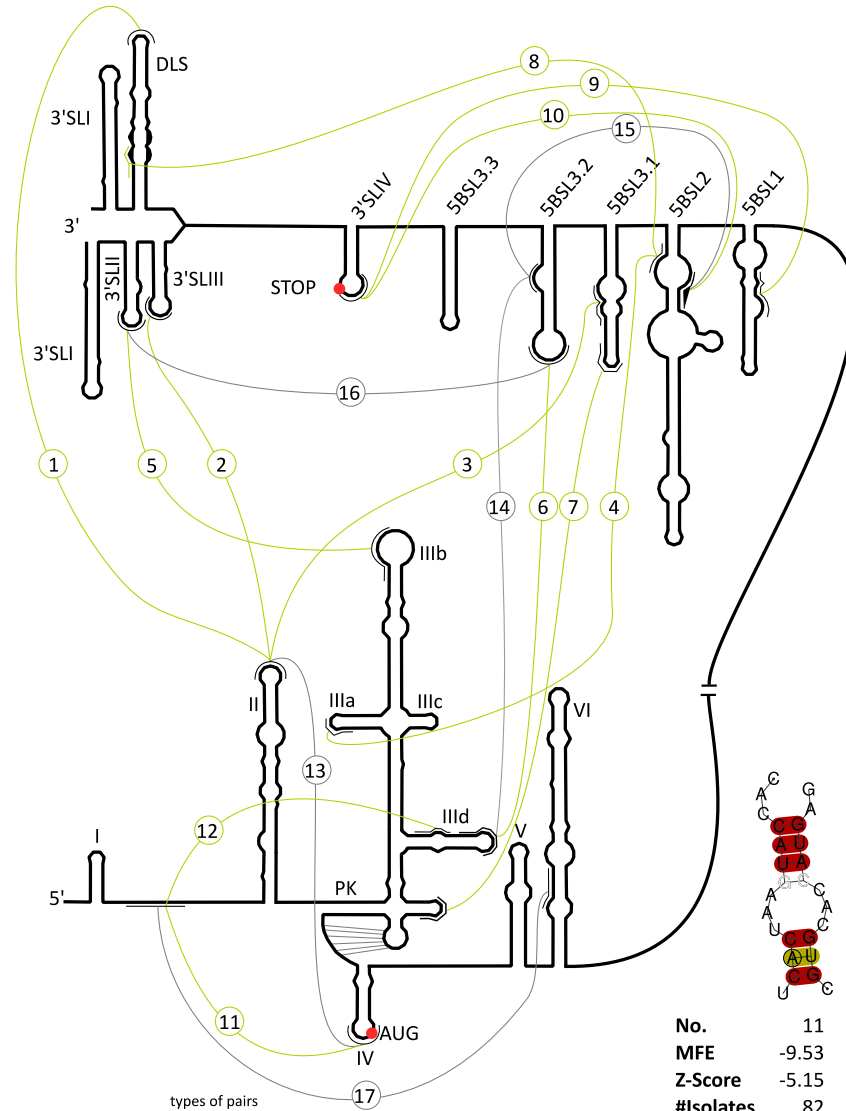
LRI

Long-range RNA-RNA interactions



According to their definition a LRI spans distances between a few hundred and several thousands of nucleotides

Conserved LRIs in HCV

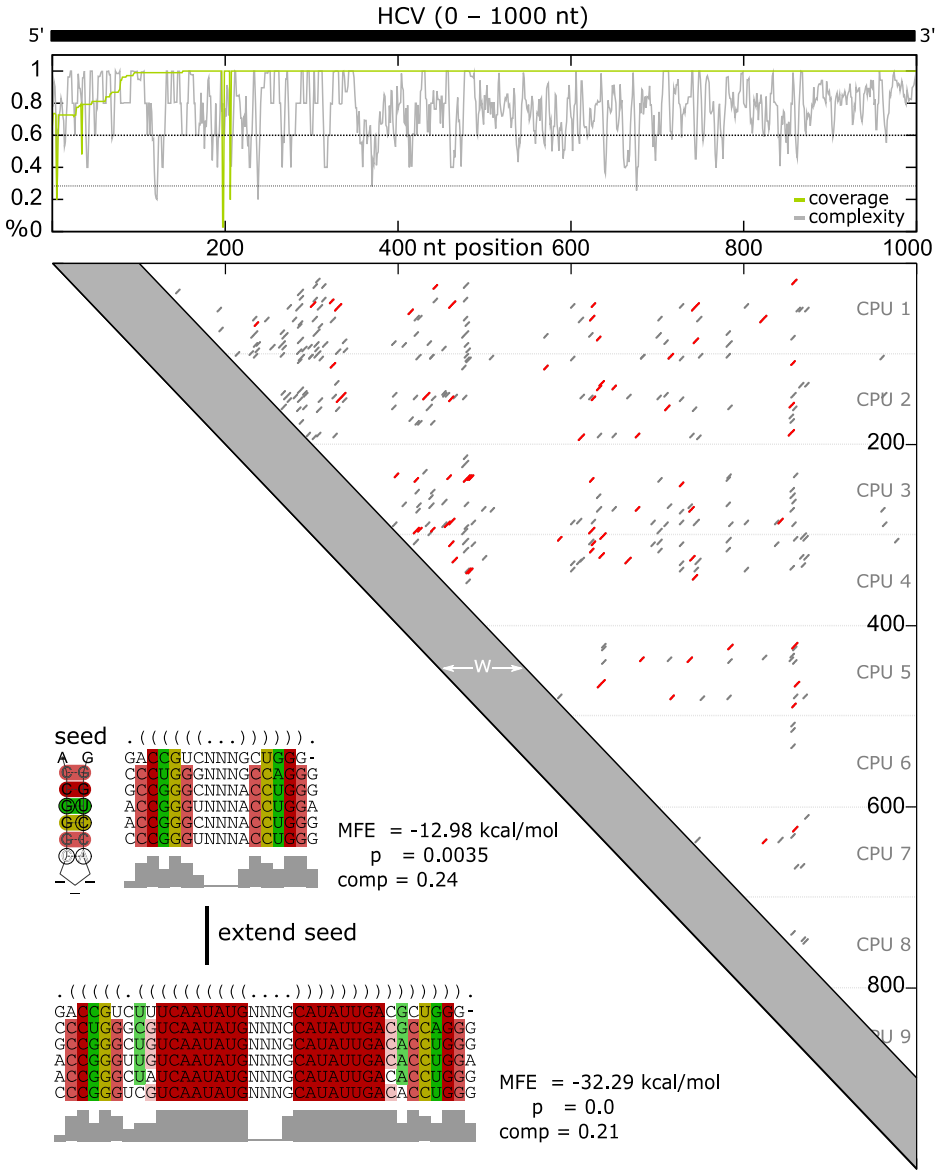


types of pairs

	1	2	3	4	5	6
incompatible pairs	0	1	2	3	4	5

No.	1	2	3	4	5		
MFE	-13.21	-6.64	-7.47	-11.83	-9.51		
Z-score	-3.64	-0.29	-3.27	-7.43	-0.96		
#Isolates	19	19	100	106	19		
No.	6	7	8	9	10		
MFE	-8.29	-6.73	-11.62	-12.75	-14.40		
Z-score	-2.88	-7.85	-9.38	-2.21	-3.54		
#Isolates	104	106	19	102	101		
No.	11	12	13	14	15	16	17
MFE	-9.53	-5.55	-16.24	-13.31	-18.59	-15.41	-22.04
Z-Score	-5.15	-3.28	-3.89	-3.20	-2.73	-7.22	-9.36
#Isolates	82	84	102	104	104	19	77





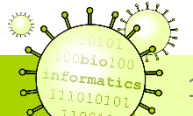
Prediction of conserved long-range RNA-RNA interaction in full viral genomes

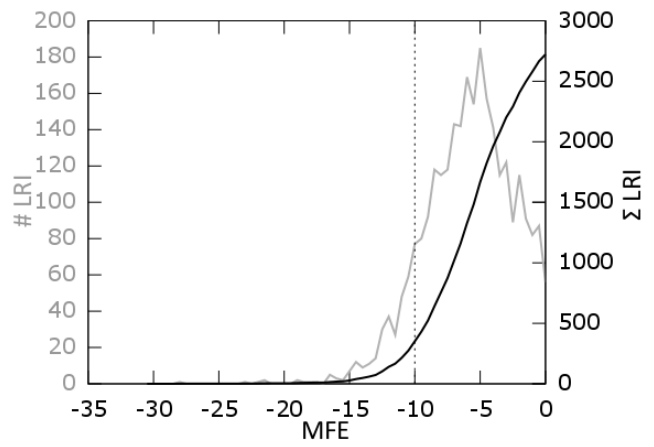
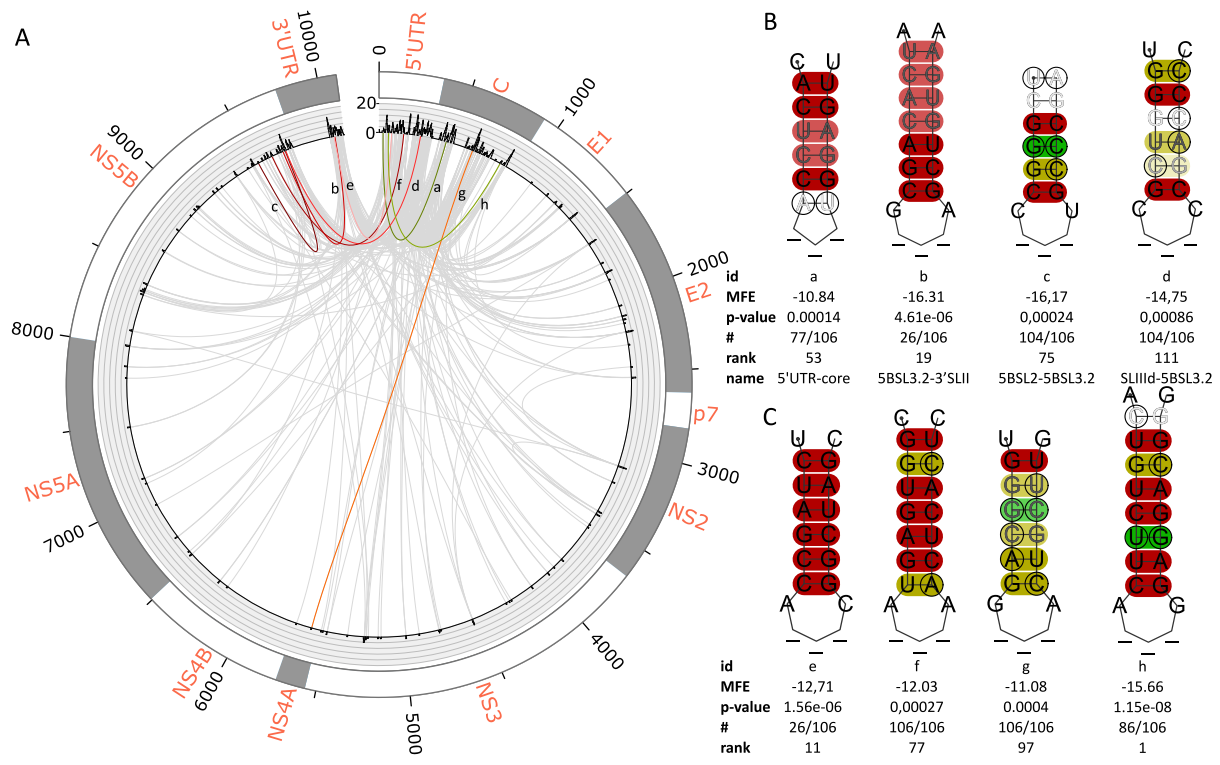
Markus Fricke¹, Manja Marz^{1,2*}



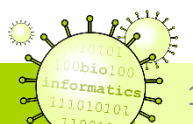
1. Calculate alignment complexity and coverage
2. Find LRI seeds with a sparse dotplot method
3. Filter LRI candidates based on MFE; calculate z-score/p-value and compensatory score.
4. Extend seed interaction

Compression:
AAAAGUUUCC = AGUC





- Tested with alignments of HCV, HIV, Flaviviruses, Tombusviruses
- Found 14 out of 16 evolutionary conserved and experimentally verified LRIs
- Known LRIs are highly ranked



Options

Show seed extension

View

types of pairs

1	2	3	4	5	6
0	1	2	3	4	5
1	2	3	4	5	6
2	3	4	5	6	7

LRi: 701-706 to 8838-8843

LRI Table

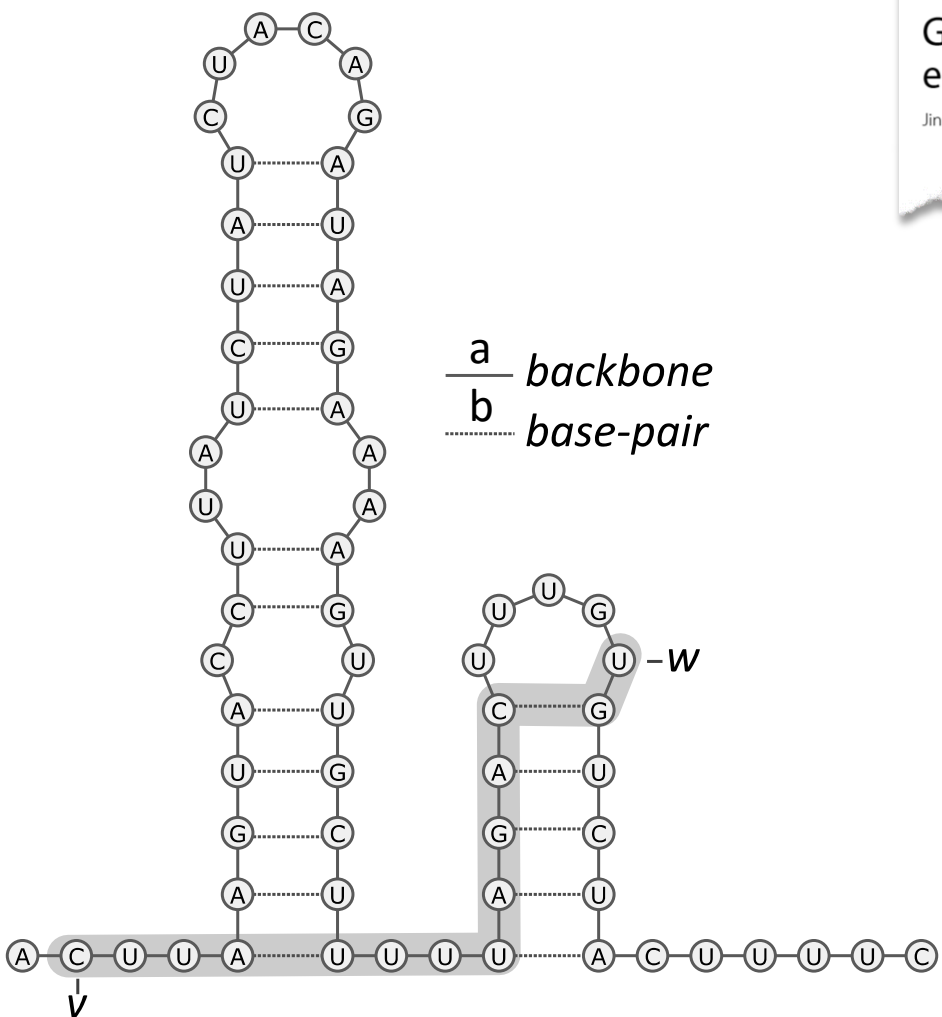
Alignment

[download](#)

id	aln	position				seed values (alignment based)							extended seed values (alignment based)			
		Ai	Aj	Bi	Bj	length	distance	complex	co	mfe	comp	p-value	length_ex	mfe_ex	comp_ex	p-value_ex
2	show	701	706	8838	8843	6	8132	0.8	6	-11.98	0.24	0	12	-17.53	0.22	0.09092548502410791
105	show	2322	2326	7173	7177	5	4847	0.7	6	-10.13	0.23	0	6	-11.59	0.23	0.20453541501857275
112	show	5550	5557	5927	5934	8	370	0.64	6	-12.96	0.22	0	15	-14.15	0.19	0.9220733261019928
88	show	9215	9221	10439	10445	7	1218	0.87	6	-10.95	0.2	2.7502418120661787e-09	9	-11.29	0.22	0.46169584550298515
138	show	160	169	11060	11069	10	10891	0.87	6	-19.18	0.18	3.4558015338248538e-09	17	-35.51	0.21	0
97	show	9969	9975	10171	10177	7	196	0.78	6	-15.68	0.25	1.574202200815833e-07	19	-23.77	0.19	0.005950087241421653
27	show	85	92	10990	10997	8	10898	0.75	6	-10.4	0.3	1.9047160548790742e-07	14	-11.29	0.29	0.021589618486253737
67	show	5060	5064	9768	9772	5	4704	0.88	6	-10.71	0.25	3.431755905003442e-07	16	-13.93	0.24	0.03489266243298761
30	show	4814	4818	5357	5361	5	539	0.69	6	-10.24	0.23	3.476131018587125e-07	14	-16.36	0.23	0.10230140532723575



Graph-distance



Qin et al. *Algorithms for Molecular Biology* 2014, 9:19
<http://www.almob.org/content/9/1/19>

AMB ALGORITHMS FOR MOLECULAR BIOLOGY

RESEARCH Open Access

Graph-distance distribution of the Boltzmann ensemble of RNA secondary structures

Jing Qin¹, Markus Fricke³, Manja Marz³, Peter F Stadler^{2,6,7,8,9} and Rolf Backofen^{4,5*}

RNA secondary structure can be transformed in a undirected graph.

$b < 3a$
 closing base pair must be shorter than a hairpin loop

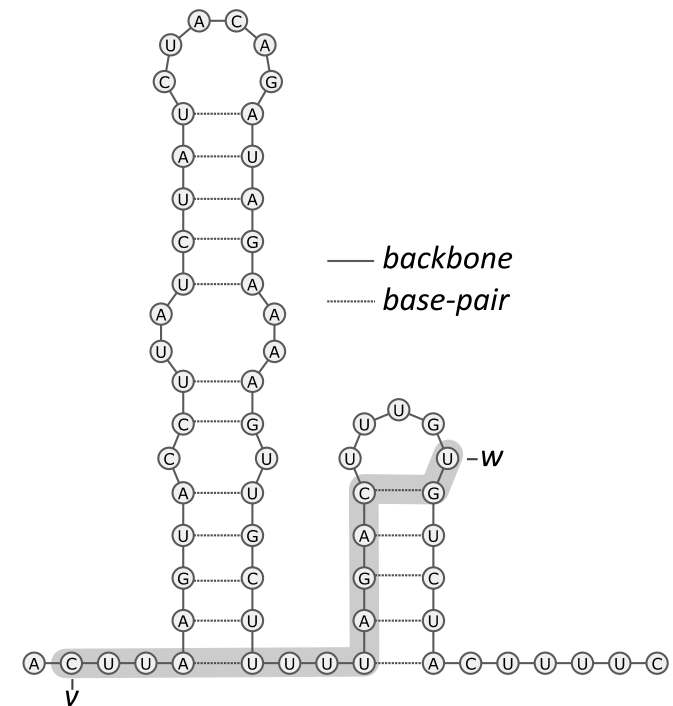
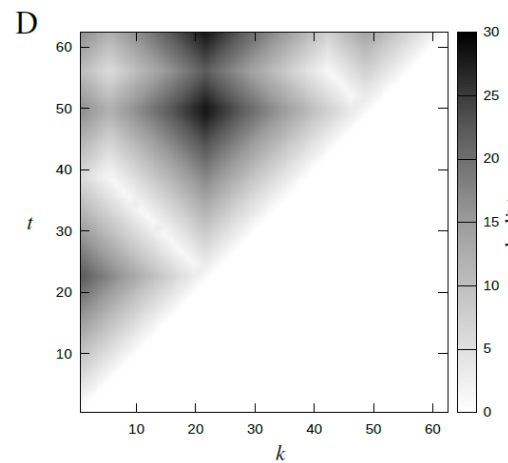
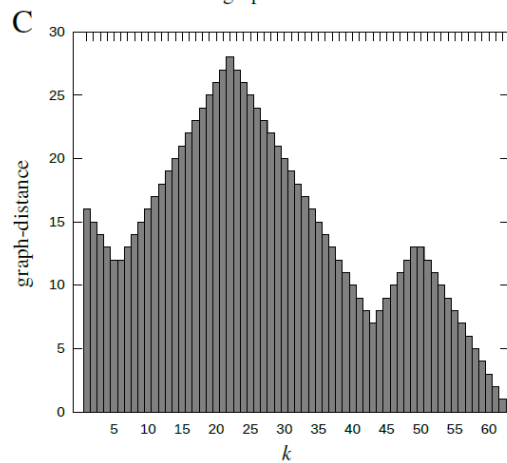
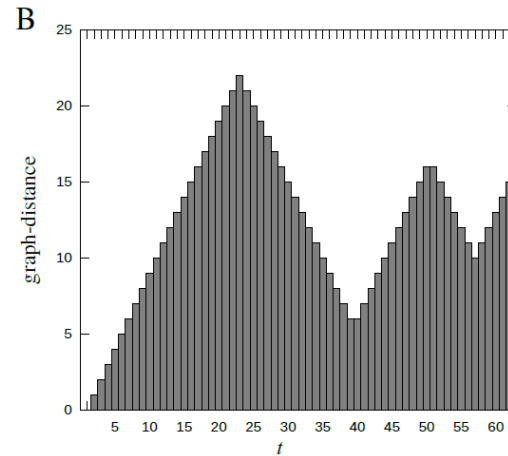
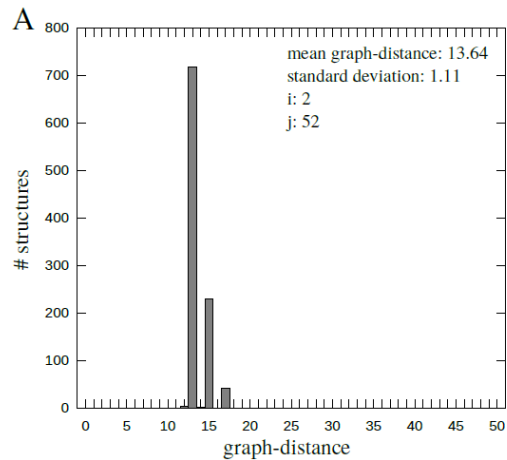
$b < b + 2a$
 closing base pair must be shorter than a stacked pair

→ polynomial-time algorithms

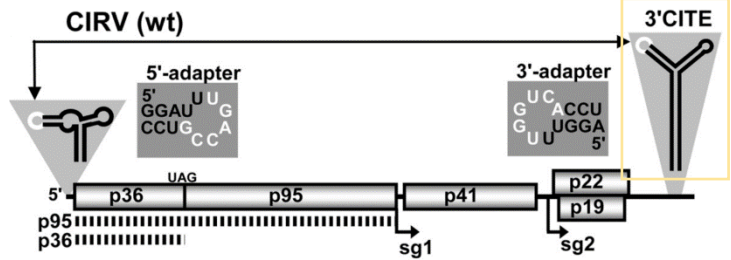
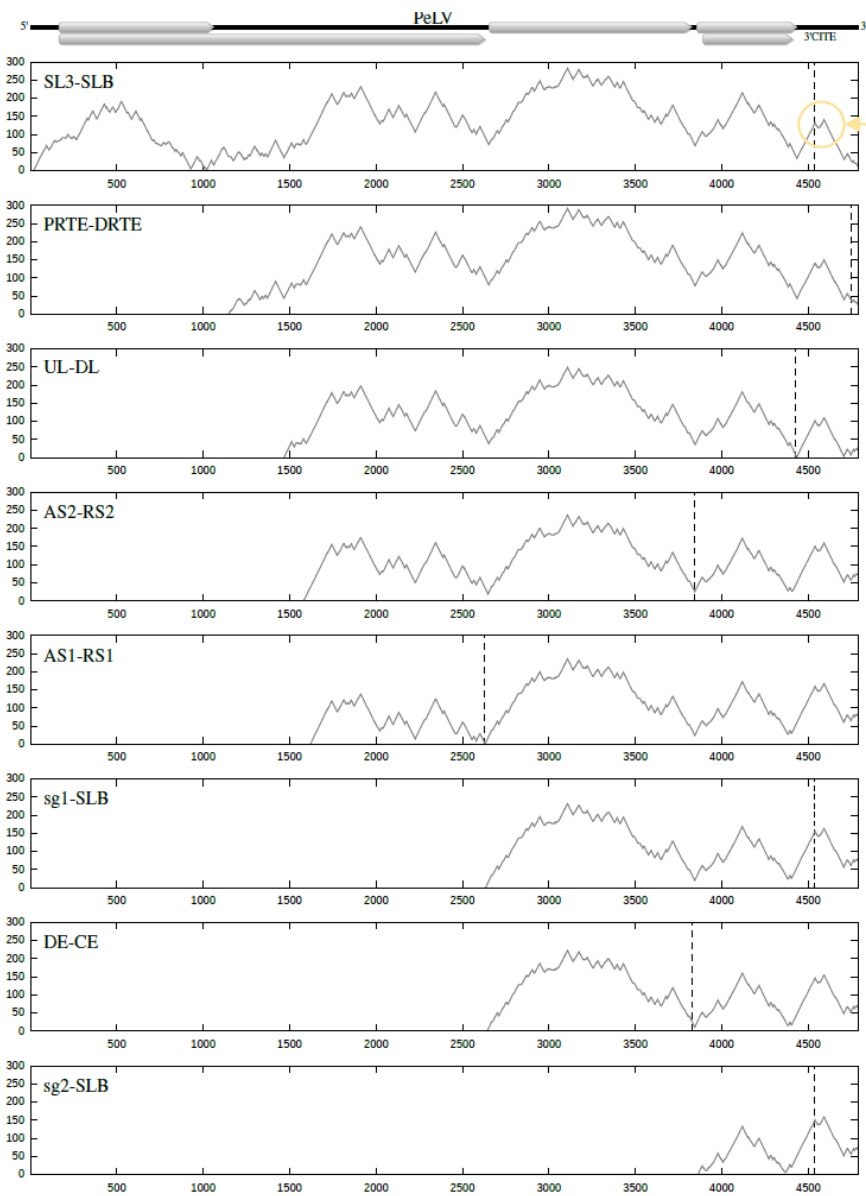


Dijkstra's algorithm in combination with a Fibonacci heap in $O(n \log n)$ time

Input: sample of Boltzmann-weighted secondary structures created with **RNAsubopt**

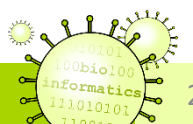


LRI in Tombusviruses

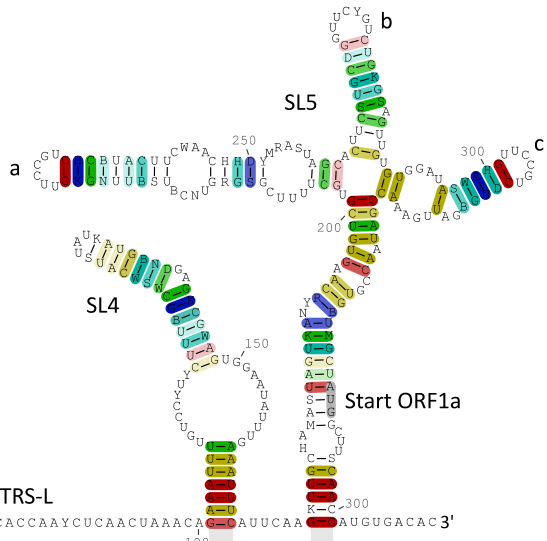
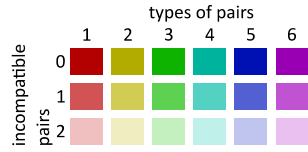


Nicholson 2013

- Tombusvirus eight LRIs are known: SL3-SLB, PRTE-DRTE, UL-DL, sg1-SLB, sg2-SLB, AS1-RS1, AS2-RS2, DE-CE
- graph-distance for the complete genome of Pear latent virus
- starting at the first interaction site of each LRI
- local minimum at second interaction site



Conserved RNA secondary in CoV – 5' UTR



Virus Research 194 (2014) 76–89

Contents lists available at ScienceDirect

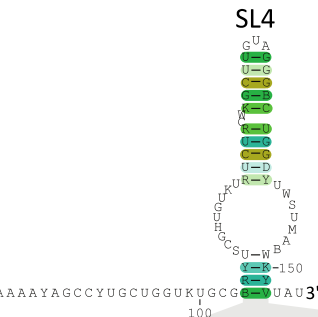
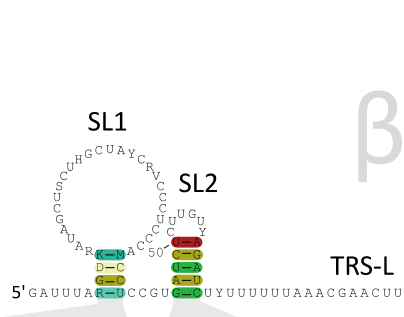
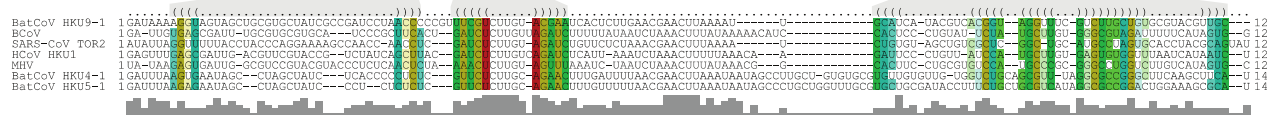
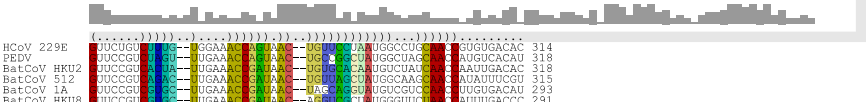
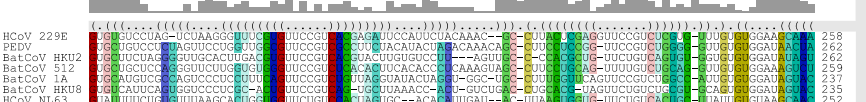
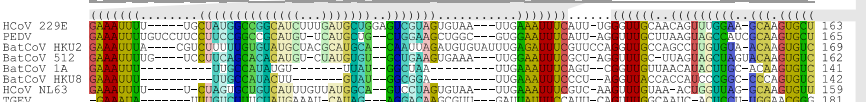
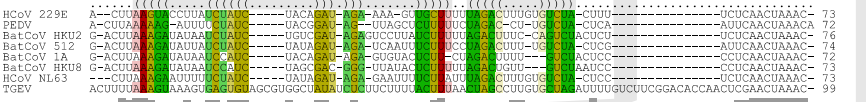
Virus Research

journal homepage: www.elsevier.com/locate/virusres

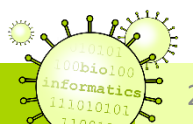
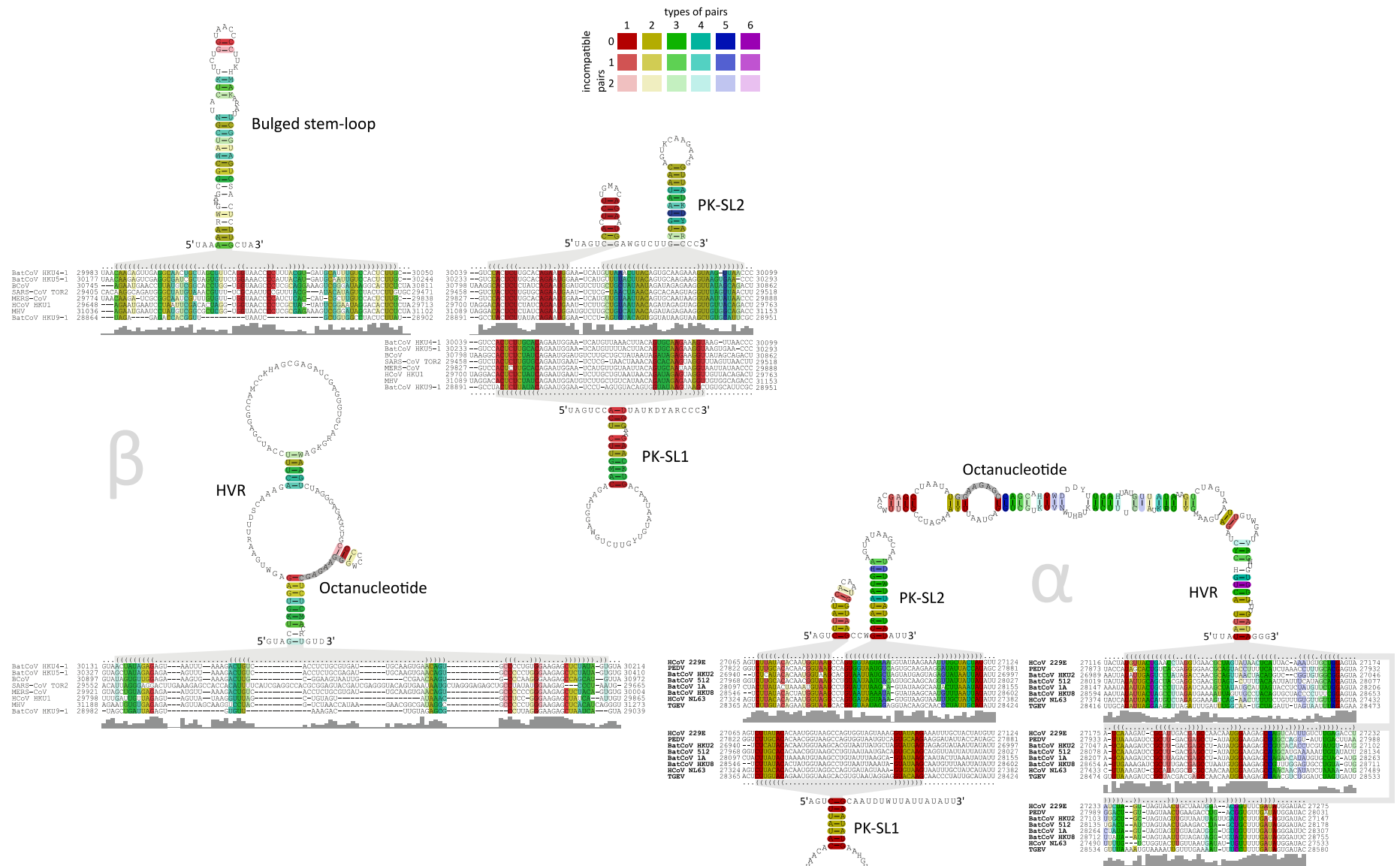
RNA structure analysis of alphacoronavirus terminal genome regions

Ramakanth Madhugiri^a, Markus Fricke^b, Manja Marz^b, John Ziebuhr^{a,*}



Conserved RNA secondary in CoV – 3' UTR



- [1] Rolf Backofen, **Markus Fricke**, Manja Marz, Jing Qin, and Peter F Stadler. “Distribution of graph-distances in Boltzmann ensembles of RNA secondary structures”. In: *Algorithms in Bioinformatics*. Springer, 2013, pp. 112–125.
- [2] Manja Marz, Niko Beerenwinkel, Christian Drosten, **Markus Fricke**, Dmitriy Frishman, Ivo L. Hofacker, Dieter Hoffmann, Martin Middendorf, Thomas Rattei, Peter F. Stadler, and Armin Töpfer. “Challenges in RNA virus bioinformatics”. In: *Bioinformatics* 30.13 (July 2014), pp. 1793–1799. DOI: [10.1093/bioinformatics/btu105](https://doi.org/10.1093/bioinformatics/btu105).
- [3] Jing Qin, **Markus Fricke**, Manja Marz, Peter F. Stadler, and Rolf Backofen. “Graph-distance distribution of the Boltzmann ensemble of RNA secondary structures”. In: *Algorithms Mol Biol* 9 (2014), pp. 19–19. DOI: [10.1186/1748-7188-9-19](https://doi.org/10.1186/1748-7188-9-19).
- [4] Ramakanth Madhugiri, **Markus Fricke**, Manja Marz, and John Ziebuhr. “RNA structure analysis of alphacoronavirus terminal genome regions”. In: *Virus Res* 194 (Dec. 2014), pp. 76–89. DOI: [10.1016/j.virusres.2014.10.001](https://doi.org/10.1016/j.virusres.2014.10.001).
- [5] **Markus Fricke**, Nadia Dünnes, Margarita Zayas, Ralf Bartenschlager, Michael Niepmann, and Manja Marz. “Conserved RNA secondary structures and long-range interactions in hepatitis C viruses.” In: *RNA* 21.7 (July 2015), pp. 1219–1232. DOI: [10.1261/rna.049338.114](https://doi.org/10.1261/rna.049338.114).
- [6] **Markus Fricke** and Manja Marz. “Prediction of conserved long-range RNA-RNA interaction in full viral genomes”. In: *Bioinformatics, submitted xx.x* (Feb. 2016), p. xxx.
- [7] **Markus Fricke**, Florian Zirkel, Christian Drosten, Junglen Sandra, and Manja Marz. “Full length de novo genome assembly of unknown RNA viruses”. In: *Bioinformatics, in preparation xx.x* (Feb. 2016), p. xxx.



Thanks for your attention!

The Marz group:

Manja Marz
Martin Hölzer
Emanuel Barth
Konstantin Riege
Franziska Hufsky
Bertram Vogel
Nelly F. Mostajo Berrospi
Michael Gaspar
Akash Srivastava
Holger Bierhoff
Konrad Sachse
Kathrin Schowtka

Cooperation partner:

Michael Niepmann
Ramakanth Madhugiri
John Ziebuhr
Nadia Dünnes
Margarita Zayas
Rolf Backofen
Jing Qin
Peter F. Stadler
Ralf Bartenschlager
Florian Zirkel
Christian Drost
Sandra Junglen