Characterization of regulatory Flavivirus RNA structure elements

Roman Ochsenreiter[1], Andrea Tanzer[1], Ivo L. Hofacker[1,2], Michael T. Wolfinger[1,3*]

[1]Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, 1090 Wien, Austria
[2]Bioinformatics and Computational Biology Research Group, University of Vienna, Währingerstraße 17, 1090 Wien, Austria
[3]Center for Anatomy and Cell Biology, Medical Unversity of Vienna, Währingerstraße 13, 1090 Wien, Austria

Contact: roman.ochsenreiter@univie.ac.at - http://www.tbi.univie.ac.at

1. Motivation

Flaviviruses (FVs) are small, single stranded positive-sense RNA viruses of 10-12kb length with **highly structured untranslated regions** (UTRs) [1]. The FV genus includes human pathogens such as Yellow fever virus (YFV), West nile virus (WNV), Japanese encephalitis virus (JEV), Dengue virus (DENV) and the recently emerging Zika virus (ZIKV).
The heavily structured UTRs are **crucial for regulation of the viral life cycle**, inducing processes such as genome circularization, viral replication, packaging, and triggering pathogenicity [2]. Here, we present a study for computational identification and characterization of **conserved structural elements** in the UTRs of mosquito-borne flaviviruses (MBFV) using **covariance models** (CMs).

2. sfRNA

Upon FV infection, accumulation of stable long non-coding viral RNAs, termed **subgenomic flaviviral RNAs (sfRNAs)** is observed. Intact sfRNAs are essential for FV survival and pathogenesis [3].

sfRNAs are produced by the host exoribonuclease Xrn1, which degrades the viral genome in 5' to 3' direction. During degradation, Xrn1 is effectively stopped at highly stable structure elements in the 3'UTR, termed **xrRNA (Xrn1-resistant RNA elements)**.

FVs typically have several xrRNA elements, each of which possessing different capacity to stop Xrn1, and thus giving rise to sfRNAs of different lengths. Most important are the so called **'stem-loop' (SL)** and **'dumb-bell' (DB) elements** [4].

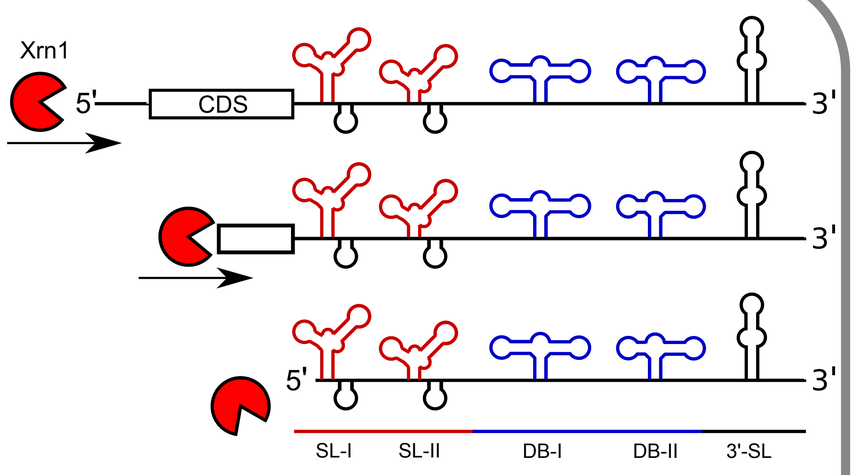
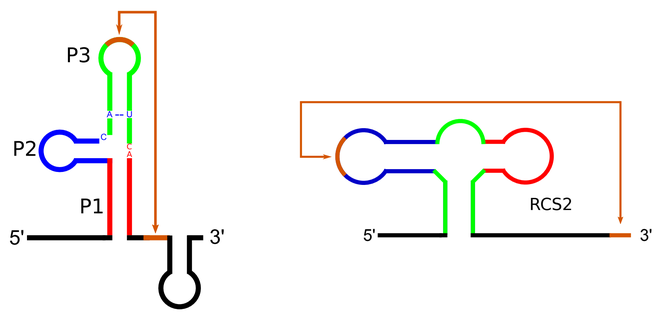


Fig. 1 Architecture of MBFV 3'UTRs. RNA elements attributed to halting Xrn1 are shown in red and blue. SL1 and SL2 correspond to previously described SL-II and SL-IV in West nile virus (WNV).



3. SL/DB RNA families

The **Rfam Database** currently contains CMs for both SL-II and DB elements. Initial screens of UTRs of all FV species could not reliably annotate elements in most species, due to the models' **specicifity for WNV and JEV**.

We manually built seed alignments for all SL and DB elements in DENV, YFV and ZIKV. Next, we iteratively refined the models by scanning all FV UTRs with our CMs using cmsearch [6] and manually aligned strong hits to our seed alignment. This procedure was repeated until no more new significant hits were obtained.

The resulting SL and DB seed alignments reveal **strong structural heterogeneity**, not only among elements, but also within individual virus families. While SL/DB elements share common functionality, they cannot be properly aligned. We therefore suggest to consider them as **RNA clans**, rather than RNA families.

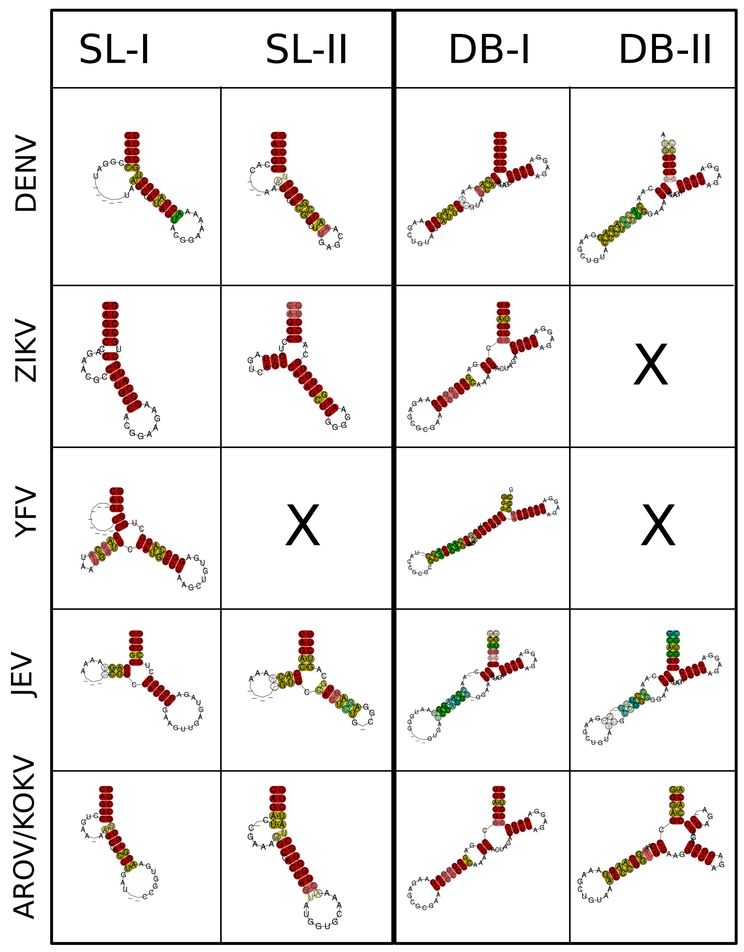


Fig. 2 Consensus RNA structures of each RNA family of the SL-I/II and DB-I/II clans. An 'X' indicates that a species has no element of this type. Consensus structure prediction has been computed with RNAalifold [5].

4. Filtering CMs

We use a **Self-Organizing-Map** (SOM) for clustering and visualization of all putative xrRNAs for **quality assessment of CMs** in our clans. Redundant or unspecific CMs can be quckly identified by frequent co-localization of descriptors, indicating low model specicifity.

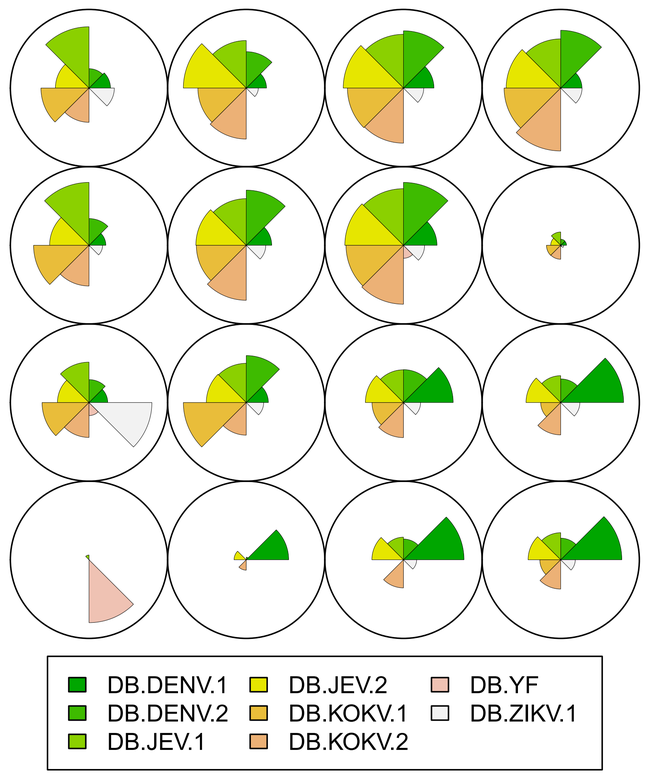


Fig. 3 Self-Organizing-Map. Each circle represents a cluster of similar DB elements. The segment sizes reflect the strength of a particular descriptor (here: cmsearch E-value of a CM) in this cluster. Frequent co-localization of two or more descriptors indicate redundant CMs (brown and green segments).

5. Results

We screened all available FV 3'UTRs and were able to characterize considerably more xrRNA elements, compared to screens based on existing Rfam families. Moreover, we found evidence for previously unknown elements (DB-II in YFV) or potentially important events in sfRNA evolution, such as rearrangement of ZIKV SL elements compared to DENV.

		Detection Ratio					
Virus	Sequences	SL-II	SL-IV	DB-I	DB-II	SL-Rfam	DB-Rfam
DENV-1	1613	1	0.99	1	0.99	0	0.99
DENV-2	1100	1	1	1	0.98	0	0.99
DENV-3	862	0.99	0.95	1	0.91	0	0.99
DENV-4	153	0.99	0	1	1	0	0.99
JEV	239	1	1	1	1	0.91	0.89
WNV	993	1	0.76	1	0.94	1	0.94
ZIKV	83	1	1	1	X	0	0.73
YFV	57	1	0.19	1	0.12	0	0.95
TMUV	35	1	1	1	1	1	0
USUV	90	1	1	1	1	0.99	0

Table 1. Detection ratios for SL/DB elements in various FVs. An 'X' indicates that a particular species has no element of this type.

6. Acknowledgements

This work was partly funded by the Austrian Science Fund FWF projects "RNA regulation of the transcriptome" (F43), "mRNAs von Viren: Evolution und Struktur-Funktionsbeziehungen" (FWF-I-1303) and the Austrian/French project "RNA-Lands" (FWF-I-1804-N28 and ANR-14-CE34-0011).

[1] Villordo, S.M., Carballeda, J.M.,Filomatori, C.V., Garmarnik, A.V. (2016), **RNA Structure Duplications and Flavivirus Host Adaptation,** Trends Microbiol
[2] Pijlman, G. P. et al. (2008), **A Highly Structured, Nucleas-Resistant, Noncoding RNA Produced by Flaviviruses Is Required for Pathogenicity,** Cell
[3] Chapman, E. et al. (2014), **RNA structures that resist degradation by Xrn1 produce a pathogenic Dengue virus RNA,** eLife
[4] Rauscher S. et al. (1997), **Secondary structure of the 3'-noncoding region of flavivirus genomes: comparative analysis of base pairing probabilities,** RNA
[5] Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., Hofacker, I.L. (2011), **ViennaRNA Package 2.0,** Algorithms for Molecular Biology
[6] Nawrocki, E.P., Eddy S.R., **Infernal 1.1: 100-fold faster RNA homology searches (2013),** Bioinformatics