Evolution of Flavivirus regulatory RNA elements

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

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

Contact: michael.wolfinger@univie.ac.at - http://www.tbi.univie.ac.at/~mtw

Background and Motivation

Arthropod-borne flaviviruses (FVs), including pathogens such as Dengue (DENV), Zika (ZIKV), Yellow fever (YFV), Japanese encephalitis virus (JEV) are a growing global health treat. FV are small (+)ssRNA viruses of 10-12kb length with **highly structured untranslated regions** (UTRs). The latter are associated with regulation of the viral life cycle, inducing genome circularization, replication, packaging, and modulating pathogenicity [1]. We present a computational approach for **automatic annotation of conserved RNA structural elements** in FV UTRs based on covariance models (CMs).

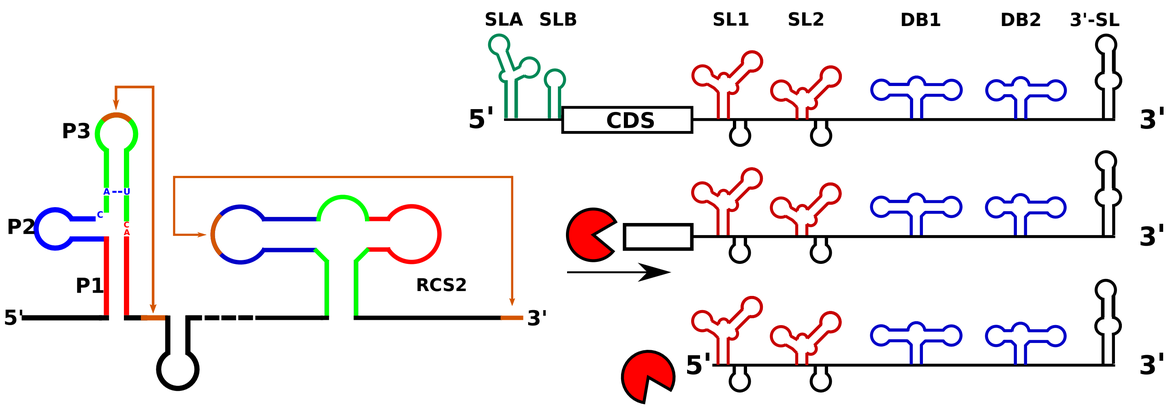


Fig 1. Schematic representation of FV genome organization (top right). Conserved xrRNA elements SL and DB (left) are located in single or tandem within 3'UTRs and efficiently stall host exonuclease Xrn1 (red pac-man). A pseudoknot interaction has been reported for some SL and DB elements (orange, left).

Upon FV infection, accumulation of stable long non-coding viral RNAs, termed **subgenomic flaviviral RNAs (sfRNAs)** is observed. sfRNAs modulate cellular function and are linked to pathogenicity. They are produced by stalling the 5'-3' host exoribonuclease Xrn1 at stable structural elements in the 3'UTR, termed Xrn1-resistant RNA (xrRNA). Mosquito-borne FV (MBFV) typically have more than one xrRNA element, each having different capacity of stalling Xrn1, thus enabling production of sfRNAs of different lengths. **Stem-loop (SL)** and **Dumbbell (DB)** elements have been attributed xrRNA functionality in MBFV.

Covariance models

Structured non-coding RNAs are evolutionary conserved. Their function depends more on their secondary or tertiary structure than on their primary sequence. Finding homologs of a set of structurally related RNAs can be achieved with **covariance models (CMs)** [2], i.e. statistical models of RNA structural alignments based on profile stochastic context free grammars (SCFG). Contrary to Hidden Markov models, paired positions in CMs depend on each other, thus allowing the profile to model covariation in base pairs.

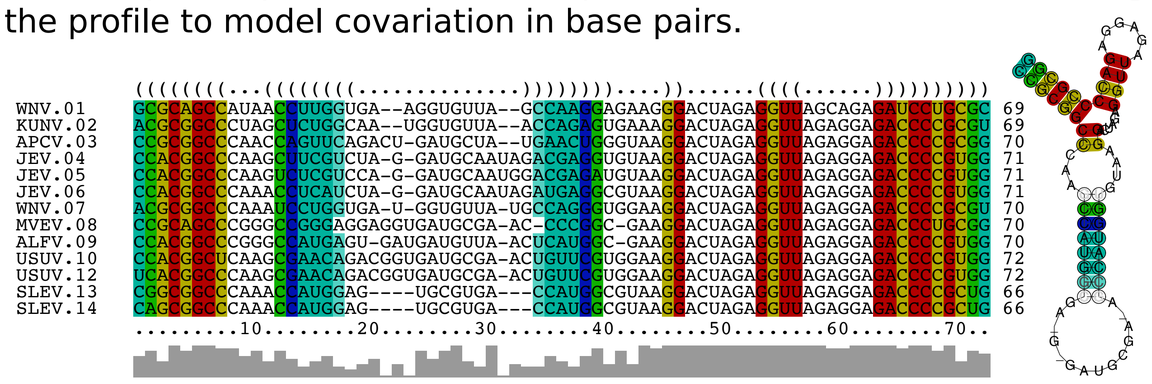


Fig 2. Coraviation in a structural alignment of JEVG viruses (left). Paired positions show a high amount of compensatory mutations. Consensus structure (right) computed with RNAalifold [3].

Covariation is observed both at inter- and intra-species levels in FV. We built **highly specific CMs** for FV 5'UTR (SLA, SLB) and 3'UTR (SL, DB, 3'SL) elements based on **all FV genomes listed in NCBI**. Our data set includes CMs for virus groups, species and serotypes that fit nicely into the concept of Rfam clans [4].

Automated genotyping of viruses

We have developed a software pipeline for automated identification and characterization of evolutionary converved RNA elements in viral genomes. Starting from RNA structural alignments, we create initial CMs and refine them iteratively. False positives are eliminated by a novel approach, **RNAaliSplit** [5], that explicitly considers structural conservation in multiple sequence alignments for classification of non-matching sequences.

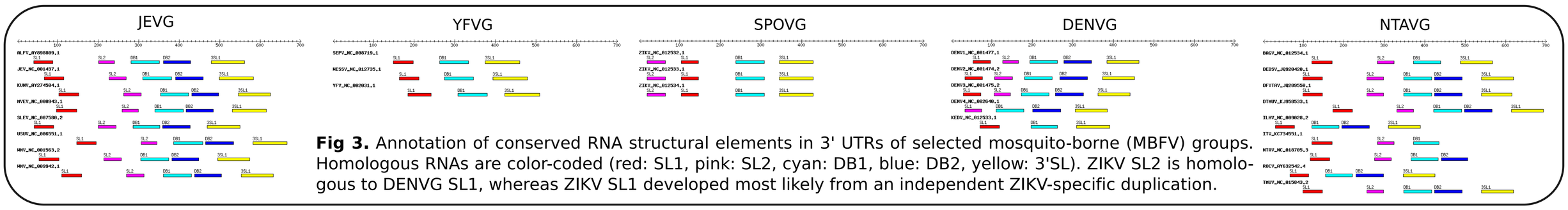
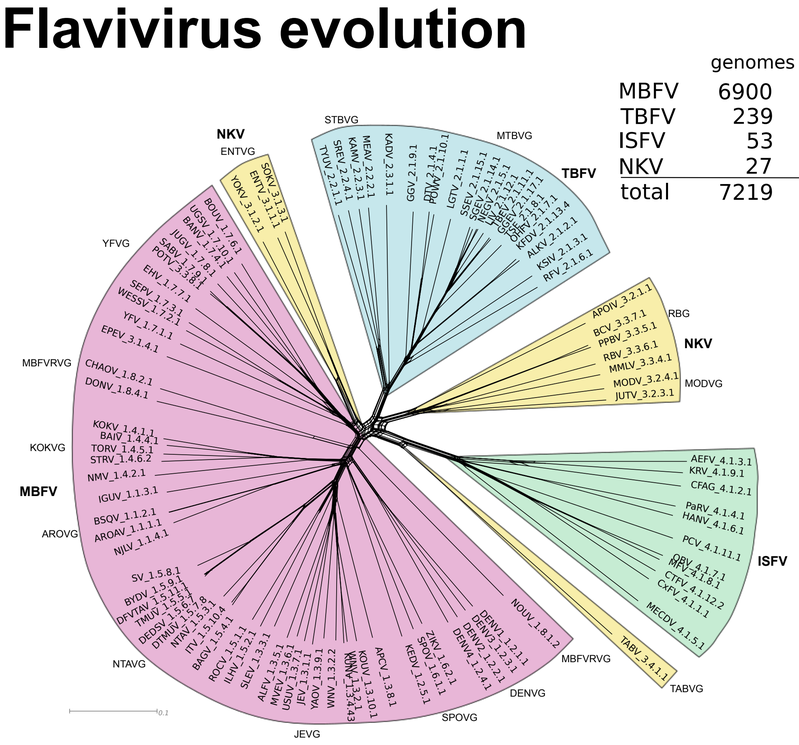


Fig 3. Annotation of conserved RNA structural elements in 3' UTRs of selected mosquito-borne (MBFV) groups. Homologous RNAs are color-coded (red: SL1, pink: SL2, cyan: DB1, blue: DB2, yellow: 3'SL). ZIKV SL2 is homologous to DENVG SL1, whereas ZIKV SL1 developed most likely from an independent ZIKV-specific duplication.

Flavivirus evolution



	genomes	hits
MBFV	6900	6391
TBFV	239	170
ISFV	53	3
NKV	27	12
total	7219	6576

Fig 4. Phylogenetic network of 95 different FV species. Neighbor net computed from ClustalO multiple sequence alignment of annotated protein coding regions.

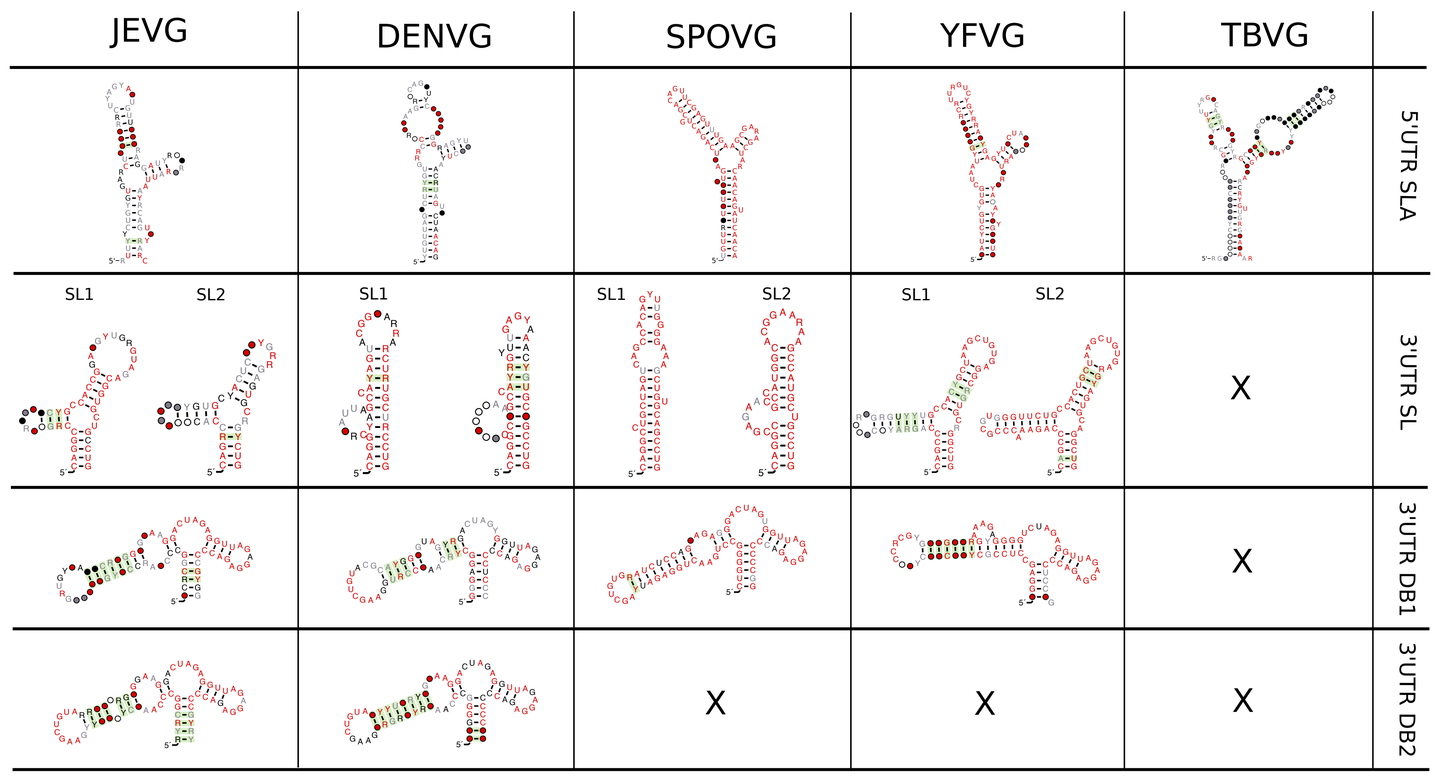


Fig 5. Regulatory elements in MBFV. Tick-borne FV have a 5'UTR SLA, but lack SL and DB elements. Consensus structures computed with RNAalifold [3]. Base pairs with significant covariation according to R-scape [6] are shown in green.

Acknowledgements: This work was partly funded by the Austrian Science Fund FWF projects F43, FWF-I-1303 and FWF-I-1804-N28.

[1] Villordo, S.M., Carballeda, J.M.,Filomatori, C.V., Garmarnik, A.V. (2016), **RNA Structure Duplications and Flavivirus Host Adaptation**. *Trends Microbiol 24(4), 270-283*.
[2] Nawrocki, E.P., Eddy S.R. (2013), **Infernal 1.1: 100-fold faster RNA homology searches.** *Bioinformatics, 29:2933-2935.*
[3] 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* 6:1, 26
[4] Nawrocki, E.P., et al. (2015), **Rfam 12.0: updates to the RNA families database**, *Nucleic Acids Reserach* 43 (D1): D130-D137.
[5] RNAaliSplit: http://github/com/mtw/Bio-RNA-RNAaliSplit
[6] Rivas, E. Clements, J. Eddy, S.R. (2016), A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. Nature Methods 14, 45-48