Functional RNA Structures in the 3'UTR of Mosquito-Borne Flaviviruses

Michael T. Wolfinger[1,2], Roman Ochsenreiter [1] and Ivo L. Hofacker[1,2]

January 14, 2021

[1] Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, 1090 Vienna, Austria;
[2] Research Group BCB, Faculty of Computer Science, University of Vienna, Währingerstraße 29, 1090 Vienna, Austria;

Correspondence: michael.wolfinger@univie.ac.at

Abstract

Recent experimental evidence revealed a thorough understanding of the involvement of functional RNA elements in the 3' untranslated regions (UTRs) of flaviviruses with virus tropism. Comparative genomics and thermodynamic modelling allow for the prediction and functional characterization of homologous structures in phylogenetically related viruses. We provide here a comprehensive overview of evolutionarily conserved RNAs in the 3'UTRs of mosquito-borne flaviviruses.

1 Introduction

Flaviviruses (genus *Flavivirus*) are enveloped, non-segmented single-stranded (+)-sense RNA viruses of approximately 50nm diameter with icosahedral or spherical geometries. They belong to the *Flaviviridae* family, which also includes the genera *Hepacivirus*, *Pegivirus* and *Pestivirus*. Flaviviruses are distributed world-wide, with certain species found only in endemic or epidemic areas, and can infect a wide variety of vertebrate and invertebrate species, rendering them a global economic burden and major health treat. Most flaviviruses are zoonotic arthropod-borne viruses (arboviruses) with ancestral transmission cycles in wildlife. Humans are typically infected via direct spillover (enzootic cycle), amplification in domesticated animals followed by spillover to humans (epizootic cycle), or a human-mosquito-human cycle, often in urban environments (urban epidemic cycle) [1]. While these transmission cycles involve hematophagous arthropod vectors, predominantly of the *Aedes* and *Culex* genera, the transmission mechanism has not been elucidated for all flaviviruses. This is reflected in the classification into groups that represent ecological host/vector associations (Figure 1): Mosquito-borne flaviviruses (MBFVs) and tick-borne flaviviruses (TBFVs) spread between invertebrate vectors (mosquitoes and ticks) and vertebrate hosts (mainly mammals and birds) and comprise the majority of all known flaviviruses. They are also referred to as vertebrate-infecting flaviviruses (VIFs) in the literature. Insect-specific flaviviruses (ISFs) only replicate in mosquitoes and split into two groups, classic insect-specific flaviviruses (cISFs) and dual-host affiliated insect-specific flaviviruses (dISFs) [2]. The first group exclusively infects mosquitoes and forms a monophyletic group within the flavivirus phylogenetic tree, while the second group comprises viruses in separate clades that are more closely related to mosquito/vertebrate flaviviruses without the capability to infect vertebrate cells. Contrary, no-known-vector flaviviruses (NKVs) comprise a phylogenetically

and ecologically diverse set of viruses that have only been found in rodents and bats, without causing disease or high viremia in these animals [3], although human infection and disease has been reported [4]. Importantly, no arthropod vector has been identified so far for these viruses. The recent discovery of novel crustacean flaviviruses that share ancestral roots with terrestrial vector-borne flaviviruses and the characterization of cephalopod flaviviruses that are the most divergent of all flaviviruses currently known highlights the importance of these marine invertebrate flaviviruses (MIFs) in our understanding of flavivirus vertebrate-invertebrate association and evolution [5].

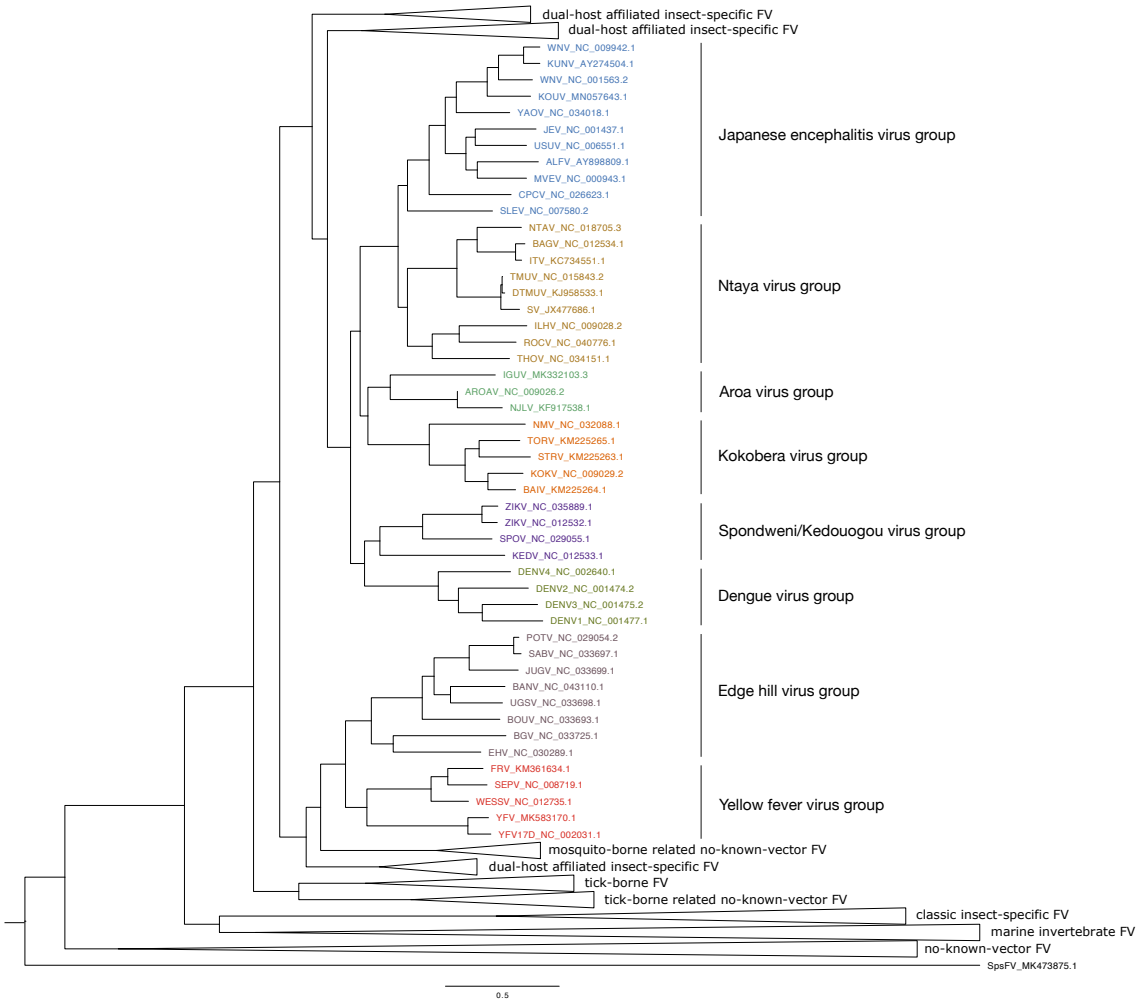

Figure 1: Maximum likelihood tree of the genus *Flavivirus*, computed from a `MAFFT` multiple sequence alignment of complete polyprotein nucleotide sequences with `iqtree`. Serologically distinguishable groups (most of them are defined by the International Committee on Taxonomy of Viruses (ICTV)) are in good agreement with sequence-based phylogenetic clustering and highlighted in color. Viral species are abbreviated together with the NCBI accession number that was used for inferring the phylogeny. `refseq` isolates were used whenever possible. For species without `refseq` annotation, the longest complete genome from `genbank` for this species was selected as representative sequence. Non-mosquito borne flavivirus clades have been collapsed to contextualize their positioning in the flavivirus phylogeny. Abbreviations follow the nomenclature discussed in the main text. Southern pygmy squid flavivirus (SpsFV, bottom) is a marine invertebrate flavivirus with the largest phylogenetic distance to all other virus taxa considered here. The tree has been rendered with `FigTree`.

Of the above groups, the vertebrate-infecting flaviviruses certainly represent a major challenge for public health systems, as they comprise re-emerging pathogens like Dengue virus (DENV), West Nile virus (WNV), Japanese encephalitis virus (JEV), and tick-borne encephalitis virus (TBEV) as well as recently emerging viruses like Zika virus (ZIKV) and Usutu virus (USUV). All of these can cause disease in human with different clinical manifestations ranging from asymptomatic to febrile illness, arthralgia, rash and hemorrhagic fever. Importantly, several mosquito-borne flaviviruses are neurotropic and can induce severe pathology of the central nervous system, including meningitis, encephalitis, and fetal microcephaly [6]. On the other side, there are only experimental antiviral drugs to treat flavivirus infections, and the number of available vaccines is limited. The most prominent is the YFV vaccine based on the live attenuated YFV-17D strain developed by Max Theiler and collaborators in 1937. Interestingly, this vaccine is still working after more than 80 years and has been used to prevent yellow fever infections in hundreds of millions of people. Besides that, inactivated JEV and TBEV vaccines are available for use in humans, and an inactivated WNV vaccine is available for use in animals [7].

In addition to the classification by ecological groups, the International Committee on Taxonomy of Viruses (ICTV) proposes a more fine-grained association of flaviviruses into antigenic complexes based on serological evidence. For the mosquito-borne flaviviruses this results in the establishment of eight distinguishable groups [8]: Japanese encephalitis virus group (JEVG), Ntaya virus group (NTAVG), Aroa virus group (AROAVG), Kokobera virus group (KOKVG), Spondweni/Kedougou virus group (SPOVG), Dengue virus group (DENVG), Edge hill virus group (EHVG) and Yellow fever virus group (YFVG). The maximum likelihood tree in Figure 1 illustrates that this classification is in good agreement with sequence-based molecular phylogenetics, which can be explained by the common genome organization shared by all flaviviruses. Flavivirus genomes consist of a single 5'-capped, non-polyadenylated RNA of approximately 10-12 kilobases in length, referred to as gRNA. The gRNA encodes a single open reading frame (ORF) of approximately 3400 codons, which is flanked by highly structured untranslated regions (UTRs) [9]. Upon translation of the ORF, a polyprotein is produced which is processed by viral and cellular enzymes, yielding structural (C, prM, E) and nonstructural (NS1, NS2A, NS2B, NS3, NS4A, 2K, NS4B, NS5) proteins [10]. Conversely, both flavivirus UTRs are crucially involved in the regulation of the viral life cycle, thereby mediating processes such as genome cyclization, viral replication, packaging and immune response [11, 12]. Although the UTRs show very little primary sequence conservation across the different ecological flavivirus groups, RNA structure conservation is eminent [13, 14]. This lines up flaviviruses with other RNA viruses that have evolved functional RNAs [15], such as Hepatitis C virus (HCV), Picorna- and Hepaciviruses with their well-studied 5'UTR internal ribosome-entry (IRES) sites [16, 17], alphaviruses that have predicted structures of unknown function in their 3'UTR [18], or coronaviruses with their cis-acting UTR elements [19, 20, 21, 22].

2 Flavivirus 3'UTR background

Biologically functional RNA elements often rely on a specific fold, which is manifested in a particular, evolutionarily conserved structure. Rather than maintaining the primary sequence, nature implements conservation typically at the level of RNA secondary structures, rendering the task of finding conserved elements essentially a variant of the *in silico* RNA structure prediction problem. Structural conservation is often achieved through compensatory mutations, where the combination of two point mutations results in maintenance of complementarity, as in a mutation AU $\rightarrow$ GC. Homology search approaches build on this evolutionary trait that is also known as covariation. We have recently employed this methodology to propose a varied landscape of RNA structure conservation in the 3'UTR of tick-borne, insect-specific and no-known-vector flaviviruses based on comparative genomics screens [23]. Here we follow up on our

previous investigations and establish a comprehensive map of functional RNAs in the 3'UTRs of mosquito-borne flaviviruses.

A hallmark of flavivirus biology that has been elucidated over the last decade is their ability to actively dysregulate the host mRNA degradation machinery. They achieve this by tolerating the degradation of parts of their genome by endogenous host enzymes, resulting in the accumulation of viral long non-coding RNA (lncRNAs) species in infected cells. These lncRNAs are referred to as subgenomic flaviviral RNAs (sfRNAs) [24] and are stable decay intermediates resulting from incomplete degradation of viral gRNA by host exoribonuclease activity [25]. sfRNAs eventually dysregulate cellular function to evade the host antiviral response and promoting viral infection [26], however, there are major differences in the sfRNA characteristics between different flavivirus groups. Evidence of sfRNA production following incomplete degradation by the 5' $\rightarrow$ 3' exoribonuclease Xrn1, an enzyme associated with the cell's RNA turnover machinery [27], has been observed in mosquito-borne [28, 29], tick-borne [30], no-known-vector [31] and dual-host affiliated insect-specific flaviviruses [32]. Mechanistically, sfRNAs are produced by halting Xrn1 at evolutionary highly conserved exoribonuclease-resistant RNA (xrRNA) elements that exhibit an ion-dependent mechanical anisotropy [33]. Stalling at xrRNAs prohibits further Xrn1 progression towards the viral 3' end and confers quantitative protection of downstream nucleotides [34].

Mosquito-borne flaviviruses have typically more than one xrRNA element, each with different exoribonuclease stalling capacity, thereby enabling the production of sfRNAs of different lengths [35]. Many of the conserved structures associated with xrRNA functionality have been computationally predicted prior to experimental validation [36, 37, 38, 17]. Specifically, the so-called stem-loop (SL) and dumbbell (DB) elements found in the 3'UTR of many flaviviruses have been related to quantitative protection of downstream virus RNA against Xrn1 degradation. Figure 2 shows the the complete Usutu virus 3'UTR (665 nt) with all its predicted RNA secondary structure elements, showcasing RNA elements that confer Xrn1-resistance.

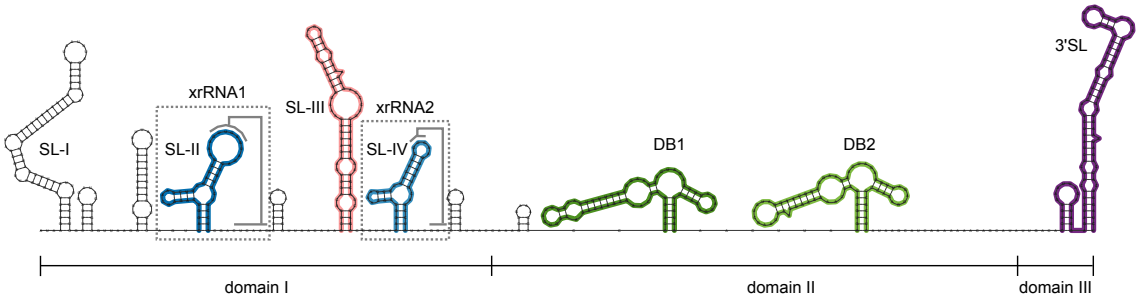

Figure 2: 3'UTR organization of mosquito-borne flaviviruses, exemplarily shown for Usutu virus, a member of the Japanese encephalitis group. The three structurally and functionally independent domains contain duplicated stem-loop (SL) and dumbbell (DB) elements as well as a terminal 3'stem-loop (3'SL) structure. SL elements in domain I are numbered from the beginning of the 3'UTR. Elements that are well conserved across multiple species are highlighted in color. The duplicated SL-II and SL-IV elements fold into a three-way junction structure that can form a transient pseudoknot with sequence regions downstream of the SL elements (depicted in grey), thereby producing exoribonuclease-resistant RNAs xrRNA1 and xrRNA2 (highlighted in boxes).

The 3'UTR of mosquito-borne flavivirus genomes comprises three autonomously folded regions, i.e. domains I-III, that contain one or two copies of conserved RNA elements with distinct functional associations, as shown in Figure 2. Domain I, which is located immediately downstream of the NS5 protein stop codon, manifests as the most variable region, containing a set of simple and branched SL elements. Importantly, due to the intrinsic volatility of this region there is no consistent naming scheme for these elements in the literature. Two of the SL elements in domain I (designated as SL-II and SL-IV here) stand out, as structural homologs can be found in all flaviviruses [23, 39]. They are associated with the formation of a three-way

junction structure that can form a transient pseudoknot by base-pairing of the apical loop with nucleotides located immediately downstream of the stem-loop element. The resulting structure is a tightly packed, mechanically stable xrRNA. Between the two xrRNA-associated SL-II and SL-IV elements, a bulged stem-loop element (SL-III) of approximately 70nt length that is interspersed with interior loops is found in species of the Japanese encephalitis virus group and Ntaya virus group as well as in several species of the dual-host affiliated insect-specific flaviviruses [32]. Although the biological function of this element is unknown, evolutionary conservation across different ecological flavivirus groups is intriguing and could be related to specific vertebrate-invertebrate associations of these viruses. Domain II contains either one or two DB elements, some of which have been reported to also form pseudoknot interactions with downstream regions [40]. These structures bear some exoribonuclease-stalling activity, although to a weaker extent than the upstream SL-derived xrRNAs [34]. An interesting property of the DB elements is the high degree of primary sequence conservation in the distal part of the central multi-loop and the adjacent hairpin loop [41], which can be explained by overlap of the DB elements with cyclisation sequences [42] that mediate the formation of long-range RNA-RNA interactions required for virus replication [43, 44]. Finally, domain III contains a small hairpin (sHP) and a larger bulged SL structure, termed 3'stem-loop (3'SL), which is thermodynamically particularly stable and conserved in all flaviviruses. Host factors acting as chaperones can destabilize the 3'SL element, thereby making cyclisation sequences accessible and mediating the process of pan-handle formation via long-range RNA-RNA interactions between the genomic 5' and 3' ends [45].

3 Materials and Methods

The computational prediction of evolutionarily conserved RNA structures in viruses has become an attractive research topic over the last two decades, particularly with the increasing availability of viral genome and metagenome data. Viral genome data, including sequences and annotation presented here was downloaded from the public National Center for Biotechnology Information (NCBI) `refseq` (`https://www.ncbi.nlm.nih.gov/refseq/`) and `genbank` (`https://www.ncbi.nlm.nih.gov/genbank/`) databases on 26 May 2020. We obtained complete viral genomes under taxonomy ID 11051 (genus *Flavivirus*) and filtered for mosquito-borne flavivirus species listed in Table 1. For phylogeny reconstruction we aimed at selecting `refseq` entries as representatives for each species. In case `refseq` isolates were not available for a particular species, we selected one of the longest complete genomes from the `genbank` set as representative sequence. A maximum-likelihood phylogeny of all species in the genus *Flavivirus* based on a `MAFFT` [46] multiple sequence alignment of complete coding region nucleotide sequences has been inferred with `iq-tree` [47] using the SYM+R10 substitution model (Figure 1).

Detection and characterization of evolutionarily conserved RNAs in MBFV 3'UTRs follows the methodology described recently for other flaviviruses [23]. Our approach is centered around finding structurally homologous RNA elements in phylogenetically narrow subgroups by means of RNA family models. We use `infernal` [48] to transform structural RNA alignments into covariance models [49], i.e. statistical models of RNA structure that extend classic Hidden-Markov-Models (HMMs) to simultaneously represent sequence and secondary structure. Covariance models can then be used to find homologous RNAs in large sequence databases. Building on stochastic context free grammars, a covariance model returns for each hit a bit score, as well as an E-value to assess hit quality.

Here we construct for each MBFV antigenic complex structural multiple sequence alignments of all available 3'UTR sequences from the `refseq` set with `locARNA` [50]. Stringent comparison of the predicted consensus structures with known functional RNAs, such as those listed in Rfam [51], allows for exact localization and re-alignment of conserved elements. In parallel,

Accession number†	Acronym	Scientific name	3'UTR length (nt)	Isolates
AY898809.1	ALFV	Alfuy virus	560	1
NC_009026.2	AROAV	Aroa virus	421	1
NC_012534.1	BAGV	Bagaza virus	566	19
KM225264.1	BAIV	Bainyik virus	371	1
NC_043110.1	BANV	Banzi virus	NA	NA
NC_033725.1	BGV	Bamaga virus	NA	NA
NC_033693.1	BOUV	Bouboui virus	NA	NA
LN849009.1	CPCV	Cacipacore virus	573	1
NC_001477.1	DENV1	Dengue virus 1	462	2210
NC_001474.2	DENV2	Dengue virus 2	451	1631
NC_001475.2	DENV3	Dengue virus 3	440	1028
NC_002640.1	DENV4	Dengue virus 4	384	260
KJ958533.1	DTMUV	Duck Tembusu virus	692	19
NC_030289.1	EHV	Edge Hill virus	NA	NA
KM361634.1	FRV	Fitzroy River virus	472	1
AY632538.4	IGUV	Iguape virus	567	5
NC_009028.2	ILHV	Ilheus virus	388	3
KC734551.1	ITV	Israel turkey meningoencephalomyelitis virus	434	5
NC_001437.1	JEV	Japanese encephalitis virus	582	289
NC_033699.1	JUGV	Jugra virus	NA	NA
NC_012533.1	KEDV	Kedougou virus	390	2
NC_009029.2	KOKV	Kokobera virus	558	2
MN057643.1	KOUV	Koutango virus	551	1
AY274504.1	KUNV	Kunjin virus	624	35
NC_000943.1	MVEV	Murray Valley encephalitis virus	614	18
KF917538.1	NJLV	Naranjal virus	NA	NA
NC_032088.1	NMV	New Mapoon virus	546	2
NC_018705.3	NTAV	Ntaya virus	565	2
NC_029054.2	POTV	Potiskum virus	NA	NA
NC_040776.1	ROCV	Rocio virus	424	3
NC_033697.1	SABV	Saboya virus	NA	NA
NC_008719.1	SEPV	Sepik virus	459	2
NC_007580.2	SLEV	Saint Louis encephalitis virus	549	44
SA-AR‡	SPOV	Spondweni virus	338	1
KM225263.1	STRV	Stratford virus	233	1
JX477686.1	SV	Sitiawan virus	NA	NA
NC_034151.1	THOV	T'Ho virus	556	2
NC_015843.2	TMUV	Tembusu virus	618	74
KM225265.1	TORV	Torres virus	414	1
NC_033698.1	UGSV	Uganda S virus	NA	NA
NC_006551.1	USUV	Usutu virus	665	287
NC_012735.1	WESSV	Wesselsbron virus	478	7
NC_009942.1	WNV	West Nile virus (lineage 1)	631	1920⋆
NC_001563.2	WNV	West Nile virus (lineage 2)	573	1920⋆
NC_034018.1	YAOV	Yaounde virus	NA	NA
MK583170.1	YFV	Yellow fever virus	657	196
NC_002031.1	YFV17D	Yellow fever virus 17D	508	23
NC_035889.1	ZIKV	Zika virus (Asian / American lineage)	429	801⋆
NC_012532.1	ZIKV	Zika virus (African lineage)	428	801⋆

Table 1: Mosquito-borne flavivirus genomes considered in this contribution. The 3'UTR length is listed for each isolate. †Representative accession number from the `refseq` database. Whenever a `refseq` genome was not available, the isolate with the longest 3'UTR from the `genbank` database was selected as representative species. ‡SPOV 3'UTR data taken from [41]. ⋆Total number of available isolates for this virus. NA: 3'UTR partial or not available.

we compute locally stable RNA secondary structures from structurally aligned UTR regions with `RNALalifold` [52] from the `ViennaRNA` package [53]. Covariance models are then built for every conserved RNA structure in each antigenic complex, resulting in a total of 37 covariance models that are subsequently used to perform a comprehensive screen in all `genbank` sequences studied here. As covariance models constructed this way are highly specific for a particular copy of a conserved structural element within an antigenic group, this methodology can discriminate paralogous copies, i.e. SL-II/SL-IV or DB1/DB2, and infer structural proximity among conserved elements in a quantitative manner. We use this approach to unambiguously characterize individual copies as SL-II/SL-IV or DB1/DB2. Structural alignments of all conserved elements presented in this contribution are available at `https://github.com/mtw/viRNA`.

To learn more about the evolutionary traits associated with the presence of single or duplicated functional elements in the 3'UTRs, we perform a quantitative comparison of covariance models based on `CMCompare` [54]. This tool computes for a pair of covariance models a link score that exhibits high values if a sequence exists that scores well in both models simultaneously. Analysis of the link scores computed from pairwise covariance model comparisons yields valuable insight into the association of conserved functional elements. In particular, this strategy allows for a fine-grained classification of elements that are found in varying copy numbers in some viruses and reveals elements that are unique to certain species or antigenic groups.

4 Results

Detailed experimental studies over the last decades have revealed the functional association of flavivirus RNAs, in particular evolutionarily conserved structural elements in the UTRs of these viruses, with pathogenesis (reviewed in [55]), leading to a broad understanding of the biological roles and implications of viral non-coding RNAs (ncRNAs). While many studies focus on examples in individual viruses, thereby elucidating properties of single sequences, a unifying view of the complex flavivirus 3'UTR architecture and the evolutionary association among homologous RNA elements is only beginning to be understood. Following up on our previous investigations in tick-borne, insect-specific and no-known-vector flaviviruses [23], we present here a complete picture of the evolutionary conservation of functional RNAs in the 3'UTR of mosquito-borne flaviviruses. Pursuing a bipartite comparative genomics strategy for elucidating conserved viral RNAs, i.e. building on previously known xrRNA, DB and 3'SL elements in specific isolates, as well as characterizing hitherto understudied elements *de novo* allows us to draw a comprehensive map of structurally homologous RNAs in this ecologic group of viruses. We constrain ourselves to major structural entities encompassing the above mentioned functional RNA elements, purposefully ignoring (small) stem-loop structures that are not conserved consistently in all MBFVs. Figure 3 shows example consensus structure predictions of these elements computed from structural alignments. We use viruses in the Japanese encephalitis virus group as examples here, as this group contains representatives of many structured RNAs found throughout the MBFV 3'UTRs.

Stem-loops II and IV (Figure 3 **b** and **c**) within domain I fold into three-way junction structures that exhibit exoribonuclease-stalling activity upon formation of a transient pesudoknot from their apical loop to downstream nucleotides (depicted in Figure 2). While both elements constitute central xrRNA building blocks, the SL-II apical loop is larger, thus facilitating more pseudoknot interactions among nucleotides than SL-IV. Importantly, structural homologs of SL-II/SL-IV are present in all MBFV species considered here, as well as in many other flaviviruses [23, 39].

Another stable structural element in domain I is stem-loop III (SL-III, Figure 3 **d**), which is conserved in all JEVG and many NTAVG viruses. Surprisingly, SL-III is also conserved in several dual-host affiliated insect-specific flaviviruses, including Nounané virus (NOUV), Barkedji virus (BJV) and Barkedji-like virus (BJLV) [32]. SL-III has a length between 66 nt and 76 nt and, within MBFVs, is always flanked by the xrRNA-associated SL-II and SL-IV elements. Although supported by rich covariation patterns (Figure 3 **c**), the biological function of this thermodynamically stable element remains elusive.

Dumbbell (DB) elements (Figure 3 **f** and **g**) are typically found in one or two copies and represent major building blocks of domain II in many MBFVs, TBFVs and NKVs [23]. These well-conserved structures are characterized by high sequence variability in the proximal (more upstream) stem-loop originating from the central multi-loop, and, conversely, high primary sequence conservation in the distal (more downstream) stem-loop. This pattern is observed in all known DB elements [41] and rooted in the fact that the distal stem-loop overlaps cyclisation sequences required by the viruses to form proper pan-handle structures during replication [42, 56].

The terminal domain III of the 3'UTR is characterized by a high degree of structural conservation throughout all ecological groups of flaviviruses. This region contains a small hairpin loop of approximately 15 nt length, which is immediately followed by a longer 3'stem-loop structure (Figure 3 **d**). We consider the two structural elements together here, designating the combined element as 3'SL.

In the following sections 4.1– 4.7 we provide a qualitative description of RNA structure conservation within MBFV 3'UTRs, without focusing on known functional details. We show for each antigenic complex a plot comprised of a phylogenetic tree computed from complete coding sequence nucleotide alignments (sensu Figure 1) and a schematic of the 3'UTR organization for

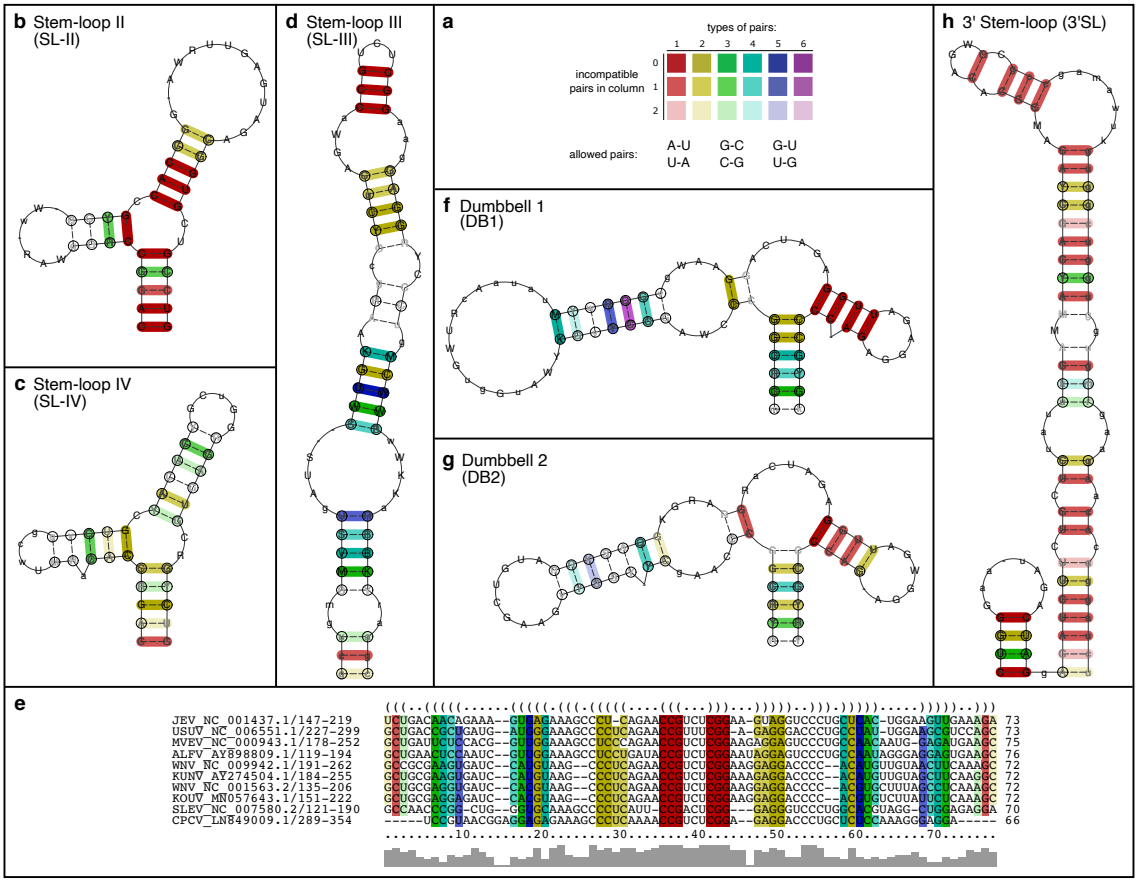

Figure 3: Consensus structures of conserved RNA elements in the 3'UTR of JEVG virus species. (**a**) Base-pair coloring in this figure indicates different levels of covariation observed at the corresponding columns in the underlying alignment, ranging from red (no covariation) over ocher, green and blue to violet (representing all six base pair combinations). (**b**) and (**c**) Stem-loop elements II and IV fold into canonical three-way junction structures and constitute building blocks for xrRNAs. (**d**) Stem-loop III is a highly conserved stem-loop structure interspersed with several interior loops. Biological function of this element is unknown. (**e**) The structural alignment of SL-III, underlying the consensus structure **d** is shown as an example. Gray bars below the alignment indicate for each column the level of primary sequence conservation. (**f**) and (**g**) Dumbbell elements 1 and 2. Sequence-level nucleotide conservation in the distal stem-loop of all species is due to overlap with cyclisation sequences. (**h**) The terminal stem-loop structure (3'SL) is found at the 3'end of all members of the genus *Flavivirus*.

each virus species. Colored regions in these plots highlight evolutionarily conserved elements. Additionally, we provide 3'UTR predictions for some species to facilitate contextualization of elements that show pervasive conservation versus those that are only found in specific groups. Features of 3'UTR regions shown here refer to virus species listed in Table 1. No data are shown for the Edge hill virus group, as our data set does not include any 3'UTR sequences for this clade.

4.1 Japanese encephalitis virus group

We begin the characterization of evolutionarily conserved RNAs in the 3'UTR of different MBFV complexes with the Japanese encephalitis virus group (Figure 4). The JEVG comprises serologically related virus species, encompassing medically important human pathogens, includ-

ing JEV, WNV, USUV, St. Louis encephalitis virus (SLEV) and Murray Valley encephalitis virus (MVEV). The 3'UTR lengths in JEVG range from 549 nt to 665 nt, with SLEV having the shortest, and USUV having the longest 3'UTR, respectively. While the available 3'UTR sequences of Koutango virus (KOUV) and Cacipacore virus (CPCV) are truncated at the 3'end, variability in the JEVG 3'UTRs lengths is largely due to variability in domain I, at the 5'end of the 3'UTR. No 3'UTR sequences are available for Yaoundé virus (YAOV).

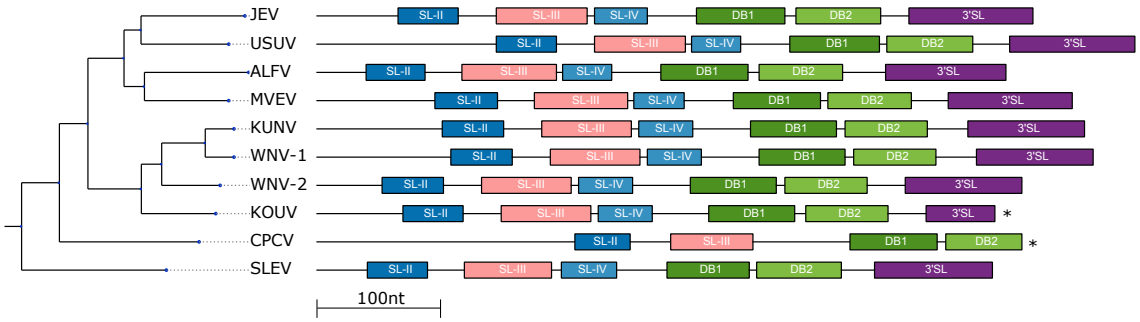

Figure 4: Annotated 3'UTR of the Japanese encephalitis virus group. All species in this group except CPCV show a conserved architectural arrangement of functional RNAs following the pattern SL-II, SL-III, SL-IV, DB1, DB2, 3'SL, with homogeneous spacer region lengths between the functional elements. Variable regions upstream of SL-II harbor structured RNAs in only a few species, e.g. SL-I in USUV (Figure 2), and are not shown here. Truncated sequences are marked by an asterisk ∗.

Comparison of the JEVG 3'UTRs highlights a common architecture of functional RNA elements. All viruses in this group, with the exception of CPCV, contain stem-loops II, III, and IV, followed by duplicated DB elements and a terminal 3'-stem loop (3'SL) element in this order (Figure 4). Domain I in CPCV stands out, as it only contains SL-II and SL-III elements, but no SL-IV structure. Intriguingly, the 3'UTR region upstream of SL-II is unexpectedly long in CPCV, with pairwise distances between the annotated elements in good agreement with other JEVG viruses. Together with the missing SL-IV, this raises the question whether CPCV has evolved different 3'UTR traits. Contrary, domain I in CPCV contains many potential stop codons, which could indicate that the effective NS5 stop codon is located more downstream.

Another interesting observation is the length bias among WNV lineages 1 (WNV-1) and 2 (WNV-2). While WNV-1 and the related Kunjin virus (KUNV) show an almost identical 3'UTR architecture with a length of 631 nt and 624 nt, respectively, domain I of WNV-2 is almost 60nt shorter. 3'UTR architectural variability among different strains or lineages of a particular species is not unusual in RNA viruses and has been observed, e.g., in tick-borne encephalitis virus [57] and chikungunya virus [18].

4.2 Ntaya virus group

The Ntaya virus group comprises human and avian pathogenic viruses that are primarily maintained in transmission cycles between *Culex* spp. mosquitoes and birds. The entire NTAVG forms a sister clade of JEVG in the complete coding sequence nucleotide-based phylogeny (Figure 1), and internally splits into two subclades. One of them contains the type species, Ntaya virus (NTAV), as well as Bagaza virus (BAGV), Israel turkey meningoencephalomyelitis virus (ITV), and Tembusu virus (TMUV) with its subtypes Duck Tembusu virus (DTMUV), Sitiawan virus (SV), and Duck egg drop syndrome virus (DEDSV, not shown here), and the other contains Ilheus virus (ILHV) together with its subtype Rocio virus (ROCV), and T'Ho virus (THOV) (Figure 5). While 3'UTR sequences are available for all of the above viruses except for SV, those of ITV and NTAV are 3' truncated in our data set. 3'UTR lengths range from 388 nt (ILHV) to 692 nt (DTMUV).

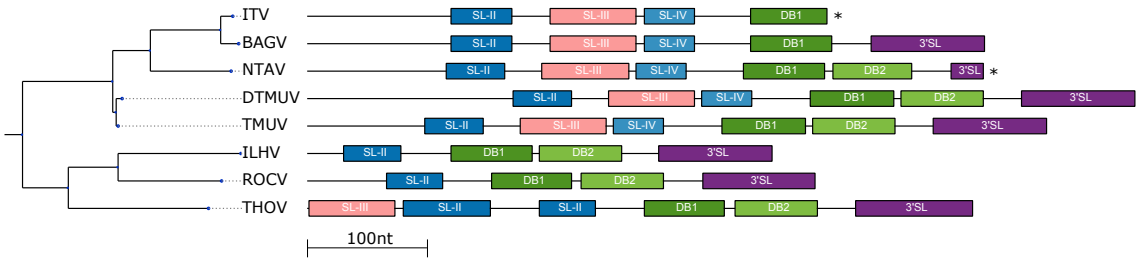

Figure 5: Annotated 3'UTR of viruses in the Ntaya virus group.

The clade that includes NTAV shows a relatively homogeneous architecture of structured elements in domain I of the 3'UTR, encompassing SL-II, SL-III and SL-IV elements. In domain II, BAGV and the truncated ITV form a single DB element, while the other viruses exhibit duplicated DB elements. In the clade comprising THOV, ILHV, and ROCV, no SL-IV elements are found. SL-III is only found in THOV, which moreover exhibits two copies of SL-II, both of which downstream of SL-III. A terminal 3'SL in commonly present in all NTAVG species.

4.3 Aroa virus group

The Aroa virus group contains four recognized species [58], Aroa virus (AROAV), Iguape virus (IGUV), Bussuquara virus (BSQV, not shown here) and Naranjal virus (NJLV). 3'UTR sequences are only available for two members, AROAV and IGUV, with a length of 421 nt and 567 nt, respectively (Figure 6). While the IGUV 3'UTR contains both SL-II and SL-IV elements in domain I, which are separated by approximately 125 nt, AROAV contains only a single SL-II element. Domains II and III contain both DB elements and a canonical 3'SL in both viruses.

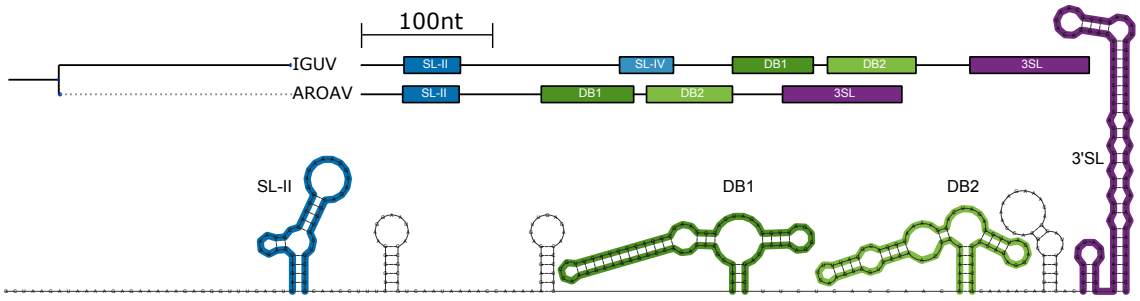

Figure 6: Top: 3'UTR organization in the Aroa virus group. Bottom: Secondary structure prediction of the AROAV 3'UTR. Conserved elements are highlighted in colors that match those in the top panel.

4.4 Kokobera virus group

The Kokobera virus group comprises five recognized species, which are all found in Australia and Papua New Guinea: Kokobera virus (KOKV), Bainyik virus (BAIV), Stratford virus (STRV), Torres virus (TORV) and New Mapoon virus (NMV). Only two of these (KOKV and NMV) have complete 3'UTR sequences in the database, with a length of 558 nt and 546 nt, respectively. The truncated 3'UTR sequences of the other species range from 233 nt (STRV) to 414 nt (TORV).

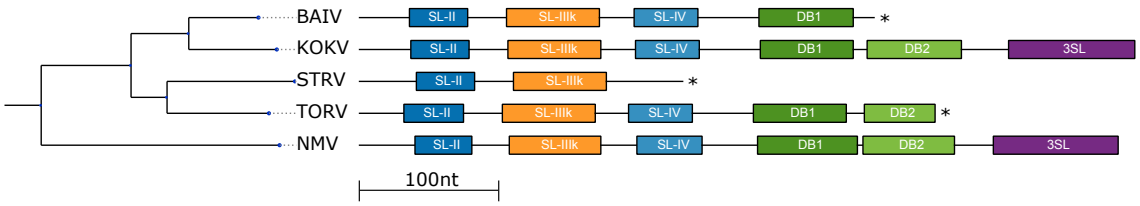

Figure 7: Annotated 3'UTR of viruses in the Kokobera virus group.

Homology search within KOKVG revealed a strictly homogeneous 3'UTR organization (Figure 7), encompassing a unique pattern of structure conservation in domain I: The SL-II and SL-IV elements enclose another stem-loop element (designated SL-IIIk here), that is exclusively present in KOKVG and does not show structural homology to the SL-III elements found in JEVG and NTAVG. The SL-IIIk element is approximately 66 nt long and folds into a bulged stem-loop structure (Figure 8). However, due to high sequence similarity in the 3'UTR of the KOKVG viruses, there is only moderate covariation support for this element (data not shown). Nevertheless, the SL-IIIk element is thermodynamically stable and its presence in all known KOKVG members hints towards unknown biological function. The remaining 3'UTR regions resemble the above discussed MBFV groups, specifically containing two DB elements and a terminal 3'SL structure.

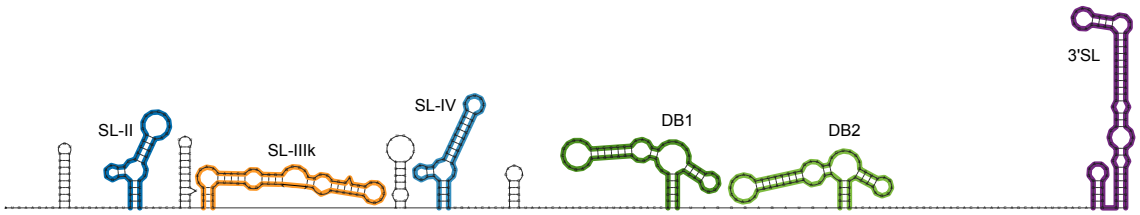

Figure 8: Secondary structure prediction of the KOKV 3'UTR. Evolutionarily conserved elements are highlighted in colors that match those in Figure 7

4.5 Dengue virus group

The Dengue virus group, which comprises four serotypes of Dengue virus (DENV1 – DENV4), is probably the best studied of all Flavivirus complexes, given that the pathology induced by the Dengue viruses represents the most prevalent arthropod-borne disease worldwide. Dengue viruses exhibit a highly homogeneous 3'UTR architecture, with DENV1 being the longest (462 nt) and DENV4 being the shortest (384 nt) sequence. The Dengue 3'UTRs typically contain SL-II and SL-IV and both DB elements, as well as the terminal 3'SL element common to all flaviviruses (Figure 9).

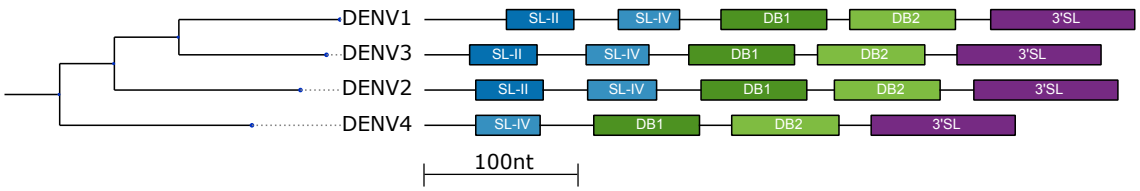

Figure 9: Annotated 3'UTR of viruses in the Dengue virus group

The two SL copies in domain I of DENV1, DENV2 and DENV3 have been extensively studied and were used as model systems to elucidate xrRNA functionality [34]. Notably, DENV4 stands out in this region of the 3'UTR, as it only contains a single SL element. Careful inspection revealed that the DENV4 SL exhibits higher similarity to SL-IV than SL-II, as verified by covariance model-based comparison to SL elements in other flaviviruses. The presence of just a single xrRNA-associated SL-IV element in DENV4 is unexpected, as it has been suggested that two SL copies account for vector/host specificity and provide backup functionality to ensure proper sfRNA generation in case exoribonuclease halting is not achieved by the first SL copy [39]. Nevertheless, the presence of a single SL-IV copy in DENV4 allows us to hypothesize that this virus has lost the other SL element due to unknown evolutionary constrains. The remaining parts of DENVG 3'UTRs are highly homogeneous, exhibiting duplicated DB elements and a terminal 3'SL structure. The duplicated DB elements present in all four Dengue serotypes have been actively studied and it was hypothesized that two DB elements are required in flaviviruses that have to replicate in different species [56].

4.6 Spondweni/Kedougou virus group

Another antigenic flavivirus complex that has attracted considerable research interest in recent years is the Spondweni virus group, containing Spondweni virus (SPOV) and Zika virus (ZIKV). The latter is considered an emerging pathogen, being responsible for a large outbreak in the Americas 2015-2017 that resulted in more than 1.5 million infections in Brazil alone and approximately 3500 cases of congenital microcephaly [59]. SPOVG shares ancestral roots with DENVG (Figure 1), and although it formally contains only two species, we consider it here together with Kedougou virus (KEDV), which phylogenetically clusters with SPOV and ZIKV. We examine the 3'UTRs of two ZIKV lineages, the ancestral African lineage based on an isolate from Uganda (ZIKV-UG, 428 nt) and the American lineage based on a Brazilian isolate (ZIKV-BR, 429 nt) (Figure 10). The 3'UTR of SPOV is 338 nt long and truncated within the 3'SL element. KEDV has a 3'UTR of 390 nt length.

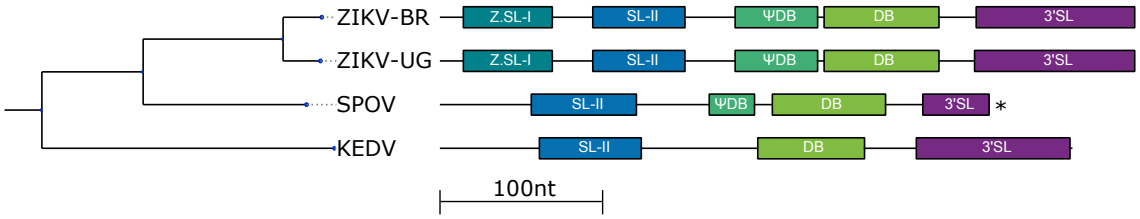

Figure 10: Annotated 3'UTR of viruses in the Spondweni virus group

Comparison of the 3'UTRs of SPOVG reveals a variable 3'UTR organization in domains I and II, and a common 3'SL element in domain III. The ZIKV 3'UTR architecture is particularly interesting, as it stands out among MBFVs. While ZIKV has duplicated xrRNA-associated SL elements in domain I, and bona fidae exoribonuclease stalling capacity has been experimentally confirmed [60], our covariance model screen revealed that the second SL element is fairly related to DENVG SL-II, while the first SL (Z.SL-I) is only remotely related to other SL elements. This suggests that Z.SL-I emerged from a ZIKV-specific duplication or recombination event. Conversely, domain I in SPOV and KEDV contain a single SL-II structure each, without evidence for the presence of a ZIKV-like SL-I element. Domain II of the SPOVG viruses is no less interesting, as it contains only a single canonical DB element preceded by a pseudo-DB (ΨDB) element in ZIKV and SPOV, but not in KEDV. Importantly, the ΨDB element not homologous to DB elements found in ZIKV or any other MBFV species (see Section 4.8). The terminal 3'SL structure is again found in all SPOVG viruses.

4.7 Yellow fever virus group

Two antigenic complexes within the MBFVs, the Yellow fever virus group and the Edge hill virus group are located ancestral to the other MBFV groups in the complete coding sequence nucleotide phylogeny (Figure 1). The YFVG includes the species Yellow fever virus (YFV), Wesselsbron virus (WESSV), Fitzroy River virus (FRV) and Sepik virus (SEPV), with 3'UTR lengths ranging from 459 nt (SEPV) to 657 nt (YFV). Several genotypes of YFV are known that differ in replicative fitness and virulence [61], and whose 3'UTRs differ in length and exhibit structural heterogeneity [62, 63]. We consider here an isolate of the YFV South American I genotype as well as the live attenuated YFV 17D strain to showcase within-species 3'UTR architectural diversity.

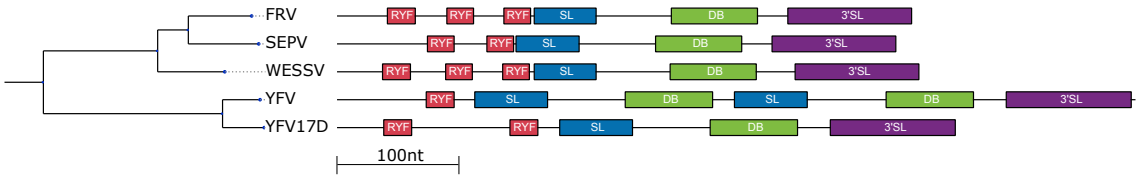

Figure 11: Annotated 3'UTR of viruses in the Yellow fever virus group

The YFVG 3'UTR organization is in qualitative agreement with that of other MBFV groups, however, there are several characteristics that make YFVG 3'UTRs unique. Probably the most obvious is the presence of a variable number of imperfect sequence repeats (RYFs) of approximately 40nt length that fold into simple stem-loop strictures at the beginning of the 3'UTR (Figure 11). While these sequence repeats have been predicted to fold into two small, consecutive hairpins (denoted G and F) in the literature [64], our predictions suggest that only the downstream part (F hairpin) forms a thermodynamically stable structure in a genomic context. The upstream repeat region (G), however, is involved in the formation of specific folds that are not structurally conserved throughout the YFVG. Focusing on structure conservation, we only consider the downstream (F) hairpin of the RYF elements with a length between 20 nt and 22 nt. RYF elements occur in variable copy numbers, depending on species and genotype, and show a good covariation support. Figure 12 shows the consensus secondary structure of the RYF hairpin and illustrates the formation of three RYF hairpins in the 3'UTR of Fitzroy River virus. RYF upstream regions (G region) of these elements do not fold into conserved structures.

Downstream of the RYF-containing region, an xrRNA-associated three-way-junction forming stem-loop element (denoted SL-E in literature) is found in many YFVG species. This element confers protection of downstream regions against exoribonuclease degradation [65] and exhibits a high similarity to the NTAVG SL-II structure. Moreover, a DB element that remotely resembles the DB elements of the other MBFV antigenic groups, but with a longer proximal arm, is found in all members of the YFVG groups, as well as the obligatory terminal 3'SL structure. Notably, the region that contains the SL-E and DB elements is duplicated in the YFV isolate discussed here (Figure 11).

While YFVG SL-E and DB elements appear to exhibit similar biological functionality to their remote homologs in the other flaviviruses, their structural divergence is likely a result of their evolutionary distance to the other MBFV groups. This is also suggested by the positioning of the YFVG clade in the phylogenetic tree relative to the other MBFV groups.

4.8 Structural diversity of conserved elements

Structural similarity among duplicated elements has previously been addressed by tree alignment approaches. In the case of DENV, it has been suggested that orthologous DB elements

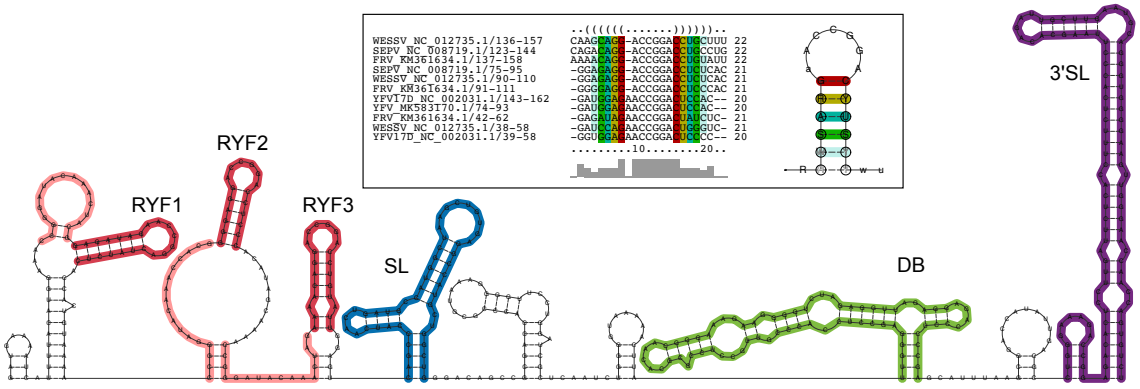

Figure 12: Secondary structure prediction of the Fitzroy River virus 3'UTR. Colored structurally conserved elements match those depicted in Figure 11. Structurally non-conserved G repeats are shown in faint red upstream of the RYF hairpins. The insert shows a structural alignment and consensus structure prediction of 3' half of RYF elements (F hairpin). Coloring of the insert follows the `RNAalifold` scheme, highlighting different covariation levels (Figure 3 **a**).

share a higher similarity than paralogous DB elements. Specifically, DB1 elements (and DB2 elements, respectively) of different Dengue serotypes are more alike among themselves than DB1 and DB2 elements from the same serotype [56]. Similar examples, encompassing xrRNA-associated SL elements, have been proposed in the four Dengue serotypes and several species of the JEVG [39], suggesting divergent evolutionary trajectories after the duplication events.

We revisit this topic here and present a comprehensive comparison of functional RNA elements across all antigenic MBFV groups, focusing specifically on DB elements and xrRNA-associated SL elements. Rather than comparing the sequences and structures of individual strains, we directly compare the covariance models built from alignments of representative sequences in each antigenic group. We use `CMCompare` [54], which for two given covariance models finds the link sequence, defined as the sequence giving the highest score on both models simultaneously. The associated link score is a measure for the similarity of the models that considers both sequence and structure similarity of the two groups. Models for closely related elements will exhibit high link scores, while low or even negative link scores indicate unrelatedness of the models.

The link scores computed from all pairwise comparisons of DB element covariance models discussed in this contribution are shown in Figure 13 as a similarity matrix. Structural proximity among DB elements as well as the SPOVG psiDB element is color-coded, highlighting a dense network of structural similarity among duplicated DB elements. Intriguingly, our data do not support the hypothesis that orthologous entities are generally more alike than paralogous entities. Rather, most DB elements exhibit high link scores to either orthologous or paralogous elements, suggesting that they could also be captured by a single covariance model.

Comparison of the SPOVG DB element with all other DB elements reveals slightly higher link scores to DB2 elements than DB1 elements. Likewise, the DB element present in the YFVG has only very weak similarity to the other DB elements, with the highest link score to the DB2 of KOKVG. Importantly, the SPOVG psiDB element has only negative or very low link scores < 2, indicating that this element is not homologous to any other DB element.

The situation is different for the comparison of xrRNA-associated SL elements (denoted SL elements here). As shown in Figure 14, our data do not suggest pervasive similarity among the majority of elements, but rather a partitioning into distinct sets that exhibit high intra-set scores, i.e. link scores to elements of the same set. Conversely, the link scores between SL elements of different sets, i.e. inter-set scores, are consistently low. The set with the highest intra-set scores comprises the SL-II and SL-IV elements of JEVG, as well as the SL-II of

Figure 13: Similarity matrix of DB elements computed from pairwise covariance model comparisons. Link score values highlighted as different shades of olive indicate the similarity levels between pairs. Diagonal entries represent the score that can be achieved by comparing a model to itself, i.e. the score of the best link sequence.

NTAVG, and the single YFVG SL element. Interestingly, the NTAVG SL-IV is not part of this set, but has only weak link scores to the other SL elements. A second group of elements with high intra-set link-scores encompasses DENVG, KOKVG and AROAVG, each with their SL-II and SL-IV elements.

While we could not identify an example where the highest link score of an element is with its tandem copy in the same species, there are cases where we see fairly high link scores among paralogous elements, e.g. the SL-II/SL-IV elements of JEVG and KOKVG. Importantly, the observation that several elements have their highest link score with orthologous elements of viruses within the same set/cluster, and at the same time these elements have particularly low link scores with their orthologs from the other cluster argues against the hypothesis that a fixed architecture of duplicated SL has been conserved from the last common ancestor of all MBFVs. Rather, these data suggest a scenario of repeated duplication and loss events of SL elements, which is also supported by the weak similarity of SPOVG SL elements: While the SPOVG SL-II exhibits the highest similarity to the DENVG SL-II element, our data show that the ZIKV SL-I structure is only very distantly related to the other SL elements, suggesting that this could have emerged from a ZIKV-specific duplication or recombination event.

5 Conclusion

Flaviviruses 3'UTRs have increasingly attracted research interest over the last years, both experimentally and theoretically. These regions are special, as they exhibit a striking level of plasticity that is manifested in a varied pattern of conserved RNA elements whose functional associations have been solved for some flaviviruses. A profound knowledge about the structural landscape of these genomic regions is crucial to understanding the mechanistic traits that modulate the viral life cycle and pathology.

The availability of large numbers of viral genome data over the last years has made available

Figure 14: Link scores computed from pairwise covariance model comparisons of xrRNA-associated SL elements. Diagonal entries show the maximal score that can be achieved when comparing a motel with itself. See text for details.

the possibility to perform comparative genomics screens at an unprecedented level of detail, allowing researchers to elucidate hitherto unknown structural traits *in silico*. In this context, comparative genomics screens based on structural alignments and RNA family models turned out as a particularly useful approach for the detection of homologous structures, thereby overcoming the limitations of sequence-based homology search approaches. Recently, an alignment-free, descriptor-based method for structured RNA pattern search has been proposed [66]. Such methods allow to include tertiary interactions like pseudoknots or base triplets that cannot be covered in covariance models. In contrast to covariance models which are automatically built from a consensus structure and sequence alignment, descriptor models have to be hand-crafted. Thus both methods have their specific advantages and can complement each other.

Many RNA viruses have successfully evolved strategies to tolerate error-prone replication, resulting in increased mutation frequencies that allow them to evade host immune responses. Likewise, there is evolutionary pressure on maintaining functional entities, such as structured elements in the UTRs that are crucially involved in virus replication and mediating pathogenesis. Knowledge about the exact location of these evolutionarily conserved elements within the 3'UTR of each antigenic group allows not only to characterize homologous structures in novel strains but also to infer the proximity between functionally conserved RNAs.

The presence of multiple copies of functional RNA elements in many flavivirus 3'UTRs likely resulted from repeated duplication and loss events, yielding diversified structures that allow the virus to adapt to its specific combnination of host and vector. The presence of tandem copies (or even higher copy numbers in some insect-specific flaviviruses [23]) of SL and DB elements raises the question why a lot of viruses maintain potentially redundant parts of their genomes in spite of the strong seletion pressure on genome length. Some evidence suggests that duplicated elements may serve specific roles in different stages of the viral life cycle, such as vector and host stages, [67, 40, 56]. Knowledge about a particular instance of an element in a specific virus can be extrapolated to related species. The obvious quest is to derive functional knowledge from studied examples and extrapolate this to newly discovered or (re-)emerging

viruses. This is particularly useful for addressing open questions like the functional association of structured RNAs with vector/host specificity. Moreover, a detailed understanding of the proximity among common structural patterns in different viruses is crucial for studying virus UTR evolution.

References

[1] Scott C Weaver, Caroline Charlier, Nikos Vasilakis, and Marc Lecuit. Zika, chikungunya, and other emerging vector-borne viral diseases. *Annu. Rev. Med.*, 69:395–408, 2018.

[2] Bradley Blitvich and Andrew Firth. Insect-specific flaviviruses: A systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. *Viruses*, 7(4):1927–1959, 2015.

[3] Peter M Howley and David M Knipe, editors. *Fields Virology: Emerging Viruses*. Lippincott Williams & Wilkins, 7th edition, 2020.

[4] Bradley Blitvich and Andrew Firth. A review of flaviviruses that have no known arthropod vector. *Viruses*, 9(6):154, 2017.

[5] Rhys Parry and Sassan Asgari. Discovery of novel crustacean and cephalopod flaviviruses: Insights into the evolution and circulation of flaviviruses between marine invertebrate and vertebrate hosts. *J. Virol.*, 93(14):e00432–19, 2019.

[6] Theodore C Pierson and Michael S Diamond. The continued threat of emerging flaviviruses. *Nature Microbiol.*, pages 1–17, 2020.

[7] Franz X Heinz and Karin Stiasny. Flaviviruses and flavivirus vaccines. *Vaccine*, 30(29):4301–4306, 2012.

[8] Gilda Grard, Gregory Moureau, Remi N Charrel, Edward C Holmes, Ernest A Gould, and Xavier de Lamballerie. Genomics and evolution of Aedes-borne flaviviruses. *J. Gen. Virol.*, 91(1):87–94, 2010.

[9] Wy Ng, Ruben Soto-Acosta, Shelton Bradrick, Mariano Garcia-Blanco, and Eng Ooi. The 5' and 3' untranslated regions of the flaviviral genome. *Viruses*, 9(6):137, 2017.

[10] Charles M Rice, Edith M Lenches, Sean R Eddy, SJ Shin, RL Sheets, and JH Strauss. Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. *Science*, 229(4715):726–733, 1985.

[11] Sergio M Villordo, Diego E Alvarez, and Andrea V Gamarnik. A balance between circular and linear forms of the dengue virus genome is crucial for viral replication. *RNA*, 16(12):2325–2335, 2010.

[12] Luana de Borba, Sergio M Villordo, Nestor G Iglesias, Claudia V Filomatori, Leopoldo G Gebhard, and Andrea V Gamarnik. Overlapping local and long-range RNA-RNA interactions modulate dengue virus genome cyclization and replication. *J. Virol.*, 89(6):3430–3437, 2015.

[13] Chang S Hahn, Young S Hahn, Charles M Rice, Eva Lee, Lynn Dalgarno, Ellen G Strauss, and James H Strauss. Conserved elements in the 3' untranslated region of flavivirus RNAs and potential cyclization sequences. *J Mol Biol*, 198(1):33–41, 1987.

[14] Margo A Brinton and Mausumi Basu. Functions of the 3' and 5' genome RNA regions of members of the genus Flavivirus. *Virus Res*, 206:108–119, 2015.

[15] Michael Kiening, Roman Ochsenreiter, Hans-Jörg Hellinger, Thomas Rattei, Ivo Hofacker, and Dmitrij Frishman. Conserved secondary structures in viral mRNAs. *Viruses*, 11(5):401, 2019.

[16] Christina Witwer, Susanne Rauscher, Ivo L Hofacker, and Peter F Stadler. Conserved RNA secondary structures in picornaviridae genomes. *Nucleic acids res*, 29(24):5079–5089, 2001.

[17] Caroline Thurner, Christina Witwer, Ivo L Hofacker, and Peter F Stadler. Conserved RNA secondary structures in flaviviridae genomes. *J Gen Virol*, 85(5):1113–1124, 2004.

[18] Adriano de Bernardi Schneider, Roman Ochsenreiter, Reilly Hostager, Ivo L. Hofacker, Daniel Janies, and Michael T. Wolfinger. Updated Phylogeny of Chikungunya Virus Suggests Lineage-Specific RNA Architecture. *Viruses*, 11:798, 2019.

[19] Dong Yang and Julian L Leibowitz. The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res.*, 206:120–133, 2015.

[20] Ramakanth Madhugiri, Markus Fricke, Manja Marz, and John Ziebuhr. RNA structure analysis of alphacoronavirus terminal genome regions. *Virus Res.*, 194:76–89, 2014.

[21] Ramakanth Madhugiri, Nadja Karl, Daniel Petersen, Kevin Lamkiewicz, Markus Fricke, Ulrike Wend, Robina Scheuer, Manja Marz, and John Ziebuhr. Structural and functional conservation of cis-acting RNA elements in coronavirus 5'-terminal genome regions. *Virology*, 517:44–55, 2018.

[22] Alexandra Popa, Jakob-Wendelin Genger, Michael D. Nicholson, Thomas Penz, Daniela Schmid, Stephan W Aberle, Benedikt Agerer, Alexander Lercher, Lukas Endler, Henrique Colaco, Mark Smyth, Michael Schuster, Miguel L. Grau, Francisco Martínez-Jiménez, Oriol Pich, Wegene Borena, Erich Pawelka, Zsofia Keszei, Martin Senekowitsch, Jan Laine, Judith H Aberle, Monika Redlberger-Fritz, Mario Karolyi, Alexander Zoufaly, Sabine Maritschnik, Martin Borkovec, Peter Hufnagl, Manfred Nairz, Günter Weiss, Michael T Wolfinger, Dorothee von Laer, Giulio Superti-Furga, Nuria Lopez-Bigas, Elisabeth Puchhammer-Stöckl, Franz Allerberger, Franziska Michor, Christoph Bock, and Andreas Bergthaler. Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Sci. Transl. Med.*, 12, 2020.

[23] Roman Ochsenreiter, Ivo L Hofacker, and Michael T Wolfinger. Functional RNA Structures in the 3' UTR of Tick-Borne, Insect-Specific and No-Known-Vector Flaviviruses. *Viruses*, 11(3):298, 2019.

[24] Gorben P Pijlman, Anneke Funk, Natasha Kondratieva, Jason Leung, Shessy Torres, Lieke Van der Aa, Wen Jun Liu, Ann C Palmenberg, Pei-Yong Shi, Roy A Hall, et al. A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. *Cell Host & Microbe*, 4(6):579–591, 2008.

[25] Benjamin M Akiyama, Daniel Eiler, and Jeffrey S Kieft. Structured RNAs that evade or confound exonucleases: function follows form. *Curr Opin Struc Biol*, 36:40–47, 2016.

[26] Andrii Slonchak and Alexander A Khromykh. Subgenomic flaviviral RNAs: What do we know after the first decade of research. *Antivir. Res.*, 159:13–25, 2018.

[27] Christopher Iain Jones, Maria Vasilyevna Zabolotskaya, and Sarah Faith Newbury. The 5'-> 3' exoribonuclease Xrn1/pacman and its functions in cellular processes and development. *Wiley Interdisciplinary Reviews: RNA*, 3(4):455–468, 2012.

[28] BD Clarke, JA Roby, A Slonchak, and AA Khromykh. Functional non-coding RNAs derived from the flavivirus 3' untranslated region. *Virus Res*, 206:53–61, 2015.

[29] Claudia V Filomatori, Juan M Carballeda, Sergio M Villordo, Sebastian Aguirre, Horacio M Pallarés, Ana M Maestre, Irma Sánchez-Vargas, Carol D Blair, Cintia Fabri, Maria A Morales, Ana Fernandez-Sesma, and Andrea V. Gamarnik. Dengue virus genomic variation associated with mosquito adaptation defines the pattern of viral non-coding RNAs and fitness in human cells. *PLoS Pathog.*, 13(3):e1006265, 2017.

[30] Andrea MacFadden, Zoe O'Donoghue, Patricia AGC Silva, Erich G Chapman, René C Olsthoorn, Mark G Sterken, Gorben P Pijlman, Peter J Bredenbeek, and Jeffrey S Kieft. Mechanism and structural diversity of exoribonuclease-resistant RNA structures in flaviviral RNAs. *Nat Commun*, 9(1):119, 2018.

[31] Rachel A. Jones, Anna-Lena Steckelberg, Matthew J. Szucs, Benjamin M. Akiyama, Quentin Vicens, and Jeffrey S. Kieft. Different tertiary interactions create the same important 3-D features in a divergent flavivirus xrRNA. *bioRxiv*, 2020.

[32] Christida E. Wastika, Hayato Harima, Sasaki Michihito, Bernard M. Hang'ombe, Yuki Eshita, Qiu Yongjin, William W. Hall, Michael T. Wolfinger, Hirofumi Sawa, and Yasuko Orba. Discoveries of Exoribonuclease-Resistant Structures of Insect-Specific Flaviviruses Isolated in Zambia. *Viruses*, 12:1017, 2020. PMID: 32933075.

[33] Xiaolin Niu, Qiuhan Liu, Zhonghe Xu, Zhifeng Chen, Linghui Xu, Lilei Xu, Jinghong Li, and Xianyang Fang. Molecular mechanisms underlying the extreme mechanical anisotropy of the flaviviral exoribonuclease-resistant RNAs (xrRNAs). *Nat Commun*, 11:5496, 2020.

[34] Erich G Chapman, Stephanie L Moon, Jeffrey Wilusz, and Jeffrey S Kieft. RNA structures that resist degradation by Xrn1 produce a pathogenic Dengue virus RNA. *elife*, 3:e01892, 2014.

[35] Anneke Funk, Katherine Truong, Tomoko Nagasaki, Shessy Torres, Nadia Floden, Ezequiel Balmori Melian, Judy Edmonds, Hongping Dong, Pei-Yong Shi, and Alexander A Khromykh. RNA structures required for production of subgenomic flavivirus RNA. *J Virol*, 84(21):11407–11417, 2010.

[36] Susanne Rauscher, Christoph Flamm, Christian W Mandl, Franz X Heinz, and Peter F Stadler. Secondary structure of the 3'-noncoding region of flavivirus genomes: comparative analysis of base pairing probabilities. *RNA*, 3(7):779–791, 1997.

[37] Ivo L. Hofacker, Martin Fekete, Christoph Flamm, Martijn A. Huynen, Susanne Rauscher, Paul E. Stolorz, and Peter F. Stadler. Automatic detection of conserved RNA structure elements in complete RNA virus genomes. *Nucleic acids res*, 26(16):3825–3836, 1998.

[38] Ivo L Hofacker, Peter F Stadler, and Roman R Stocsits. Conserved RNA secondary structures in viral genomes: a survey. *Bioinformatics*, 20(10):1495–1499, 2004.

[39] Sergio M Villordo, Juan M Carballeda, Claudia V Filomatori, and Andrea V Gamarnik. RNA structure duplications and flavivirus host adaptation. *Trends Microbiol*, 24(4):270–283, 2016.

[40] Joanna Sztuba-Solinska, Tadahisa Teramoto, Jason W Rausch, Bruce A Shapiro, Radhakrishnan Padmanabhan, and Stuart FJ Le Grice. Structural complexity of dengue virus untranslated regions: cis-acting RNA motifs and pseudoknot interactions modulating functionality of the viral genome. *Nucleic acids res*, page gkt203, 2013.

[41] Adriano de Bernardi Schneider and Michael T. Wolfinger. Musashi binding elements in Zika and related Flavivirus 3'UTRs: A comparative study *in silico*. *Sci. Rep.*, 9(1):6911, 2019. PMCID: PMC6502878.

[42] Alexander A Khromykh, Hedije Meka, Kimberley J Guyatt, and Edwin G Westaway. Essential role of cyclization sequences in flavivirus RNA replication. *J. Virol.*, 75(14):6719–6728, 2001.

[43] Diego E Alvarez, María F Lodeiro, Silvio J Luduena, Lía I Pietrasanta, and Andrea V Gamarnik. Long-range RNA-RNA interactions circularize the dengue virus genome. *J. Virol.*, 79(11):6631–6643, 2005.

[44] Claudia V Filomatori, Maria F Lodeiro, Diego E Alvarez, Marcelo M Samsa, Lía Pietrasanta, and Andrea V Gamarnik. A 5' RNA element promotes dengue virus RNA synthesis on a circular genome. *Gene Dev*, 20(16):2238–2249, 2006.

[45] Susann Friedrich, Susanne Engelmann, Tobias Schmidt, Grit Szczepankiewicz, Sandra Bergs, Uwe G Liebert, Beate M Kümmerer, Ralph P Golbik, and Sven-Erik Behrens. The host factor AUF1 p45 supports flavivirus propagation by triggering the RNA switch required for viral genome cyclization. *J. Virol.*, 92(6):e01647–17, 2018.

[46] Kazutaka Katoh and Daron M Standley. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*, 30(4):772–780, 2013.

[47] Lam-Tung Nguyen, Heiko A Schmidt, Arndt von Haeseler, and Bui Quang Minh. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*, 32(1):268–274, 2014.

[48] Eric P. Nawrocki and Sean R. Eddy. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, 29:2933–2935, 2013.

[49] Sean R Eddy and Richard Durbin. RNA sequence analysis using covariance models. *Nucleic Acids Res*, 22(11):2079–2088, 1994.

[50] Sebastian Will, Kristin Reiche, Ivo L Hofacker, Peter F Stadler, and Rolf Backofen. Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comp Biol*, 3(4):e65, 2007.

[51] Sam Griffiths-Jones, Alex Bateman, Mhairi Marshall, Ajay Khanna, and Sean R Eddy. Rfam: an RNA family database. *Nucleic Acids Res*, 31(1):439–441, 2003.

[52] Stephan H Bernhart, Ivo L Hofacker, Sebastian Will, Andreas R Gruber, and Peter F Stadler. RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics*, 9(1):474, 2008.

[53] Ronny Lorenz, Stephan H Bernhart, Christian Hoener Zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F Stadler, and Ivo L Hofacker. Viennarna Package 2.0. *Algorithm Mol Biol*, 6(1):26, 2011.

[54] Christian Höner zu Siederdissen and Ivo L Hofacker. Discriminatory power of RNA family models. *Bioinformatics*, 26(18):i453–i459, 2010.

[55] Clément Mazeaud, Wesley Freppel, and Laurent Chatel-Chaix. The multiples fates of the Flavivirus RNA genome during pathogenesis. *Front. Genet.*, 9:595, 2018.

[56] Luana de Borba, Sergio M Villordo, Franco L Marsico, Juan M Carballeda, Claudia V Filomatori, Leopoldo G Gebhard, Horacio M Pallarés, Sebastian Lequime, Louis Lambrechts, Irma Sánchez Vargas, Carol D. Blair, and Andrea V. Gamarnik. RNA Structure Duplication in the Dengue Virus 3' UTR: Redundancy or Host Specificity? *mBio*, 10(1):e02506–18, 2019.

[57] Vladimir A Ternovoi, Anastasia V Gladysheva, Eugenia P Ponomareva, Tamara P Mikryukova, Elena V Protopopova, Alexander N Shvalov, Svetlana N Konovalova, Eugene V Chausov, and Valery B Loktev. Variability in the 3' untranslated regions of the genomes of the different tick-borne encephalitis virus subtypes. *Virus Genes*, 55(4):448–457, 2019.

[58] Gregory Moureau, Shelley Cook, Philippe Lemey, Antoine Nougairede, Naomi L Forrester, Maxim Khasnatinov, Remi N Charrel, Andrew E Firth, Ernest A Gould, and Xavier De Lamballerie. New insights into flavivirus evolution, taxonomy and biogeographic history, extended by analysis of canonical and alternative coding sequences. *PLoS One*, 10(2):e0117849, 2015.

[59] Thalia Velho Barreto de Araújo, Ricardo Arraes de Alencar Ximenes, Demócrito de Barros Miranda-Filho, Wayner Vieira Souza, Ulisses Ramos Montarroyos, Ana Paula Lopes de Melo, Sandra Valongueiro, Cynthia Braga, Sinval Pinto Brandão Filho, Marli Tenório Cordeiro, et al. Association between microcephaly, Zika virus infection, and other risk factors in Brazil: final report of a case-control study. *Lancet Infect Dis*, 18(3):328–336, 2018.

[60] Benjamin M Akiyama, Hannah M Laurence, Aaron R Massey, DavidKamit A Costantino, Xuping Xie, Yujiao Yang, Pei-Yong Shi, Jay C Nix, J David Beckham, and Jeffrey S Kieft. Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease. *Science*, page aah3963, 2016.

[61] ADT Barrett and EA Gould. Comparison of neurovirulence of different strains of yellow fever virus in mice. *J Gen Virol*, 67(4):631–637, 1986.

[62] Juliet E Bryant, Pedro FC Vasconcelos, Rene CA Rijnbrand, JP Mutebi, Stephen Higgs, and Alan DT Barrett. Size heterogeneity in the 3' noncoding region of South American isolates of yellow fever virus. *J Virol*, 79(6):3807–3821, 2005.

[63] Raphaëlle Klitting, Carlo Fischer, Jan Drexler, Ernest Gould, David Roiz, Christophe Paupy, and Xavier de Lamballerie. What Does the Future Hold for Yellow Fever Virus? (II). *Genes*, 9(9):425, Aug 2018.

[64] John-Paul Mutebi, René CA Rijnbrand, Heiman Wang, Kate D Ryman, Eryu Wang, Lynda D Fulop, Rick Titball, and Alan DT Barrett. Genetic relationships and evolution of genotypes of yellow fever virus and other members of the yellow fever virus group within the flavivirus genus based on the 3' noncoding region. *J. Virol*, 78(18):9652–9665, 2004.

[65] Patrícia AGC Silva, Carina F Pereira, Tim J Dalebout, Willy JM Spaan, and Peter J Bredenbeek. An RNA pseudoknot is required for production of yellow fever virus subgenomic RNA by the host nuclease Xrn1. *J Virol*, 84(21):11395–11406, 2010.

[66] Alan Zammit, Leon Helwerda, René CL Olsthoorn, Fons J Verbeek, and Alexander P Gultyaev. A database of flavivirus RNA structures with a search algorithm for pseudoknots and triple base interactions. *Bioinformatics*, page btaa759, 2020.

[67] Mark Manzano, Erin D Reichert, Stephanie Polo, Barry Falgout, Wojciech Kasprzak, Bruce A Shapiro, and Radhakrishnan Padmanabhan. Identification of cis-acting elements in the 3'-untranslated region of the dengue virus type 2 RNA that modulate translation and replication. *J Biol Chem*, pages jbc–M111, 2011.