View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Available online at www.sciencedirect.com

ScienceDirect

Virology 358 (2007) 258–265

VIROLOGY

www.elsevier.com/locate/yviro

Rapid Communication

Direct repeats in the flavivirus 3′ untranslated region; a strategy for survival in the environment?

T.S. Gritsun *, E.A. Gould

Centre for Ecology and Hydrology, Mansfield Rd, Oxford 0X1 3SR, UK

Received 9 August 2006; returned to author for revision 7 September 2006; accepted 12 September 2006
Available online 24 October 2006

Abstract

Previously, direct repeats (DRs) of 20–70 nucleotides were identified in the 3′ untranslated regions (3′UTR) of flavivirus sequences. To address their functional significance, we have manually generated a pan-flavivirus 3′UTR alignment and correlated it with the corresponding predicted RNA secondary structures. This approach revealed that intra-group-conserved DRs evolved from six long repeated sequences (LRSs) which, as ~200-nucleotide domains were preserved only in the genomes of the slowly evolving tick-borne flaviviruses. We propose that short DRs represent the evolutionary remnants of LRSs rather than distinct molecular duplications. The relevance of DRs to virus replication enhancer function, and thus survival, is discussed.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Flaviviruses; Alignment "by eye"; Direct repeats; 3′UTR; Promoter; Enhancer

Introduction

The flaviviruses (family Flaviviridae genus *Flavivirus*) are subdivided into three groups based on their ecological characteristics—mosquito-borne, tick-borne and no-known vector flaviviruses (MBFV, TBFV and NKV) respectively (Heinz et al., 2000). Three non-classified flaviviruses viruses (NCFV) that share similar genome organization, namely cell fusion agent virus (CFAV), Kamiti River virus (KRV) and Tamana bat virus (TBV), are currently considered to be tentative species of this genus (de Lamballerie et al., 2002). Virions (~50 nm) contain a core consisting of C protein and positive-sense ssRNA (~11 kb), and a membrane with viral E and M glycoproteins. A single open reading frame (ORF) encodes a polyprotein of about 3400 amino acids which is processed into the individual structural and nonstructural proteins; the latter provide proteolytic and RNA-replicative functions. The ORF is flanked by 5′ and 3′UTRs that initiate virus translation and replication. A distinct feature of the 3′UTR is the presence of direct repeats (DRs) consisting of 20–70 nucleotide duplicated sequences separated by varying lengths of non-repeated regions. Initially, DRs were identified as conserved sequences (CSs) and their repeats (RCSs) within the MBFV groups (Hahn et al., 1987) but subsequently they were described under different names in other flavivirus groups (Charlier et al., 2002; Crabtree et al., 2003; Gritsun et al., 1997; Wallner et al., 1995). The use of infectious clones to introduce genetic changes revealed an apparent contradiction between intra-group conservation of DRs and their redundancy for virus viability:; the deletion of individual DRs did not result in the loss of virus viability (Mandl et al., 1998; Men et al., 1996). In an attempt to explain these observations we initially traced the origin and evolution of the 3′UTR by constructing robust alignments for each flavivirus group (Gritsun and Gould, 2006a, 2006b, 2006c). Assuming that the evolution of the 3′UTRs, in addition to single substitutions, was mostly based on deletions and duplications we adjusted these alignments "by eye" introducing long gaps to reveal maximum homology between distantly located regions. This method revealed 6 long repeated sequences (LRS, ~200 nucleotides) in the 3′UTR of the TBFV group each of which, as we concluded, subsequently evolved to shorter DRs designated R1, R2 and R3 (Gritsun and Gould, 2006a). In the NCFV group we demonstrated that the 3′UTR of

* Corresponding author.
E-mail address: tsg@ceh.ac.uk (T.S. Gritsun).

0042-6822/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2006.09.033

KRV originated by duplication of the entire 3′UTR and we proposed that subsequent evolution of these duplicates reduced the homology to two 67-nucleotide DRs (Gritsun and Gould, 2006b). Alignment of the three MBFV groups, i.e., Japanese encephalitis virus (JEV), Dengue virus (DENV) and Yellow fever virus (YFV) groups, revealed that all their DRs (i.e., CSs/RCSs) evolved from an MBFV precursor lineage and subsequently were reduced or totally deleted within the individual MBFV groups. These data indicated that the DRs within the different flavivirus groups probably originated from the LRSs that we identified in the TBFV group. By developing a pan-flavivirus alignment, we now demonstrate more extensive homology between distantly related viruses than was recognized previously. The DRs that were previously attributed to individual flavivirus groups are now shown to share homology and while under laboratory conditions they may not be essential for virus viability, in the more demanding conditions of the environment they may play an important role in fitness for survival by mediating rapid replication rates.

Results and discussion

Construction of a pan-flavivirus alignment

A 3′UTR pan-flavivirus alignment (Fig. 1) was generated manually by aligning the alignments produced previously for each of the four individual flavivirus groups, i.e., MBFV (YFV, JEV and DENV) (Gritsun and Gould, 2006c), TBFV (TBEV, POW, LGTV, LIV, OHFV KFDV) (Gritsun and Gould, 2006a), NKV (MODV, RBV, APOIV and MMLV) (Supplementary Fig. 2) and NCFV (KRV, CFAV and TBV) (Gritsun and Gould, 2006b) and edited by "eye". The alignment process was initially carried out in the 3′ → 5′direction since the 3′terminal regions contained more homology than those in close proximity to the stop codon and the results are also described in the 3′ → 5′ direction (Supplementary Fig. 1a). The alignment was then correlated with the corresponding RNA secondary structures (Fig. 3). Sequences conserved between distantly related groups were also used as anchors for the subsequent upstream and downstream alignment of more variable regions. Thus, regions of high homology were separated by regions of lower homology. However, the major "anchors" for this pan-flavivirus alignment were the LRSs, consisting of ~200-nucleotide domains that were previously identified in the ORF/3′UTR of the TBFV (Gritsun and Gould, 2006a). Each LRS (except in the case of LRS1/LRS2), aligned with the corresponding sequences of MBFV, NKV and NCFV groups, is presented on a separate page (Supplementary Fig. 1a). The KRV and CFAV 3′UTRs are included in the alignment as two aligned halves corresponding to the duplicated sequence (Gritsun and Gould, 2006b).

Correspondence of RNA secondary structures with the pan-flavivirus alignment

Previously, computer simulations produced two models of RNA secondary structure in the flavivirus 3′UTR (reviewed in Gritsun et al., 2006; Markoff, 2003). However, they have never been directly compared. Therefore, we compared the RNA conformations for the different flavivirus groups (Fig. 3) and used them to construct the pan-flavivirus alignment (Fig. 1) and also to deduce their evolution (Fig. 4). The first model predicted linear (independent from the 5′UTR) folding of the 3′UTR in which shared and distinct stem-loop (SL) conformations were identified between the four flavivirus groups (Figs. 3a–d). The most conserved conformation was the terminal long stable hairpin (3′LSH) with an adjacent short stem-loop 2 (SL2). The 3′LSH contains a conserved pentanucleotide CACAG that is exposed as a loop and is essential for virus infectivity. The 3′LSH interacts with viral and cellular proteins within the polymerase complex thus initiating replication of the viral genome. For the MBFV and TBFV, the structures located upstream of the 3′LSH-SL2 are different (Figs. 3a–d). The first shared MBFV conformation resembles a dumbbell (DB1) that exposes the CS2, i.e., a DR of the MBFV (Figs. 3a–b). Viruses within the JEV and DENV groups formed pseudo-duplicated DB2, also with exposed RCS2. For YFV the region equivalent to DB2 presented a different conformation. The CS3/RCS3, the DRs found only in the JEV group, and the three YFV-specific tandem repeats (YFV-R1, -R2 and -R3) also formed distinct local SLs. In the TBFV the RNA structure upstream of the 3′LSH-SL2 (Fig. 4c) is Y-shaped, with loops 3 and 4 conserved between the TBFV and NKV (Fig. 4d). The Y-shaped structure and DB1/DB2, described previously (Charlier et al., 2002), were identified in the 3′UTR of the NKV (Fig. 3d).

The second model of stable secondary RNA structures, known as the cyclisation model (Fig.3e), has been confirmed experimentally (reviewed in Gritsun et al., 2006; Markoff, 2003). It predicts interaction between complementary sequences located in the 3′UTR and 5′UTR (or in the C gene for the MBFV) and results in the formation of a long dsRNA stem (reviewed in Thurner et al., 2004). Some of these cyclisation sequences, referred to as 3′CYCL and 5′CYCL (Supplementary Fig. 1a) (Hahn et al., 1987), are conserved within the MBFV or TBFV groups and therefore might provide a signaling function in the single or double-stranded form. In this cyclisation model the top part of the 3′LSH folds as an independent structure ("3′LSH", Fig. 3e) implying that linear and cyclisation models overlap, possibly representing two transitional stages in the formation of a complex RNA promoter (Gritsun et al., 2006).

Deletion of the Y-shaped secondary RNA structure from the MBFV group

Initially, alignment was carried out in the region of LRS1 and LRS2 (Supplementary Fig. 1a, page 8) using three conserved sequence anchors; (1) pentanucleotide CACAG, (2) loop 3 (AATTGGC) and (3) loop 4 (TTT) of the Y-shaped structure, shared between the TBFV and the NKV (Charlier et al., 2002). The boundaries of the 3′LSH, SL2 and Y-shaped structure are also specified. To align the MBFV group a large gap was required in the region of the Y-shaped structure indicating that the MBFV probably lost this structure during their evolution. The NCFV shared homology with the TBFV and NKV groups in

this region and RNA folding predictions also demonstrated the presence of a Y-shaped structure in the NCFV (not presented). This conservation, between three ecologically divergent flavivirus groups, indicates that the Y-shaped structure is a vitally important RNA conformation in correspondence with the experimental data using infectious clones (Mandl et al., 1998; Pletnev, 2001).

The alignment also revealed evolution of the "mosquito-borne" CS1, a conserved, but non-repeated element (Supplementary Fig. 1a, page 8). This CS1 region shared homology with the NKV and TBFV and also with a longer equivalent region within the NCFV. Each of the groups, i.e., the MBFV, TBFV and NKV preserved remnants of pre-CS1 that aligned individually with different stretches of the pre-CS1. However, one specific 20-nucleotide deletion was identified in all three groups implying separation of the TBFV/MBFV/NKV lineages from the NCFV lineage.

Notably, the boundaries of the 3′LSH of all flaviviruses virtually coincide with the boundaries between the LRS1 and LRS2 (Supplementary Fig. 1a, page 8), as described for the TBFV group (Gritsun and Gould, 2006a). The boundaries of the Y-shaped structures also almost coincide with the boundary between LRS2/LRS3. The boundaries that determine the difference between viable and nonviable TBFV (Mandl et al., 1998) and DENV (Men et al., 1996) also map in close proximity to the boundary between LRS2/LRS3. These observations imply that different RNA conformations of different flaviviruses might have evolved from duplicated "ancient" LRSs which only the TBFV group has preserved.

Evidence of a common origin for the flavivirus DRs

This was found in the region of the LRS3/LRS4 where the DRs of the MBFV/NKV (i.e., CS2/RCS2) were used as anchors (Supplementary Fig. 1a, pages 6–7 and Fig. 1b). Remarkably, all the short DRs, i.e., CS2/RCS2 of the MBFV and NKV (Charlier et al., 2002; Hahn et al., 1987), the R3 (DR1/DR2) and R2 repeats of the TBFV group (Gritsun et al., 1997; Wallner et al., 1995) and the two NCFV repeats (Crabtree et al., 2003) show homology with the TBFV LRS3 and LRS4. The NCFV repeats, R1/R2, virtually coincide with the TBFV region R3+R2, within the LRS4. These data confirmed our previous predictions that all flavivirus DRs originated from a duplicated ancient LRSs, rather than from a *de novo* duplication in each flavivirus group (Fig. 1b). Preservation of the duplicated region corresponding to LRS3 or LRS4 occurred several times during the evolution of the different flaviviruses. The first preservation presumably occurred in an early flavivirus lineage, which is reflected in the more rigid conservation of LRS3/LRS4 in the TBFV (Gritsun and Gould, 2006a). Subsequently in KRV, the LRS3 was almost completely deleted (Supplementary Fig. 1a, page 7), but the LRS4 was only reduced (Supplementary Fig. 1a, page 6); this was followed by duplication of the entire 3′UTR and further deletion-driven mutagenesis; however, the KRV-R1/R2 were left intact. In CFAV the large deletion of 510 nucleotides shortened the 3′UTR but R1/R2 were again preserved (Gritsun and Gould, 2006b). Finally, YFV strains were shown to have a large duplication that includes the region equivalent to LRS3 (LRS4) (Bryant et al., 2005). The most likely significance of this is the selection of double DB1/DB2 structures, described for NKV and MBFV, each encoded by regions homologous to the LRS3 and LRS4 (Supplementary Fig. 1a, pages 5–6). Although the LRS3 in the TBFV group folds as the second Y-shaped structure, no prediction equivalent to those of the MBFV/NKV has been made for the LRS3/LRS4 region of TBFV and NCFV (Supplementary Fig. 1a, page 7). They may represent an early double Y-shaped/DB-like conformation from which the other extant flavivirus RNA conformations were derived.

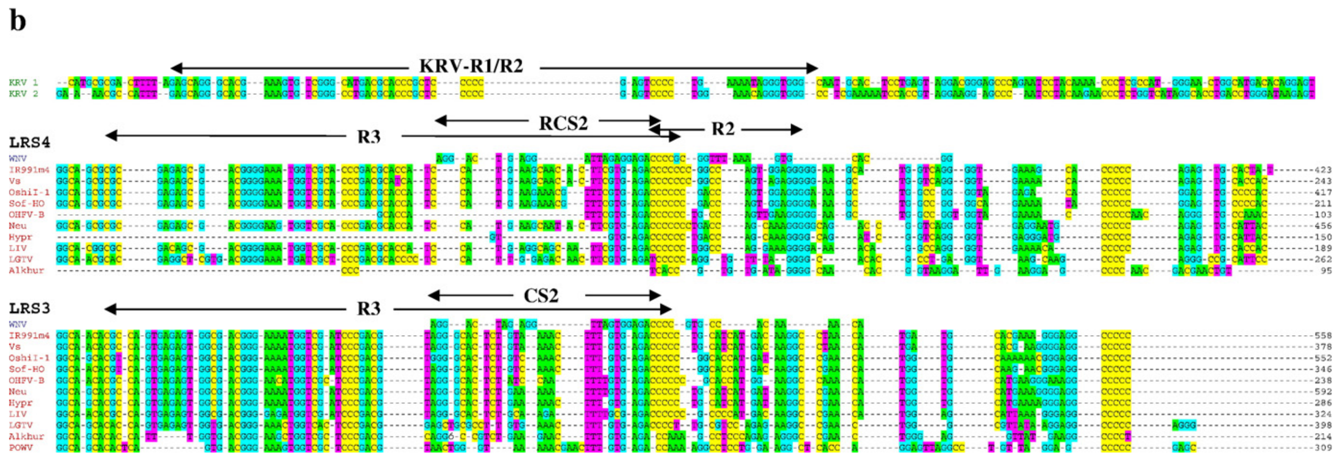

Fig. 1. (a) Full-length 3'UTR alignment. Available as supplementary material (http) to view with magnification 130-150%. Viruses (colored abbreviated names) are specified by their accession numbers and grouped according to the phylogenetic classification: TBFV (red), MBFV (blue), NKV (lilac) and NCFV (green). The DRs and boundaries of RNA secondary structures for each flavivirus group are specified. The boundary between viable/nonviable engineered mutants for TBEV (Mandl et al., 1998; Pletnev, 2001) and DENV (Men et al., 1996) are indicated. To assist observation and interpretation, the alignments for TBFV, NKV and NCFV were placed both above and below the MBFV alignment on each page. Pan-flavivirus alignment of 3′UTR. (b) Abbreviated alignment corresponding to the region LRS3/LRS4. The DRs of MBFV, NCFV and TBFV are indicated. See text for explanations. (c) Schematic presentation (see overleaf) of pan-flavivirus alignment. Each flavivirus group is indicated in colors matching those in Supplementary Fig. 1a. The RNA conformations are shown as in Fig. 3.

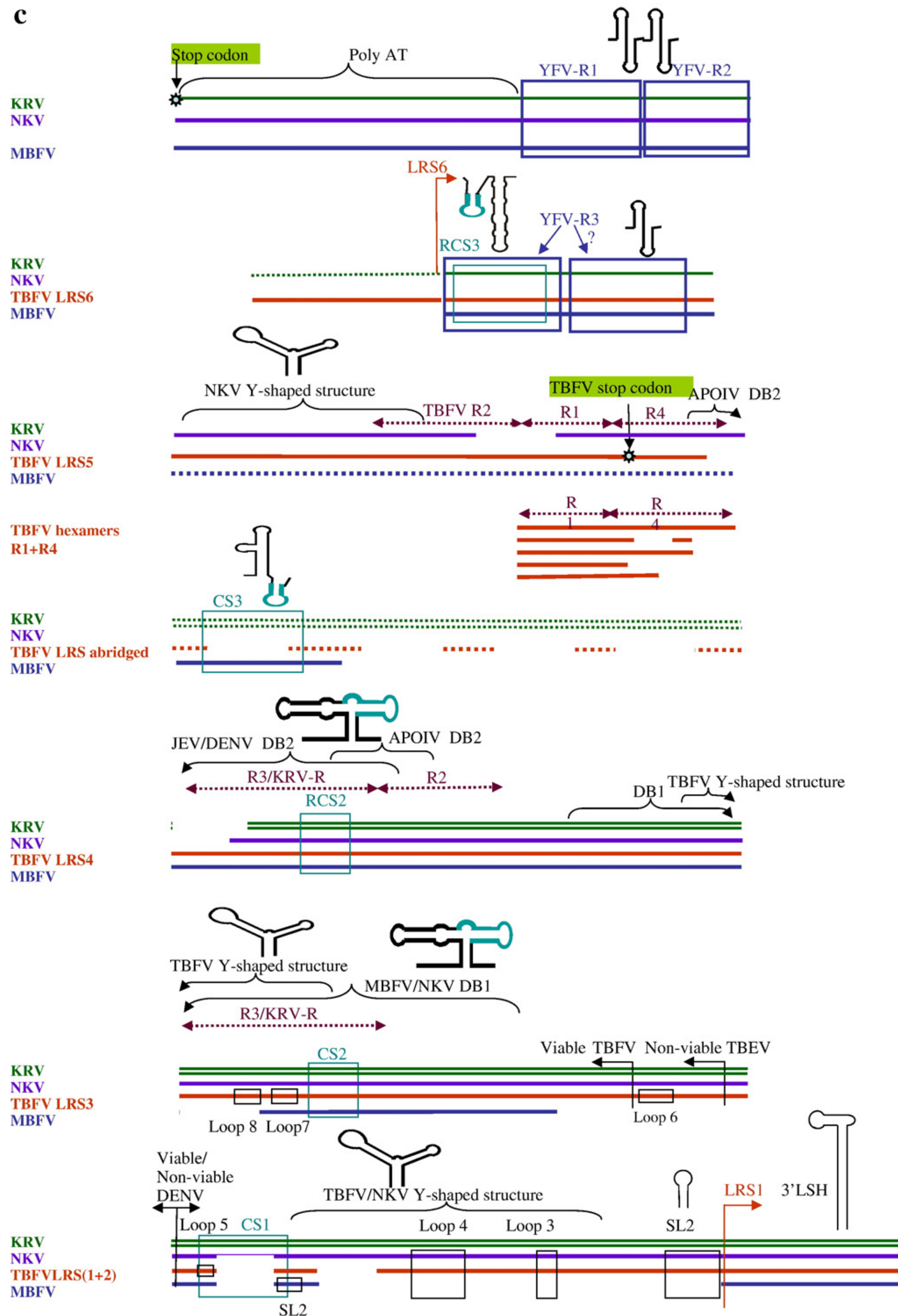

Fig. 1 (*continued*).

The previously "non-aligned" TBFV region originated from an LRS

The region upstream of LRS4 was initially defined as "non-aligned" since no homology was found with any of the TBFV LRSs (Gritsun and Gould, 2006a). Surprisingly, NCFV could be aligned with this region (Supplementary Fig. 1a, page 5), indicating that it was present in the early flavivirus lineage. Perhaps the additional LRS, originally located between LRS4 and LRS5, was reduced extensively by deletions (abridged LRS).

The formation of six R1 + R4 sequences is unique to the TBFV group

The region with six R1 + R4 repeats (Gritsun and Gould, 2006a), located upstream of the abridged LRS, appears to be present only in the TBFV group (Supplementary Fig. 1a, page

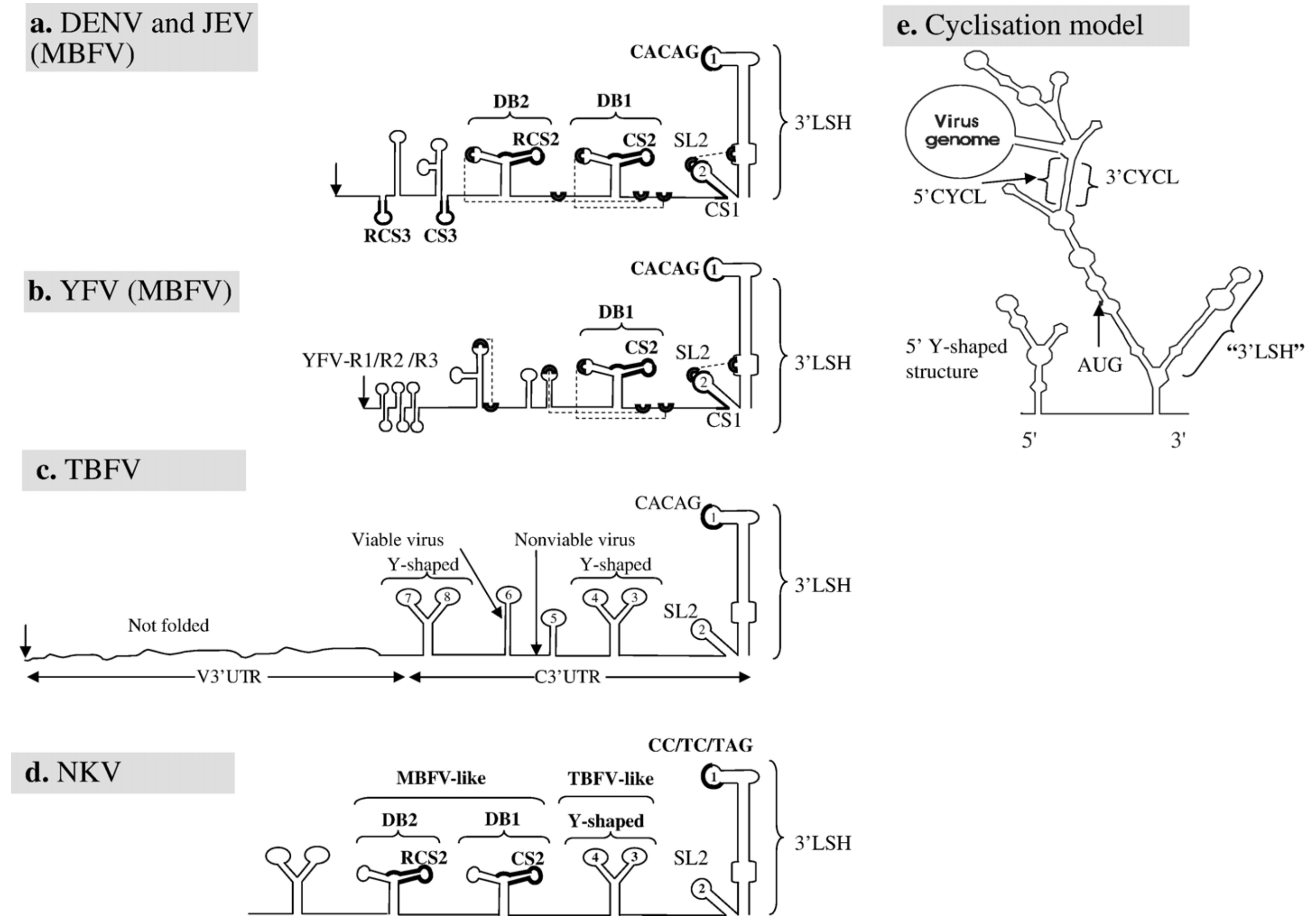

Fig. 3. Linear (a, b, c, d) and cyclisation (e, f) models of RNA structures in the 3′UTR for different flavivirus groups (specified in shadowed boxes). The positions and conformations of the RNA secondary structures are indicated. The CS2/RCS2, CS3/RCS3 and conserved pentanucleotide are highlighted by thick lines. Pseudoknot (thick semi-oval lines) interactions in the MBFV group are shown by dashed lines. The boundaries between viable/nonviable TBFV mutants (c) are indicated. In the cyclisation model (e) the Y-shaped structure at the 5′UTR is folded independently from the cyclisation stem and the part of the side stem-loop that overlaps with the 3′ LSH is specified as "3′LSH".

4) and possibly arose after the TBFV diverged from the other flaviviruses. Viruses that do not contain these sequences are viable, with no reduction in virulence (Mandl et al., 1998). However, their presence in freshly isolated TBFV strains implies that they are essential for TBFV replication in ticks (Hayasaka et al., 2001; Wallner et al., 1995).

The RCS3, but not the CS3, aligns with the TBFV LRS

The region between CS3 and RCS3 (Supplementary Fig. 1a, page 2/3), aligns poorly between all flaviviruses. However, homology was identified between the RCS3 and LRS6; the TBFV group demonstrated higher homology with the JEV group than with the DENV/YFV groups. These regions were used as anchors to identify homology in-between the CS3 and RCS3, i.e., in the region of the LRS5. The alignment of the TBFV ORF (LRS5/LRS6 are parts of the ORF) with the 3′UTR of other flavivirus groups implies that an early 3′UTR lineage was probably longer than the 3′UTR of extant flaviviruses. Thus, originally there were probably 9 LRSs, with 2 complete in the ORF and 6 complete and 1 "abridged" in the 3′UTR of the flavivirus precursor lineage (Fig. 4). Subsequently, as the TBFV group evolved, two LRSs proximal to the stop codon were entirely lost; this region is not essential for virus infectivity in engineered virus mutants (Mandl et al., 1998). Following the linear order of the alignment, CS3 was placed against the "abridged" LRS. Since RCS3 showed homology to LRS6, we placed RCS3, LRS5 and LRS6 (with a reduced number of viruses) into this "unnatural" location (figure brackets on Supplementary Fig. 1a, page 5) to illustrate how the evolution might have occurred in this region. Therefore, both the CS3 and RCS3 could be a part of the TBFV LRSs but the TBFV homologue for CS3 as an "abridged" LRS has been significantly reduced. An additional observation is that CS3/RCS3 and CS2/RCS2 probably represent two different parts of the original LRSs since it is difficult to imagine two identical sequences originating independently of each other on two separate occasions.

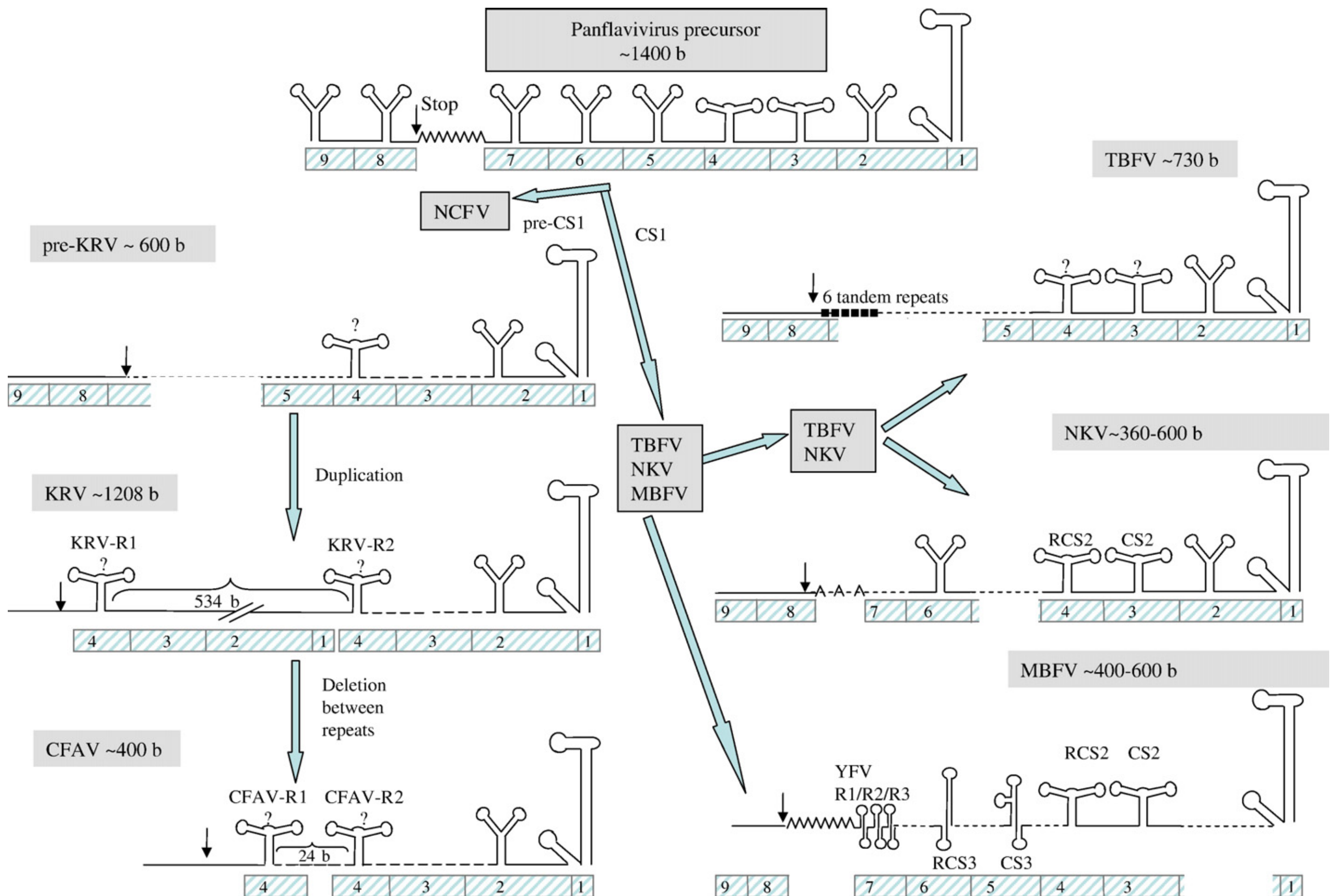

Fig. 4. The possible development of RNA secondary structures in the 3′UTR from the flavivirus precursor lineage. It probably originally contained nine LRSs, with developed 3′LSH-SL2, Y-shaped structures, two DB-like conformations and a poly-AT stretch (zigzag line). The hypothetical structure corresponding to the LRS is depicted as Y-shaped. The "obsolete" regions are depicted by dashed lines. Possible (not predicted) RNA conformations are specified by "?". See text for more explanations.

The poly-AT region appeared early in the evolution of the flaviviruses

Further alignment in the 3′ → 5′ direction became increasingly difficult due to the high divergence even between closely related viruses (Wallner et al., 1995). Therefore, we anchored the stop codons of the MBFV, NKV and NCFV groups and generated the alignment in the 5′ → 3′ direction towards the already aligned sequences (Supplementary Fig. 1a, page 1). As demonstrated previously for the MBFV (Gritsun and Gould, 2006c), there is a variable poly-AT-enriched region that presumably developed as the result of multiple stuttering/duplications of viral RNA polymerase around the TAA stop codon. Although DENV and YFV do not have a poly-AT, it was possibly present in early MBFV lineages since YOKV, a member of the YFV group has a poly-AT. The presence of remnants of poly-AT in the distantly related NKV and NCFV (Supplementary Fig. 1a, page 1) indicates that poly-AT might have been present in early flavivirus lineages. The preservation of poly-AT only within the JEV group probably reflects a selective advantage for virus circulation in the specific ecological niches of these viruses. In molecular terms, AT enrichment might promote the formation of unstable secondary structures acting as spacers to ensure independent folding of the 3′UTR (Gritsun and Gould, 2006c). The TCC-enriched region located in an equivalent part of KRV and TBV could also serve the same purpose (Supplementary Fig. 1a, page 1).

The three YFV repeat sequences are also present in the NCFV group

Three 44-nucleotide tandem repeats YFV-R1, -R2 and -R3, located next to the stop codon, were assigned only to the YFV group and originally were discovered in the West-African (WA) strains of YFV (Hahn et al., 1987). One to 3 copies of these repeats were also found in the YFV Central/East-African (CE) and South American strains (SA) and also in the YFV-related flaviviruses, Sepik (SEPV), Uganda S (UGSV) and Banzi (BANV) (Mutebi et al., 2004) (Supplementary Fig. 1a). Subsequently, in two other MBFV groups, i.e., JEV and DENV, sequences were identified that were interpreted as remnants of YFV tandem repeats possibly presented in the MBFV precursor lineage (Gritsun and Gould, 2006c). We have now investigated the possible presence of YFV repeats in other flavivirus groups. Indeed some residual homology was

identified between YFV-R1/R2 and NCFV (Supplementary Fig. 1a, page 1). The short A(TT)A region between YFV-R2 and YFV-R1 was also similar between the YFV and NCFV groups. The region between the YFV-R2 and YFV-R3 (Supplementary Fig. 1a, page 2) contains a long insertion (Gritsun and Gould, 2006c) that also shows some homology with the NCFV and also with the TBFV ORF region upstream of the LRS6. In this region the alignment between the flavivirus groups is not as robust as in other regions and as discussed previously small duplications, CTGT and TAGG (underlined in Supplementary Fig. 1a, page 2) contributed to its formation (Gritsun and Gould, 2006c).

The assigned position of YFV-R3 is tenuous and for the JEV- and DENV-related groups there are two possibilities; one overlaps with RCS3 and the other is downstream of this position (Supplementary Fig. 1a, page 3; figure brackets specify the possible second, "unnatural" position of YFV-R3; the gap is the "natural" position in this region). It is quite possible that more copies of YFV repeats were present in earlier lineages and this could explain the difficulty of identifying the YFV-R3 position. The significant variability of this region, even between closely related MBFV, leaves us with doubts that YFV repeats were present in the flavivirus precursor lineage. The isolation of other "exotic" flaviviruses could shed more light on this problem.

Evolution of flavivirus DRs and secondary RNA structures in the 3′UTR

Fig. 1c presents a schematic version of the pan-flavivirus alignment and Fig. 4 provides a flowchart of the predicted evolutionary events leading to the extant 3′UTRs. The earliest flavivirus lineages probably had either no 3′UTR or they had a very short 3′UTR similar to the 3′LSH that also mediated cyclisation of the virus genome. The function of the 3′UTR appears to have overlapped at least partially with the function of the ORF. Multiple duplications of the terminal region of the ORF LRS led to the enlargement of the 3′UTR; the advantage for the virus was the capability for independent specialization of the ORF and 3′UTR (Gritsun and Gould, 2006a). The homology of three distinct flavivirus groups (MBFV, NKV, NCFV) with the TBFV group beyond the stop codon indicates that duplication occurred at least 8 times thus producing 9 LRSs. Six recognizable LRSs were preserved only in the TBFV, with two in the ORF and four in the 3′UTR; in the other flavivirus groups mutagenesis was more extensive, leaving only short DRs, derived from LRSs, as indicators of commonality. The higher level of preservation of LRSs in the TBFV group probably reflects the lower rate of evolution of these viruses as described previously (Gritsun and Gould, 2006a).

In relation to RNA secondary structures, it is possible that the early lineages had a fully developed 3′LSH-SL2 region, Y-shaped structure (found in TBFV and NKV), two DB-like structures (found in MBFV and NKV) and a poly-AT (found in the MBFV, NKV and NCFV). The deletion of some linear and conformational features probably reflects the diversification of the flaviviruses into the four groups, i.e., TBFV, MBFV, NKV and NCFV. The formation of CS1 from pre-CS1 could have taken place during diversification of the NCFV and the loss of the Y-shaped structure in the MBFV probably occurred with formation of the precursor MBFV lineage. It is not clear if the DB-like structures are still present in the TBFV group since no equivalent folding predictions have been made for the LRS3/LRS4; probably the DB1/DB2 evolved from the two tentative, ubiquitous, Y-shaped structures encoded by an ancient LRS lineage. In the case of the NCFV, reduction by deletion and subsequent duplication of the entire 3′UTR occurred after these viruses diverged from the TBFV/MBFV/NKV precursor lineage.

Previously we proposed that the flavivirus 3′UTR might be subdivided into promoter and enhancer (Gritsun and Gould, 2006c; Gritsun et al., 2006). The promoter comprises a region that is essential for RNA replication and virus viability; as established experimentally for the TBFV group it corresponds roughly to the 3′LSH-SL2-Y-shaped structure and in the MBFV group to the 3′LSH-SL2-DB1 complex. The region between the stop codon and promoter was defined as an enhancer that is not essential for viability of laboratory-maintained viruses. For the MBFV we suggested that DRs are essential elements of the enhancer that might interact with cellular/virus proteins to maximize the rate of virus RNA synthesis required for efficient virus transmission (Gritsun and Gould, 2006c). Extending this analogy for the entire genus we now propose that evolution of the 3′UTR was associated with diversification of the promoter/enhancer region as the means of virus adaptation to different hosts (Fig. 4). The reason for preservation of the DRs as a double signal is still not clear; they are possibly recognized by proteins, the activity of which depends on dimerization. The DRs might represent RNA recognition signals interacting with cellular/viral proteins that traffic flavivirus RNA to the appropriate cellular compartments to accelerate the assembly of the virus replicase complex (Gritsun and Gould, 2006c). Although the enhancer function may be non-essential for virus viability under experimental conditions, it could play a significant role in the more demanding natural environment where high rates of virus replication could be critical for virus transmission between hosts thus ensuring their success and strategy for survival.

In summary, we have constructed a 3′UTR pan-flavivirus alignment that has revealed conserved and variable elements of promoters and enhancers that probably formed early in the evolution of the flaviviruses. We propose that the ancestral 3′UTR was formed by multiple duplications of a 200-nucleotide region of the ORF, 6 mutated copies of which (LRSs) were preserved in the genome of the slowly evolving TBFV group. More extensive evolution in the 3′UTR of the MBFV, NKV and NCFV groups resulted in the gradual reduction of LRSs to smaller regions that are now observed as short DRs. The early LRSs might encode the RNA conformation, from which a variety of other conformations, distinct for each flavivirus group, subsequently evolved. We suggest that the DRs, functioning as enhancers of RNA replication, provide the strategy for flavivirus survival in a variety of hosts. Finally, our pan-flavivirus alignment is a valuable tool with which to predict conserved functionally important elements within the UTRs of

flaviviruses and to design genetic experiments under laboratory conditions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.virol.2006.09.033.

References

Bryant, J.E., Vasconcelos, P.F., Rijnbrand, R.C., Mutebi, J.P., Higgs, S., Barrett, A.D., 2005. Size heterogeneity in the 3′ noncoding region of South American isolates of yellow fever virus. J. Virol. 79 (6), 3807–3821.

Charlier, N., Leyssen, P., Pleij, C.W.A., Lemey, P., Billoir, F., Van Laethem, K., Vandamme, A.M., De Clercq, E., de Lamballerie, X., Neyts, J., 2002. Complete genome sequence of Montana Myotis leukoencephalitis virus, phylogenetic analysis and comparative study of the 3′ untranslated region of flaviviruses with no known vector. J. Gen. Virol. 83 (Pt. 8), 1875–1885.

Crabtree, M.B., Sang, R.C., Stollar, V., Dunster, L.M., Miller, B.R., 2003. Genetic and phenotypic characterization of the newly described insect flavivirus, Kamiti River virus. Arch. Virol. 148 (6), 1095–1118.

de Lamballerie, X., Crochu, S., Billoir, F., Neyts, J., de Micco, P., Holmes, E.C., Gould, E.A., 2002. Genome sequence analysis of Tamana bat virus and its relationship with the genus *Flavivirus*. J. Gen. Virol. 83 (Pt. 10), 2443–2454.

Gritsun, T.S., Gould, E.A., 2006a. The 3′ untranslated region of tick-borne flaviviruses originated by the duplication of long repeat sequences within the open reading frame. Virology 350 (2), 269–275.

Gritsun, T.S., Gould, E.A., 2006b. The 3′untranslated region of Kamiti River virus and Cell fusing agent virus originated by self-duplication. J. Gen. Virol. 87, 2615–2619.

Gritsun, T.S., Gould, E.A., 2006c. Direct repeats in the 3′ untranslated regions of mosquito-borne flaviviruses: possible implications for virus transmission. J. Gen. Virol. 87 (11), 3297–3305.

Gritsun, T.S., Venugopal, K., Zanotto, P.M., Mikhailov, M.V., Sall, A.A., Holmes, E.C., Polkinghorne, I., Frolova, T.V., Pogodina, V.V., Lashkevich, V.A., Gould, E.A., 1997. Complete sequence of two tick-borne flaviviruses isolated from Siberia and the UK: analysis and significance of the 5′ and 3′-UTRs. Virus Res. 49 (1), 27–39.

Gritsun, T.S., Tuplin, A.K., Gould, E.A., 2006. Origin, evolution and function of flavivirus RNA in untranslated and coding regions: implications for virus transmission. In “Flaviviridae: Pathogenesis, Molecular Biology and Genetics” (M. Kalitzky, and P. Borowski Eds.) Horizon Scientific Press, Norwich, UK, pp. 47–99.

Hahn, C.S., Hahn, Y.S., Rice, C.M., Lee, E., Dalgarno, L., Strauss, E.G., Strauss, J.H., 1987. Conserved elements in the 3′ untranslated region of flavivirus RNAs and potential cyclisation sequences. J. Mol. Biol. 198 (1), 33–41.

Hayasaka, D., Ivanov, L., Leonova, G.N., Goto, A., Yoshii, K., Mizutani, T., Kariwa, H., Takashima, I., 2001. Distribution and characterization of tick-borne encephalitis viruses from Siberia and far-eastern Asia. J. Gen. Virol. 82 (Pt. 6), 1319–1328.

Heinz, F.X., Collett, M.S., Purcell, R.H., Gould, E.A., Howard, C.R., Houghton, M., Moormann, R.J.M., Rice, C.M., Thiel, H.J., 2000. Family Flaviviridae. In: Fauquet, C.M., Bishop, D.H.L., Carstens, E., Estes, M.K., Lemon, S., Maniloff, J., Mayo, M.A., McGeogch, D., Pringle, C.R., Wickner, R.B. (Eds.), Virus Taxonomy. 7th International committee for the Taxonomy of Viruses. Academic Press, San Diego, pp. 859–878.

Mandl, C.W., Holzmann, H., Meixner, T., Rauscher, S., Stadler, P.F., Allison, S.L., Heinz, F.X., 1998. Spontaneous and engineered deletions in the 3′ noncoding region of tick-borne encephalitis virus: construction of highly attenuated mutants of a flavivirus. J. Virol. 72 (3), 2132–2140.

Markoff, L., 2003. 5′- and 3′-noncoding regions in flavivirus RNA. Adv. Virus Res. 59, 177–228.

Men, R., Bray, M., Clark, D., Chanock, R.M., Lai, C.J., 1996. Dengue type 4 virus mutants containing deletions in the 3′ noncoding region of the RNA genome: analysis of growth restriction in cell culture and altered viremia pattern and immunogenicity in rhesus monkeys. J. Virol. 70 (6), 3930–3937.

Mutebi, J.P., Rijnbrand, R.C., Wang, H., Ryman, K.D., Wang, E., Fulop, L.D., Titball, R., Barrett, A.D., 2004. 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.

Pletnev, A.G., 2001. Infectious cDNA clone of attenuated Langat tick-borne flavivirus (strain E5) and a 3′ deletion mutant constructed from it exhibits decreased neuroinvasiveness in immunodeficient mice. Virology 282 (2), 288–300.

Thurner, C., Witwer, C., Hofacker, I.L., Stadler, P.F., 2004. Conserved RNA secondary structures in Flaviviridae genomes. J. Gen. Virol. 85 (Pt. 5), 1113–1124.

Wallner, G., Mandl, C.W., Kunz, C., Heinz, F.X., 1995. The flavivirus 3′-noncoding region: extensive size heterogeneity independent of evolutionary relationships among strains of tick-borne encephalitis virus. Virology 213 (1), 169–178.