CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Elsevier - Publisher Connector

Available online at www.sciencedirect.com
ScienceDirect
Virology 366 (2007) 8–15
VIROLOGY
www.elsevier.com/locate/yviro

Rapid Communication

Origin and evolution of flavivirus 5′UTRs and panhandles: Trans-terminal duplications?

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

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

Received 27 February 2007; returned to author for revision 3 April 2007; accepted 9 April 2007
Available online 20 July 2007

Abstract

Flavivirus replication is mediated by interactions between complementary ssRNA sequences of the 5′- and 3′-termini that form dsRNA cyclisation stems or panhandles, varying in length, sequence and specific location in the mosquito-borne, tick-borne, non-vectored and non-classified flaviviruses. In this manuscript we manually aligned the flavivirus 5′UTRs and adjacent capsid genes and revealed significantly more homology than has hitherto been identified. Analysis of the alignments revealed that the panhandles represent evolutionary remnants of a long cyclisation domain that probably emerged through duplication of one of the UTR termini.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Flaviviruses; Alignment; 5′UTR; Promoter; Enhancer; Panhandle; Cyclisation

Introduction

Viruses in the genus *Flavivirus* (family *Flaviviridae*) are subdivided into three groups—mosquito-borne, tick-borne and no-known vector flaviviruses (MBFV, TBFV and NKV) respectively. Cell fusion agent virus (CFAV), Culex flavivirus (CuFV) (Hoshino et al., 2006), Kamiti River virus (KRV) and Tamana bat virus (TBV) are tentative members of the genus and will be referred to as non-classified flaviviruses (NCFV) (de Lamballerie et al., 2002). Virions (∼50 nm) contain a core formed by a capsid (C) protein and positive-sense ssRNA (∼11 kb) which is enclosed in a cell-derived membrane containing envelope (E) and membrane (M) glycoproteins. A single open reading frame (ORF) encodes a polyprotein of about 3400 amino acids which is processed into the structural (C, M and E) and nonstructural (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5) proteins. The ORF is flanked by 5′ and 3′ untranslated regions (UTRs), that directly interact to circularise virus genomes with the formation of short double stranded regions, i.e. panhandles, the compulsory prerequisite for initiation of replication (reviewed in Gritsun et al., 2006).

Previously we constructed a pan-flavivirus alignment for the 3′ UTRs and correlated it with predicted RNA secondary structures and experimental data. This approach demonstrated that evolution of the 3′UTR was based on multiple duplications of the terminal parts of the viral genome and the subsequent formation of extended RNA domains evolved as promoters and enhancers of virus replication determined by the selective requirements of the vertebrate and invertebrate hosts (Gritsun and Gould, 2007a). In this paper we have applied a similar approach to trace the origin and mode of evolution of flaviviruses in relation to the 5′UTR and cyclisation (panhandle) domain.

Results and discussion

Construction of a 5′UTR pan-flavivirus alignment

A 5′UTR pan-flavivirus alignment (Supplementary Fig. 1a) was initially generated by manually aligning the alignments produced previously for MBFV (YFV, JEV and DENV) and TBFV (TBEV, POW, LGTV, LIV, OHFV KFDV) (Gritsun et al., 1997, 2006). Subsequently the alignments produced for NKV (MODV, RBV, APOIV and MMLV) and NCFV (KRV, CFAV and TBV) were added to the alignment (Supplementary Fig. 1b).

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

0042-6822/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2007.04.011

The C gene region of ~240 bases, with the corresponding amino acid sequences was also included. The sequences of the predicted RNA secondary structures (Fig. 2) were then mapped onto the alignments (Supplementary Fig. 1b and Fig. 2d).

RNA secondary structures

Fig. 2 summarises the flavivirus RNA cyclisation models reviewed in (Gritsun et al., 2006) and the models for previously unstudied flaviviruses produced using MFold (http://www.bioinfo.rpi.edu/~zukerm/) (available on request). All flaviviruses formed dsRNA regions, known in other viruses as panhandles, the result of direct interaction between complementary sequences in the 5′UTR and 3′UTR. In addition, the proximal regions of the 5′UTRs of all 4 flavivirus groups formed a second homologous, Y-shaped structure that for most flaviviruses folded independently of the panhandle (Fig. 2). Experiments have shown that the Y-shaped structure is required for the initiation of RNA replication from the 3′end, in compliance with recognised mechanisms for other RNA viruses (Filomatori et al., 2006; Gritsun et al., 2006). Although the Y-shaped conformation is also believed to regulate flavivirus translation, its conservation amongst flaviviruses is probably largely attributed to its role in virus replication. Three elements of the 5′UTR Y-shaped structure have been defined, i.e. stem 0, stem-loops 1 and 2 (SL1 and SL2). The length of stem 0 varies between 8 and 13 bp and this is one of the most conserved elements within the 5′UTR of the flaviviruses. A similar but not identical Y-shaped-like structure is present in the 5′UTR of West Nile virus (WNV) and Kunjin virus (KUNV) for which a bulging region in the middle of the longer stem 0 could represent an extensively degraded SL2, with the complete loss of nucleotide complementarity in this region. Engineered deletions within SL2 in DENV did not result in loss of infectivity, correlating with the significant variability of this structure in flaviviruses; the MBFV SL2 is much shorter than the TBFV SL2 and among the TBFV it has variable lengths (reviewed in Gritsun et al., 2006). A 5′Y-shaped-like structure (with a bulge instead of an SL2) was predicted for KRV, but not for other NCFV for which the corresponding region of the 5′UTR was involved in the formation of a panhandle. However, at suboptimal free energy values, a Y-shaped-like structure was predicted for all NCFV, implying that NCFV might share RNA structural organisation with other flaviviruses.

The formation of panhandles results in complete or partial alteration of RNA structures predicted for independently folded 3′UTRs (Fig. 2) and interpreted as promoters/enhancers of flavivirus replication (Gritsun et al., 2006). The only RNA structure that retains its predicted folding during the formation of the panhandle, is the 3′LSH (long stable hairpin; formed by the terminal ~100 nucleotides) that was identified as a virus promoter. In DENV and NCFV (Fig. 2), the 3′LSH was preserved completely during panhandle formation, but in TBFV, YFV and JEV conservation was only partial with the lower part merging into the panhandle structure. The sequences involved in formation of the lower part of the 3′LSH may function as a single-stranded rather than a duplex region (Kofler et al., 2006). However in the NKV group, creation of the panhandle completely abolished the 3′LSH (Charlier et al., 2002) implying that panhandle, rather than 3′LSH formation might be the dominating event at this stage of formation of the replication promoter. This possibly reflects adaptation of NKV replication to single mammalian hosts in contrast with the double, i.e. vertebrate and invertebrate hosts of the TBFV and MBFV.

Evolution of the flavivirus 5′UTR and panhandle sequences

Multiple deletions and duplications are largely responsible for the evolution of flavivirus 3′UTRs (Gould and Gritsun, 2006; Gritsun and Gould, 2006b; Gritsun et al., 2006). Similar molecular perturbations of flavivirus genomes were also detected in the 5′UTR; a series of short sequences with high homology in the four flavivirus groups, were separated by gaps; duplications also contributed to genome variability (Supplementary Figs. 1a, b). Sequences involved in stem 0 formation were the most conserved in all flavivirus groups. Some SL2 sequences were partially preserved between only the TBFV and JEV groups. However JEV, DENV and YFV had deletions resulting in overall shortening of the MBFV SL2.

As RNA models have demonstrated, all flaviviruses form panhandles, but their sequences, specific location and lengths vary between the different groups (Supplementary Fig. 1b). The panhandle of the TBFV group (5′CYCL) is formed by a perfectly conserved continuous 21-nucleotide sequence located in the 5′UTR. The equivalent highly conserved 5′CYCL of the MBFV is limited to an 8-nucleotide region located in the C gene although the entire panhandles for each of the three MBFV subgroups are much longer and formed by variable sequences (Supplementary Fig. 1b and Fig. 1d). For the NKV and NCFV groups their 5′CYCL-equivalent, i.e. conserved regions have not yet been located.

Overall, MBFV, NKV and NCFV panhandles are longer than those of the TBFV group consisting of discontinuous dsRNA regions within the 5′UTR and C gene. However the alignments reveal homology between all four flavivirus groups, the most striking being between the 5′CYCL of the TBFV group and a corresponding region for YFV (Supplementary Fig. 1b). Remnants of the MBFV 5′CYCL are also detectable in MODV (NKV) and the TBFV C gene immediately following the introduced large gap located in the region of high nucleotide variability between the MBFV, NKV and NCFV groups (Supplementary Fig. 1b). One would normally predict that amino acid substitutions rather than deletions would be occurring in this region since it is the ORF, but the presented data imply that the first half of the C protein is under lower selection pressure than the second half. As an example we aligned the nucleotides and the amino acids of CFAV, CuFV and KRV (Fig. 3). The level of identity for the first 64 amino acids of the C protein between CFAV/CuFV was 26% whereas for the last 72 amino acids it was 62%. The corresponding nucleotide alignment showed that evolution within the first half was largely mediated by deletions and multiple frame shifts rather than single amino acid substitutions. Similarly, the corresponding

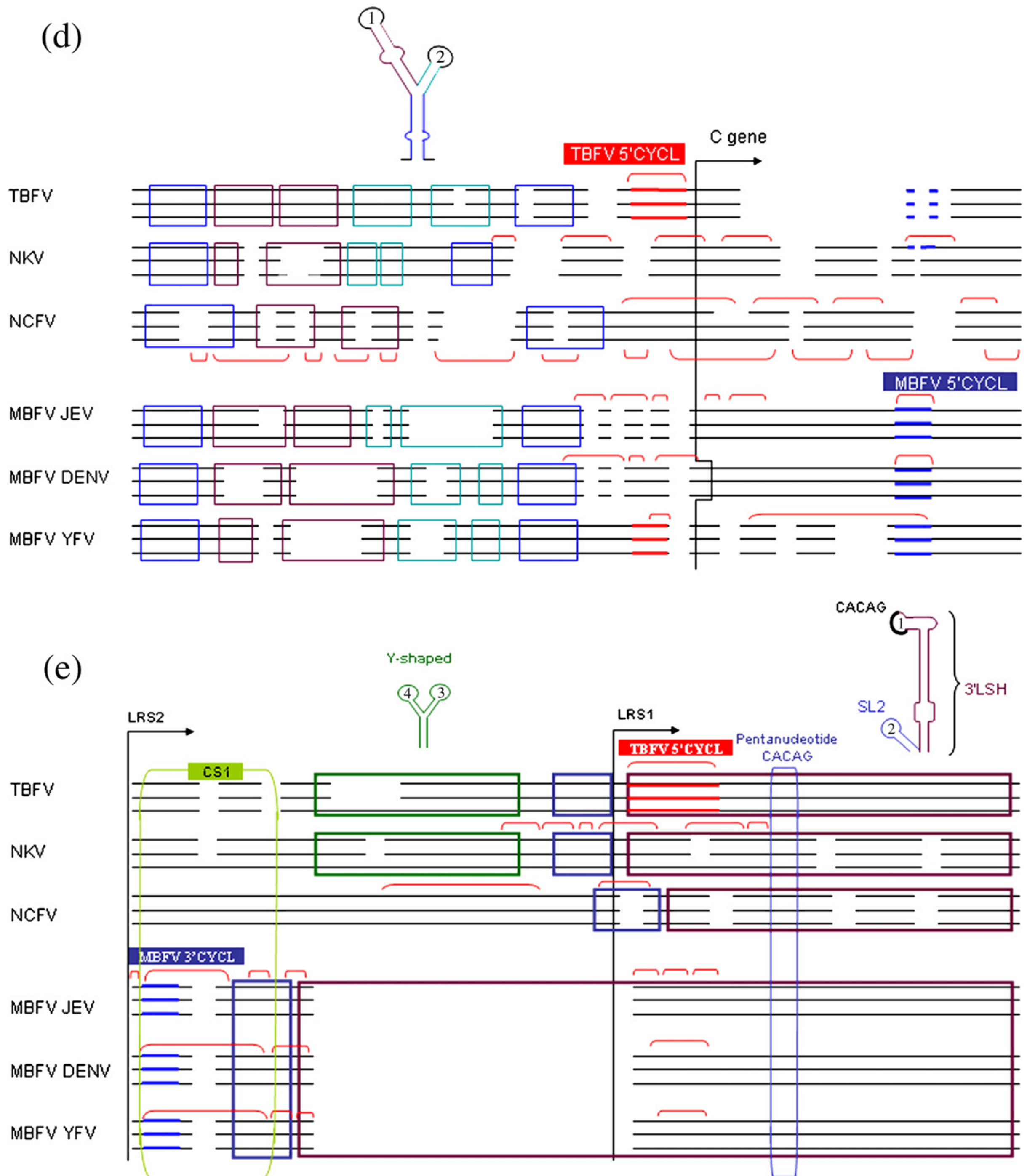

Fig. 1. Mapping of flavivirus cyclisation sequences. Alignments of the 5′UTR, with a fragment of the adjacent C gene (a,b), and 3′UTR (c) are available as supplementary material (doi:10.1016/j.virol.2007.04.011) to view with magnification 130–150%; their schematic image is depicted in D and E respectively. Viruses are identified by their accession numbers and grouped according to the classification scheme defined in the text of the manuscript: TBFV, MBFV, NKV and NCFV. Cyclisation sequences are indicated by red brackets. (a) Alignment between TBFV and MBFV 5′UTRs with a fragment of adjacent C gene. Duplicated regions, imperfect repeats, 5′CYCL, 3′CYCL and SL6 (see text) are indicated. The alignment for the TBFV is placed above and below MBFV alignment. (b) Pan-flavivirus 5′UTR alignment, with a fragment of adjacent C gene. Double stranded regions of stem 0, SL1 and SL2 (underlined for individual viruses) are enclosed in colored boxes (blue, brown and green correspondingly). The double-stranded regions of the SL6 are enclosed in black-line boxes. The conserved hexanucleotide TGCCAA in loop 6 (presented as an inverted repeat in the 3′UTR) is underlined. Amino acids in the ORF are indicated above the uppermost sequence. Amino acids used as anchors in the alignment are highlighted in dark green. Duplicated protein Pro-Gly-enriched sequences (see Fig. 3) are highlighted in grey. The sequence within the capsid gene that could be deleted without total loss of virus viability (Kofler et al., 2002) is indicated. (c) Pan-flavivirus alignment of 3′UTR corresponding to the LRS1/2-3-4 (pages 3-2-1 respectively) with the specified boundaries of RNA secondary structures (3′LSH-SL2, Y-shaped structures, DB1) for each flavivirus group (Gritsun and Gould, 2007a). Linear conserved elements CS1, CS2, RCS2 and pentanucleotide CACAG are shown. (d) Schematic presentation of pan-flavivirus alignment for the 5′UTR/C gene as detailed in Fig. 1b. The Y-shaped structure is depicted as in Fig. 2; double-stranded regions of stem 0, SL1 and SL2 are boxed in blue, brown and green respectively. The TBFV 5′CYCL and its remnants in YFV are depicted as red lines; the MBFV 5′CYCL and its remnants in TBFV and NKV groups is depicted in blue lines. (e) Schematic presentation of pan-flavivirus alignment for the 3′UTR as detailed in Fig. 1c. The positions of flavivirus conserved features, i.e. 3′CYCL for MBFV (thick blue lines) and TBFV (thick red lines), CS1 and pentanucleotide CACAG are indicated.

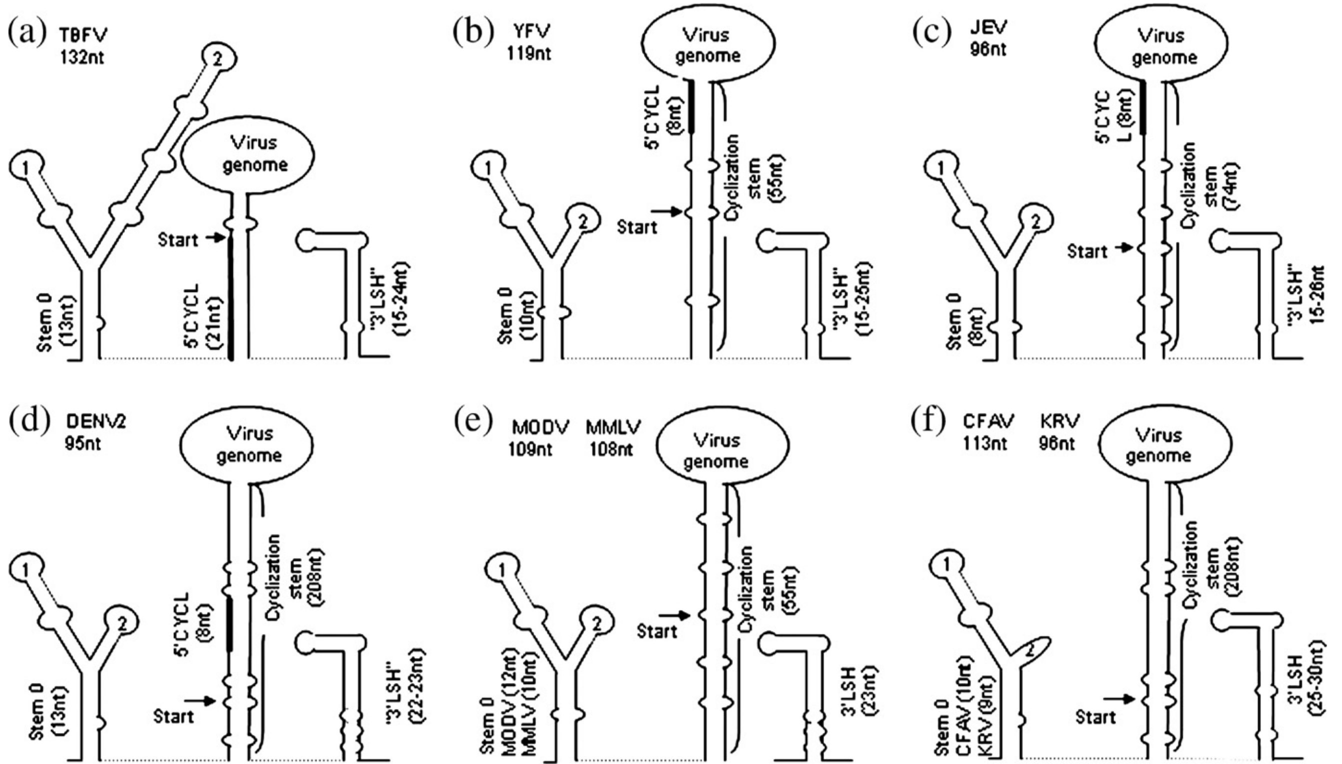

Fig. 2. Comparison of RNA secondary structures between TBFV (a), MBFV (b, c, d), NKV (e) and NCFV (f). The lengths of the 5′UTR are specified in the top-left corner. The major elements of the Y-shaped structure (stem 0, SL1 and Sl2) and 3′LSH, the length of the panhandle and position of start codon are indicated.

region of YFV contains CAAA-like-sequences duplicated 5 times (Supplementary Fig. 1b) that probably replaced the deleted region homologous to other MBFV. There are also imperfect repeats of Pro-Gly-enriched peptides in the hypervariable region of the C protein of the NCFV and also in the NKV and MBFV groups (Supplementary Fig. 1b and Fig. 3). An additional example of structural flexibility within the C gene is the region of conserved stable SL6 structure (Fig. 1) predicted

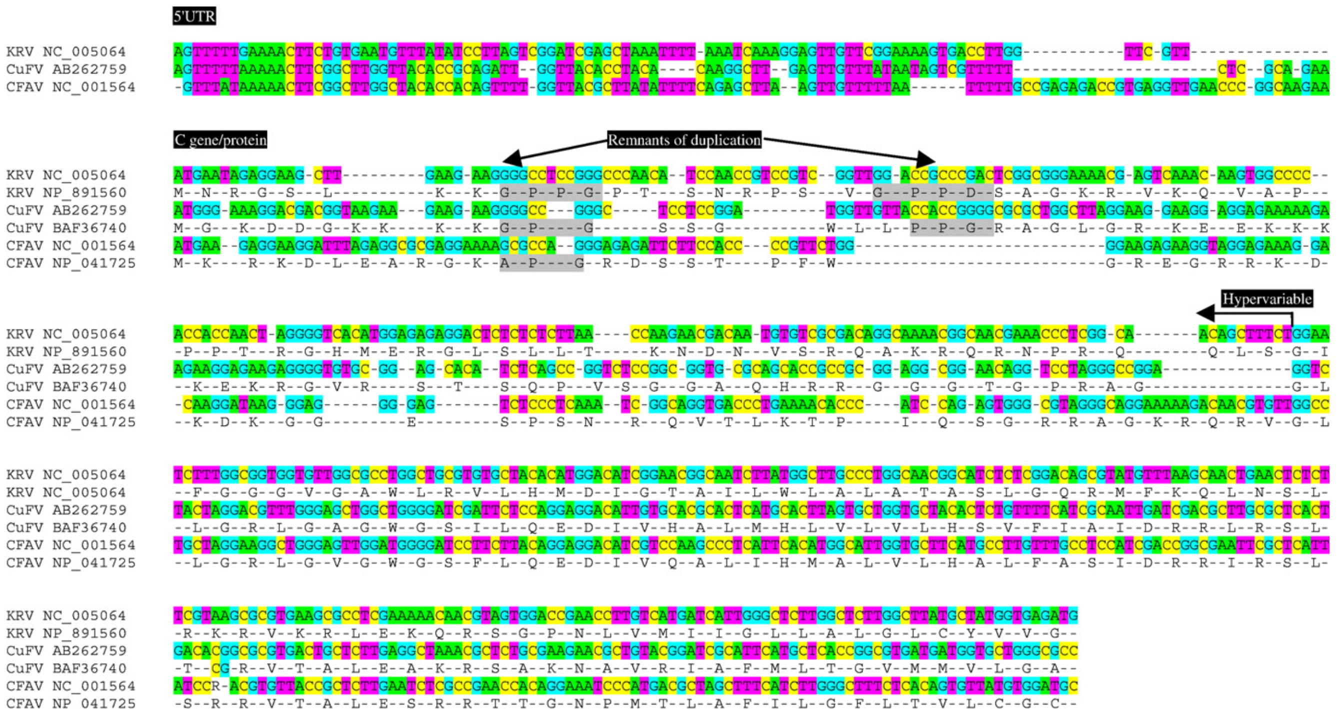

Fig. 3. Alignment of the nucleotide and amino acid 5′UTR-C-gene regions between NCFV. The remnants of possible duplicated amino acid PG-enriched regions are highlighted in grey. The boundary between the hypervariable and more conserved region within C protein is identified.

for the TBFV in the C gene (Gritsun et al., 1997) that was formed by duplication of a CCAATG sequence (Supplementary Fig. 1a). The inverted complementary copy of the TBFV-conserved sequence TGCCAA in loop 6 is also present in loop 4 of the 3′UTR revealing kissing loops (Gritsun et al., 2006). Remarkably, specific deletion in this region (Supplementary Fig. 1b) did not abolish virus infectivity (Kofler et al., 2002), implying that this region can tolerate significant nucleotide and amino acid alterations. Overall, we conclude that the first half of the C gene, being flexible as a protein, allows this region to function also as an RNA structural domain, with promoter and enhancer activities and might correspond to the disordered domains in capsid proteins of other viruses that are associated with RNA chaperone activity (Mir and Panganiban, 2006).

Previously we developed a pan-flavivirus alignment of the 3′ UTR in which the highly conserved 3′CYCLs of the MBFV and TBFV groups mapped to non-overlapping regions (Gritsun and Gould, 2007a). However when all double-stranded RNA cyclisation sequences were superimposed onto this alignment, other regions of homology were revealed (Supplementary Fig. 1c and Fig. 1e).

Thus during early evolution of the genus *Flavivirus* the cyclisation region was probably much longer and only the relatively slowly evolving TBFV group has preserved it as a continuous sequence. In other flavivirus groups, the original long continuous panhandle probably evolved further leaving discontinuous double-stranded RNA regions of various lengths and locations.

Alignment between flavivirus 5′ and 3′UTRs

To support this hypothesis we aligned the 5′UTR/C gene region with the 3′UTR of TBEV, the rationale being that the relatively slowly evolving TBFV might have retained the remnants of earlier flavivirus lineages (Gritsun and Gould, 2006a). As we demonstrated, the formation of the TBFV 3′UTR was mediated by multiple duplication of sequences approximately 200 nucleotides in length that were subsequently progressively reduced during evolution. Accordingly, we subdivided the 3′UTR of the TBFV into the four regions, known as long repeated sequences (LRS1, 2, 3 and 4; Supplementary Fig. 1c) that were homologous with each other (Gritsun and Gould, 2006a). Fig. 4a (supplement) illustrates the alignment between the most terminal TBFV LRS1–LRS2 region (inverted complement) with the TBFV/MBFV 5′UTR/C gene region. The 5′CYCL and loop 6 sequences were used as anchors to reveal more homology in regions outside the cyclisation zone. A series of homologous regions separated by gaps was identified indicating again that the initial cyclisation region was probably longer than is detected in current viruses.

Remarkably, the length of the LRS1 is virtually identical to that of the Y-shaped structure in the 5′UTR (Supplementary Fig. 4a). It is possible that the 5′UTR originated by template switching of the replicative complex (RC) from the 3′ end of the (+)strand RNA template, with incomplete nascent (−)strand, onto the 3′ end of the neighboring (−)strand of dsRNA template (Replicative Form, or RF). Alternatively template-switching might occur when the RC with a nascent (+) strand is translocated from the (−)strand of the RF template to the 3′ end of the (+) template to initiate the synthesis of the (−)strand according to recognised mechanisms (Lazzarini et al., 1981). The subsequent elongation would result in formation of RNA with long self-annealed ends. This might be advantageous for the virus since panhandles are present in many RNA viruses (see below). The dsRNA extended region evolved further until the optimal transient structure was derived, in terms of sequence and free energy, to be unwound by the RC to proceed to the subsequent steps of viral RNA synthesis. For the TBFV group, the 5′UTR and 3′UTR sequences directly involved in cyclisation are located after the 5′ Y-shaped structure and the 3′LSH respectively. The terminal 5′UTR and 3′UTR regions not involved in cyclisation also show homology, implying that they might have evolved from duplicated domains.

The C gene alignment revealed a series of regions homologous with the LRS2 (Supplementary Fig. 4a). Previously we established that all LRSs could be aligned with each other by introducing multiple gaps. LRS1 and LRS2 were shorter than the other LRSs and they had lower homologies since they evolved by extensive deletions in this region. Therefore we aligned the 5′UTR and adjacent C gene with the longer and better preserved LRS3 (Supplementary Fig. 4b), and revealed homology between the LRS3 and 5′UTR/C gene region. Notably, the cyclisation sequences and sequences flanking the cyclisation domains of individual MBFV subgroups (JEV, DENV and YFV) overlapped with each other and with the TBFV group. Overall this overlapping cyclisation region is much longer than that presented in individual flavivirus groups.

Thus, robust alignments have demonstrated homology between cyclisation sequences of one virus group with non-cyclisation sequences of other virus groups and also homology between the 5′ and 3′UTRs outside all cyclisation domains. The 5′UTR was possibly derived from the 3′ UTR terminal region (or *vice versa*) using template-switching similar to that described for the generation of panhandle-, snapback- or compound-type defective-interfering particles for Vesicular stomatitis virus (Lazzarini et al., 1981).

Function of panhandles in different RNA virus groups

The proximity of 5′ and 3′ termini is potentially a universal molecular event in the replication cycle of all RNA viruses. For viruses with (+)RNA genomes the RC recognises the 5′ end of the RNA template and initiates RNA synthesis from the 3′ terminus that is mediated by the formation of the protein bridge between the 5′ and 3′ termini. In some groups of (+)RNA viruses including flaviviruses this is accomplished by direct interaction between complementary RNA sequences at the 5′- and 3′-termini with the formation of the panhandle. In the picornaviruses (Barton et al., 2001; Herold and Andino, 2001) and alphaviruses (Gorchakov et al., 2004) direct interaction between the 5′ and 3′ termini is limited to protein bridges and double-stranded panhandles are not considered to contribute to virus replication.

Computer predictions and experimental studies provide evidence for the formation of panhandles for some *Mononegavirales*. For the hantaviruses (*Bunyaviridae*) the panhandle (32 bp long) interacts with the nucleocapsid protein that as the result of intrinsic RNA chaperone activity, unwinds the two RNA strands of the panhandle and remains bound to the single-stranded sequences at the 5′ end, thus releasing the 3′end for interaction with the RC, in effect regulating RNA synthesis and encapsidation (Mir and Panganiban, 2006). Formation of a long panhandle (43 bp) has also been predicted for Ebola virus (*Filoviridae*), in which point mutations do not appear to stop virus replication although some mutations reduce virus replication rate (Crary et al., 2003). In contrast, for the arenaviruses (*Arenaviridae*) replication appears to be quite sensitive to point mutations within the panhandle region (~20 bp) (Perez and de la Torre, 2003). For Respiratory syncytial virus (*Paramyxoviridae*) computer analysis predicts the formation of a long panhandle (38 bp) where the signals for transcription, replication and encapsidation overlap (Fearns et al., 2002; McGivern et al., 2005). For influenza virus (*Orthomyxoviridae*) a short panhandle (5–7 bp) comprises the distal part of a more complex “corkscrew” model of virus promoter; any mutations disrupting the double-stranded panhandle are lethal or severely reduce promoter activity (Flick and Hobom, 1999). However the involvement of “hinge” unpaired nucleotides adjacent to the panhandle, in virion encapsidation and packaging has also been demonstrated (Tchatalbachev et al., 2001). The formation of a panhandle (~30 bp) and its significance in virus replication has been demonstrated for the double-stranded RNA rotaviruses (Patton, 2001). These examples emphasise the possible overlapping role of the panhandle, in viral replication, transcription and encapsidation although the specific molecular details vary between different virus groups.

The replicating (−)genomic and (+)antigenomic RNA molecules of all *Mononegavirales* are encapsidated but that is not the case for (+)RNA viruses. Indeed, the construction of flavivirus replicons does not require the capsid C gene (Khromykh et al., 2001). Therefore different roles for panhandles in the life cycles of (+) and (−) RNA viruses might be expected. The integrity of the panhandle (particularly 5′CYCL/3′CYCL) as a double-stranded RNA region rather than single-stranded RNA sequence, is essential for flavivirus RNA synthesis although none of the mutated replicons (with compensatory mutations in the double-stranded panhandle) showed the wild-type level of replication (reviewed in Gritsun et al., 2006). These experiments contrast with the high preservation of 5′CYCL/3′CYCL sequences among the flaviviruses indicating that this region probably plays a double function as in mononegaviruses. As a double-stranded RNA region it is required for the initiation of RNA synthesis; as a single-stranded RNA region it might be a signal for encapsidation and these two processes are known to be coupled (Khromykh et al., 2001). It is also quite possible that the RC remains bound with the single-stranded panhandle sequence after dissociation of the RNA duplex and this might explain why mutated panhandle sequence, albeit with a re-constituted double-stranded RNA region, reduced the efficiency of replication in comparison with the wild type replicon. Drawing parallels with the mononegaviruses, one can predict that the panhandle of flaviviruses also regulates processes of translation, replication and encapsidation that are carried out on the same template. For flaviviruses it is not clear which protein(s) interact with the panhandle, viral NS3 helicase, NS5 polymerase or capsid C protein via their intrinsic ability to bind nucleic acids; more experiments are required to understand the precise role of the panhandle which presumably goes beyond formation of a double-stranded RNA region.

Origin and evolution of the cyclisation function

A remarkable parallel has been drawn between the structural features of non-virion intracellular polymerase complexes of (+) sense RNA viruses and replicative cores of extracellular virions of double-stranded RNA and reverse-transcribing viruses indicating that these three virus groups share common features during genomic replication. According to this observation, these viruses all follow a common strategy for RNA synthesis but derive their infectious particles from different RNA intermediates (Ahlquist, 2006). Following this logic, the presence of the panhandle in different groups of RNA viruses could be interpreted as a universal replication strategy inherited by different viruses from a precursor lineage. The primordial panhandle might have emerged randomly due to the intrinsic property of RNA polymerase to switch templates during RNA synthesis. The formation of panhandles might have been advantageous for early viruses if it accelerated virus replication rate, for example by improving the assembly of viral RC and/or coordination of the 3′termini into the template tunnel within the viral replicase and/or regulation of replication and encapsidation. During the extensive RNA virus evolution that resulted in the formation of different RNA virus families, the role of panhandles as important elements of the virus replication process has evolved in different ways, the extreme example being the complete loss of direct RNA–RNA interactions.

As an alternative to the concept of a monophyletic origin for panhandles, early viruses might not have had panhandles. Thus, different panhandles would have had to emerge independently in each virus family. The absence of panhandles in some extant viruses would appear to support this possibility although this equally might be the result of regressive evolution. The detection of defective interfering particles with a variety of terminal panhandles in Vesicular stomatitis virus also implies that the re-emergence of panhandles could occur independently in different virus families using different template-switching mechanisms (Lazzarini et al., 1981). Additionally, the occurrence of long-terminal panhandles, with few unpaired nucleotides, supports the likelihood that panhandles have emerged relatively recently, leaving little time for accumulation of mutations. This could explain why their conservation appears to contradict the ease with which some panhandles can be modified experimentally without significant loss of replication rate.

Whilst tenuous, the balance of the evidence falls in favour of panhandles arising as the result of random independent

evolutionary events. However the commonality of panhandles in distantly related RNA virus groups is unlikely to be a random coincidence and probably reflects an ancient dependence of virus replication on a similar if not universal dsRNA panhandle being present in early viruses. In other words, panhandles may have been present in primordial RNA viruses and were possibly intimately associated with the regulation of the virus replication/encapsidation machinery as we now observe in some virus groups (Mir and Panganiban, 2006). During subsequent diversification some virus families preserved panhandles as essential elements of their life cycle, some lost them permanently (relying only on protein bridges) and some lost them but then re-acquired them. This would explain both the presence and absence of panhandles in different virus families and their variety of lengths in closely related viruses. In relation to flaviviruses it is quite possible that the precursor 5′ terminal sequence was derived from a 3′ terminal sequence (or *vice versa*) and then both evolved until the ancient flavivirus precursor "escaped" from the biologically dynamic mammalian and/or mosquito hosts into the quiescent tick (Gritsun et al., 1995). This significantly delayed the molecular evolution of the TBFV (whereas other flavivirus groups evolved at a faster rate) and enabled us to observe the early steps of sequence perturbation within flavivirus genomes, such as long duplications and extensive deletions described here and previously (Gritsun and Gould, 2006a, 2006b, 2007a, 2007b).

In summary, we have constructed robust alignments of the 5′-terminal RNA domain of four divergent flavivirus groups and demonstrated that extant flavivirus panhandles varying in sequence, length and location, might be descendants of a long cyclisation region derived by template-switch duplication of one of the terminal RNA domains present in an early lineage before diversification. In this and previously published manuscripts, manually constructed alignments have proved to be powerful tools with which to analyse the mode of UTR evolution based on multiple deletions and duplications rather than point mutations. There is clearly a need for computer algorithms that can perform these types of analysis but until they become available manual adjustment of alignments appears to be the only method for adequately analyzing the evolution and structure–function relationships of UTRs.

Acknowledgment

This work was funded by Vizier EU Contract LSHG-CT-2004-511960.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.virol.2007.04.011.

References

Ahlquist, P., 2006. Parallels among positive-strand RNA viruses, reverse-transcribing viruses and double-stranded RNA viruses. Nat. Rev. Microbiol. 4 (5), 371–382.

Barton, D.J., O'Donnell, B.J., Flanegan, J.B., 2001. 5′ Cloverleaf in poliovirus RNA is a cis-acting replication element required for negative-strand synthesis. Embo J. 20 (6), 1439–1448.

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.

Crary, S.M., Towner, J.S., Honig, J.E., Shoemaker, T.R., Nichol, S.T., 2003. Analysis of the role of predicted RNA secondary structures in Ebola virus replication. Virology 306 (2), 210–218.

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.

Fearns, R., Peeples, M.E., Collins, P.L., 2002. Mapping the transcription and replication promoters of respiratory syncytial virus. J. Virol. 76 (4), 1663–1672.

Filomatori, C.V., Lodeiro, M.F., Alvarez, D.E., Samsa, M.M., Pietrasanta, L., Gamarnik, A.V., 2006. A 5′ RNA element promotes dengue virus RNA synthesis on a circular genome. Genes Dev. 20 (16), 2238–2249.

Flick, R., Hobom, G., 1999. Interaction of influenza virus polymerase with viral RNA in the 'corkscrew' conformation. J. Gen. Virol. 80 (Pt 10), 2565–2572.

Gorchakov, R., Hardy, R., Rice, C.M., Frolov, I., 2004. Selection of functional 5′ cis-acting elements promoting efficient sindbis virus genome replication. J. Virol. 78 (1), 61–75.

Gould, E.A., Gritsun, T.S., 2006. Taxonomy and evolution of the flaviviridae. In: Kalitzky, M., Borowski, P. (Eds.), Flaviviridae: Pathogenesis, Molecular Biology and Genetics. Horizon Scientific Press, Norwich, 497 UK, pp. 17–45.

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 regions of Kamiti river virus and cell fusing agent virus originated by self-duplication. J. Gen. Virol. 87 (Pt 9), 2615–2619.

Gritsun, T.S., Gould, E.A., 2007a. Direct repeats in the flavivirus 3′ untranslated region; a strategy for survival in the environment? Virology 358 (2), 258–265.

Gritsun, T.S., Gould, E.A., 2007b. Origin and evolution of 3′UTR of flaviviruses: long direct repeats as a basis for the formation of secondary structures and their significance for virus transmission. Adv. Virus Res. 69, 203–248.

Gritsun, T.S., Holmes, E.C., Gould, E.A., 1995. Analysis of flavivirus envelope proteins reveals variable domains that reflect their antigenicity and may determine their pathogenesis. Virus Res. 35 (3), 307–321.

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: Kalitzky, M., Borowski, P. (Eds.), Flaviviridae: Pathogenesis, Molecular Biology and Genetics. Horizon Scientific Press, Norwich, UK, pp. 47–99.

Herold, J., Andino, R., 2001. Poliovirus RNA replication requires genome circularization through a protein–protein bridge. Mol. Cell 7 (3), 581–591.

Hoshino, K., Isawa, H., Tsuda, Y., Yano, K., Sasaki, T., Yuda, M., Takasaki, T., Kobayashi, M., Sawabe, K., 2006. Genetic characterization of a new insect flavivirus isolated from Culex pipiens mosquito in Japan. Virology.

Khromykh, A.A., Varnavski, A.N., Sedlak, P.L., Westaway, E.G., 2001. Coupling between replication and packaging of flavivirus RNA: evidence derived from the use of DNA-based full-length cDNA clones of Kunjin virus. J. Virol. 75 (10), 4633–4640.

Kofler, R.M., Heinz, F.X., Mandl, C.W., 2002. Capsid protein C of tick-borne encephalitis virus tolerates large internal deletions and is a favorable target for attenuation of virulence. J. Virol. 76 (7), 3534–3543.

Kofler, R.M., Hoenninger, V.M., Thurner, C., Mandl, C.W., 2006. Functional analysis of the tick-borne encephalitis virus cyclization elements indicates

major differences between mosquito-borne and tick-borne flaviviruses. J. Virol. 80 (8), 4099–4113.

Lazzarini, R.A., Keene, J.D., Schubert, M., 1981. The origins of defective interfering particles of the negative-strand RNA viruses. Cell 26 (2 Pt 2), 145–154.

McGivern, D.R., Collins, P.L., Fearns, R., 2005. Identification of internal sequences in the 3′ leader region of human respiratory syncytial virus that enhance transcription and confer replication processivity. J. Virol. 79 (4), 2449–2460.

Mir, M.A., Panganiban, A.T., 2006. The bunyavirus nucleocapsid protein is an RNA chaperone: possible roles in viral RNA panhandle formation and genome replication. RNA 12 (2), 272–282.

Patton, J.T., 2001. Rotavirus RNA replication and gene expression. Novartis Found Symp. 238, 64–77 (discussion 77–81).

Perez, M., de la Torre, J.C., 2003. Characterization of the genomic promoter of the prototypic arenavirus lymphocytic choriomeningitis virus. J. Virol. 77 (2), 1184–1194.

Tchatalbachev, S., Flick, R., Hobom, G., 2001. The packaging signal of influenza viral RNA molecules. RNA 7 (7), 979–989.