Direct repeats in the 3’ untranslated regions of mosquito-borne flavviruses: possible implications for virus transmission

T. S. Gritsun and E. A. Gould

Centre for Ecology and Hydrology, Mansfield Road, Oxford OX1 3SR, UK

Direct repeats (DRs) of 20–45 nucleotide conserved sequences (CS) and repeated CS (RCS), separated by non-conserved sequences up to 100 nucleotides long, were previously described in the 3’ untranslated region (3’UTR) of the three major mosquito-borne flavivirus (MBFV) subgroups, represented by Japanese encephalitis virus, Yellow fever virus and Dengue virus. Each subgroup exhibits a specific pattern of DRs, the biological significance of which has not yet been adequately addressed. The DRs were originally identified using conventional alignment programs based on the assumption that genetic variation is driven primarily by nucleotide substitutions. Since there are no recognized alignment programs that can adequately accommodate very divergent sequences, a method has been devised to construct and analyse a substantially improved 3’UTR alignment between these highly divergent viruses, based on the concept that deletions and/or insertions, in addition to substitutions, are important drivers of 3’UTR evolution. This ‘robust alignment’ approach demonstrated more extensive homologies in the 3’UTR than had been recognized previously and revealed the presence of similar DRs, either intact or as sequence ‘remnants’, in all the MBFV subgroups. The relevance of these observations is discussed in relation to (i) the function of DRs as elements of replication enhancement, (ii) the evolution of RNA secondary structures and (iii) the significance of DRs and secondary structures in MBFV transmissibility between vertebrate and invertebrate hosts.

INTRODUCTION

The genus Flavivirus (family Flaviviridae) contains four groups, the tick-borne flaviviruses (TBFV), the mosquito-borne flaviviruses (MBFV), the no-known-vector flaviviruses (NKVFV) and the non-classified flaviviruses (NCFV). The MBFV are further subdivided into three subgroups, the Japanese encephalitis virus (JEV), Yellow fever virus (YFV) and Dengue virus (DENV) subgroups, each of which includes human pathogens (Heinz et al., 2000; Gould et al., 2001). They are enveloped viruses (50 nm in diameter) with single-stranded positive-sense RNA genomes of approximately 11,000 bases. The RNA contains a single-stranded open reading frame (ORF) that encodes a polyprotein (about 3400 amino acids) which is processed co-translationally to produce the structural (envelope, membrane and capsid) and non-structural proteins, the latter providing the proteolytic and replicative functions (Lindenbach & Rice, 2001).

The ORF is flanked by the 5’ and 3’ untranslated regions (UTRs). In common with other positive-stranded RNA viruses, the 5’UTR initiates virus translation and the 3’UTR might also contribute to this process (Li & Brinton, 2001; Edgil et al., 2003; Holden & Harris, 2004; Chiu et al., 2005; Tilgner et al., 2005). The 3’UTR initiates virus replication, also mediated by interaction with the 5’UTR (You & Padmanabhan, 1999; Ackermann & Padmanabhan, 2001; Charrel et al., 2001; Khromykh et al., 2001; You et al., 2001; Conver et al., 2003; Lo et al., 2003). The mode of replication of flavivirus RNA is asymmetrical, with an excess of positive-strand RNA and no free minus strands in infected cells (Westaway et al., 2002, 2003).

Previous studies of the MBFV 3’UTR identified short direct repeats (DRs) as 20–45 nucleotide regions conserved between the JEV, YFV and DENV subgroups and separated by more divergent sequences. The MBFV DRs were initially specified as CSs and RCSs (conserved/repeated conserved sequences) and also YFV-R (YFV repeats) (Hahn et al., 1987). CS1 and CS2 sequences were present in all three MBFV subgroups, whereas the RCS2 was identified only in the JEV and DENV subgroups (Hahn et al., 1987). Surprisingly, CS2 and RCS2 were also present in the 3’UTR of the NKVFV group (Charlier et al., 2002), which is phylogenetically quite distant from the MBFV. In addition, CS3 and RCS3 were identified only in the JEV subgroups, whereas the three tandem repeats YF-R1, YF-R2 and YF-R3

Supplementary figures are available in JGV Online.
were specific for the YFV and YFV-related subgroup (Hahn et al., 1987; Mutebi et al., 2004).

Experiments using infectious clones and replicons revealed that, with the exception of CS1, all DRs could be deleted without the engineered viruses losing their replicative capacity or viability (Men et al., 1996; Khromykh & Westaway, 1997; Mandl et al., 1998; Pletnev, 2001; Bredenbeek et al., 2003; Lo et al., 2003; Tilgner et al., 2005). These observations raised a number of questions: why are DRs so well preserved, why are they preserved as repeats and how did they originate?

We have constructed an alignment between the 3′UTRs of MBFV based on the principle that the evolutionary constraints in the UTR could be different from those in the ORF, allowing greater nucleotide perturbations such as large deletions and/or insertions. This approach enabled us to reveal sequence ‘remnants’ of CSs/RCSs in all MBFV subgroups, leading us to propose that all CSs/RCSs were present in early MBFV lineages and that their preservation within the individual subgroups was driven by their specific function as enhancers of virus replication. These findings advance our understanding of the molecular organization of flavivirus UTRs and in particular their evolution and relevance to virus transmission and dissemination between vertebrate and invertebrate hosts.

**METHODS**

The MBFV subgroups are highly divergent, and the mean pairwise nucleotide sequence identity of the 3′UTRs for the JEV, DENV and YFV subgroup viruses is too low (<50%) to enable automated construction of a robust alignment. To overcome this problem, we initially constructed three comparative alignments for available 3′UTR sequences of 55 viruses of the JEV subgroup, 137 viruses of the DENV subgroup and 39 viruses of the YFV subgroup using CLUSTAL X followed by manual editing. The 3′UTRs of the YFV-related viruses Banzi virus (BANZV), Uganda S virus (UGSV), Sepik virus (SEPV) (Mutebi et al., 2004) and Yokose virus (YOKV) (Taijima et al., 2005) were aligned separately with YFV, and Kokobera virus (KOKV) and Stratford virus (STRV) (Poidinger et al., 1996; Nisbet et al., 2005) with JEV. Subsequently, identical or nearly identical 3′UTR sequences were excluded from the subgroup alignments by visual inspection. Accordingly, the remaining 3′UTR sequences show maximal virus divergence within each subgroup and they were used for manual construction of a pan-MBFV alignment (see Supplementary Figs S1 and S2 available in JGV Online). Multiple large and small gaps were introduced manually in the alignments using CSs/RCSs as anchors. The viruses used for the alignments are specified by their accession numbers (Supplementary Figs S1 and S2).

**RESULTS AND DISCUSSION**

In published 3′UTR alignments, the homology between all the MBFV was limited to 20–45 nucleotide DRs, i.e. CSs/RCSs and YFV repeats identified as highly conserved regions separated by 50–100 base variable regions. However, available alignment programs do not search for regions of homology that are separated by extended non-homologous regions although, for many RNA viruses, the evolution of UTRs is mediated not only by substitution mutagenesis, but also by deletions and duplications (Copper et al., 1974; Faragher & Dalgarno, 1986; Bryan et al., 1992; Lai, 1992; Santagati et al., 1994; Pilipenko et al., 1995; Hajjou et al., 1996; Peerenboom et al., 1997; Shi et al., 1997; Warren & Murphy, 2003).

The results of construction of robust 3′UTR alignments for the MBFV are presented starting with the alignment for YFV and YFV-related viruses that were previously produced using conventional computer programs (Mutebi et al., 2004). Our manually corrected alignment revealed a number of important gaps and demonstrated that, even between closely related YFV, 3′UTR evolution involves deletions and insertions rather than substitutions.

**3′UTR alignment of YFV subgroup viruses**

Supplementary Fig. S1(a) (available in JGV Online) illustrates a 3′UTR alignment between selected strains of the three recognized YFV genotypes, i.e. West African (W), Central/East African (CE) and South American (SA); it also includes three YFV-related viruses, BANZV, UGSV and SEPV (Wang et al., 1996; Mutebi et al., 2001, 2004; Mutebi & Barrett, 2002). Although YOKV is also recognized as a YFV-related virus, we included it in the pan-MBFV alignment due to the presence of a poly-AT stretch that aligns only with JEV (see below).

Originally, three 45-nucleotide-long tandem repeats were described for W strains of YFV (Hahn et al., 1987). It was subsequently demonstrated that EC and SA strains have lost one (YFV-R2) and two (YFV-R1 and YFV-R2) repeats, respectively (Wang et al., 1996; Mutebi et al., 2004).

Our alignment (Supplementary Fig. S1a) shows that the process of losing the repeats occurred gradually, with a series of deletions that reduced, but did not completely delete, the repeat elements in different YFV strains. The YFV W strains contain all three intact YFV-specific repeat sequences, whilst the CE strains contain intact YFV-R1/R3 and partial YFV-R2. The SA strains contain intact YFV-R3 and partial YFV-R2/R1 sequences.

The possibility that multiple template switching of RNA polymerase, rather than point mutations, shaped the phenotype of the YFV 3′UTR was supported by comparison with the YFV-related viruses BANZV, UGSV and SEPV (Supplementary Fig. S1a). Previously, it had been reported that they also contained YFV-specific repeats, with two repeats for SEPV, one for UGSV and three for BANZV (Mutebi et al., 2004). The alignment in Supplementary Fig. S1(a) illustrates that all three repeats, in a complete or truncated form, were present in all the viruses. However, to identify these repeats, it was necessary to introduce two gaps, a short G1 and a long G2 between YFV-R2 and YFV-R3.

Mutebi et al. (2004) suggested that a short sequence of the YFV-repeat, 5′-AACCGGATACAAC-3′, was acquired from the Drosophila melanogaster genome. However, our
alignment reveals an alternative possibility, that each repeat was initially formed by multiple duplications and subsequent mutagenesis of an AACCGGG sequence that was initially produced by ‘stuttering’ of the RNA polymerase. Each YFV repeat could have contained four of these initial AACCGGG repeats. The sequences between YFV repeats were also probably part of the longer repeats. For example, there is a TAAAT(T)GG repeat in G1 and G2 only for BANZV, whereas TAAATGG is present in the G2 of other YFV. Thus, TAAAT(T)GG sequences may have been present in the G1 of all YFV, but were subsequently deleted. Another example is the repeat sequence CTGT that is present in all three YFV-related viruses in G2 and also downstream of YFV-R3 (Supplementary Fig. S1a). These examples also support the suggestion that insertional mutagenesis could contribute to the formation of the G2 region.

In order to assess whether or not the YFV repeats are part of longer repeat sequences, they were placed under each other and aligned (Supplementary Fig. S1b available in JGV Online). The TGAAAC sequence that follows YFV-R2 within the G2 region and which is observed only in YFV (not in SEPV, UGSV or BANV) might be relocated from G2 into the G0 region (compare Supplementary Figs S1a and S1b). Less homology is observed in the original position of this sequence within G2, and therefore the ambiguous position of the sequence TGAAAC in either G2 or G0 is indicated (boxed in Supplementary Fig. 1b). These examples suggest that YFV repeats were originally longer. Furthermore, the introduction of large gaps enabled us to detect remnants of YFV-R4 that was probably also present in a YFV ancestral lineage.

Examination of the alignment downstream of YFV-R3 also indicates that the evolution of YFV and YFV-related viruses occurred primarily through deletions and insertions rather than substitutions. A 216 nucleotide duplication has also been demonstrated in the 3′UTR of numerous South American genotype 1 YFV isolates (Bryant et al., 2005). Whilst the origin of CE strains is not known (beyond the phylogenetic prediction), the evolution of SA strains from W strains is assumed to have resulted from the introduction of virus into South America during the slave trade, 100–400 years ago (Gould et al., 2003). Therefore, even during this relatively short period, deletions and duplications have substantially changed the 3′UTR of closely related viruses.

Thus, the deduced consensus sequence might represent a longer 3′UTR, of potentially 760 nucleotides, that was present in an ancestral lineage of the YFV and YFV-like viruses. The current longest identified YFV 3′UTR contains 511 nucleotides; the precise length of the 3′UTR of YFV-related viruses has not been determined. On the basis of this alignment and our interpretation of the findings, a schedule for the possible development of the YFV repeats from a common ancestral lineage is presented in Fig. 1.

3′UTR alignment of all MBFV subgroups

A comparative 3′UTR alignment between all available MBFV is demonstrated in Supplementary Fig. S2 (available in JGV Online). The CSs and RCSs were used as anchors to search for homology between them. The elements of the predicted RNA secondary structures, superimposed on the alignment, include (i) the 3′LSH with a conserved pentanucleotide CACAG in loop1 (Brinton et al., 1986), (ii) a short upstream SL2 that interacts with the 3′LSH in the pseudoknot (Shi et al., 1996), (iii) dumbbell-like structures, DB1 (exposing CS2), shared between JEV, DENV and YFV, and DB2 (exposing RCS2), shared between JEV and DENV (Proutski et al., 1997a; Olsthoorn & Bol, 2001), and (iv) the

---

**Fig. 1.** Schematic organization of repeats YFV-R1, -R2 and -R3 as deduced from the alignment in Supplementary Fig. S1 within the MBFV 3′UTR; adapted from previous publications (Hahn et al., 1987; Mutebi et al., 2004) with modifications. Intact YFV repeats are shown as shaded boxes, whereas partial YFV repeats are shown as open boxes. Solid lines depict extended regions of homology whereas dotted lines indicate regions where homology is confined to short nucleotide regions. Gaps between YFV repeats are specified as G0, G1 and G2 and reduced YFV-R4, in correspondence with the alignment in Supplementary Fig. S1.
3’CYCL sequence, CATATTGA, which interacts with the inverted repeat 5’CYCL sequence in the capsid gene to circularize the virus genome (Hahn et al., 1987), required for the initiation of replication. The experimentally established boundaries that define whether or not a virus is viable (Men et al., 1996) are also indicated in Supplementary Fig. S2. The observations made from this alignment are described below.

**CSs and RCSs are present in all the MBFV.** Although the CSs/RCSs were previously the only regions of recognized homology between all MBFV, the alignment in Supplementary Fig. S2 reveals a number of other shorter regions of homology which we interpret as remnants of longer preserved elements between the CSs and RCSs. YFV sequence was similar in some regions to the DENV or JEV subgroup outside the CSs, but not necessarily with both virus groups in the same regions. A schematic illustration of the presence of CSs/RCSs in all MBFV is presented in Fig. 2.

Notably, YOKV, which does not have a recognized arthropod vector but diverged from the mosquito-borne YFV-related virus lineages (Kuno et al., 1998; Gould et al., 2003; Tajima et al., 2005; Cook & Holmes, 2006), showed significant homology with YFV and the YFV-related viruses which was identified following the introduction of multiple deletions and insertions. Extensive deletions were also apparent in KOKV and STRV (Supplementary Fig. S2), both of which evolved from JEV-related evolutionary lineages (Poidinger et al., 1996; Kuno et al., 1998; Gould et al., 2001; Nisbet et al., 2005).

**Fragments of the YFV repeat sequences are present in other MBFV.** In previous publications, no homology between YFV repeats and other MBFV was demonstrated. We examined this in more detail by introducing multiple gaps manually into the alignment. However, only short remnants of homologous sequences of each YFV repeat were found in the DENV and JEV subgroups (Supplementary Fig. S2). The region between YFV-R1 and YFV-R2 also contained traces of similarity between the different MBFV subgroups (Supplementary Fig. S2). The alignment in this region is not robust, but the short regions of homology could still represent the remnants of ancient sequences common to all MBFV in accordance with their monophyletic origin (Gould et al., 2001). The most likely scenario is that the three YFV repeats were originally present in an early lineage and then subsequently evolved by deletions, insertions and substitutions into the

---

**Fig. 2.** Scheme for the development of the CSs and RCSs of the MBFV deduced from the alignments in Supplementary Fig. S2. Regions of high nucleotide conservation (CS, RCS and YFV repeats) are boxed and regions of nucleotide variability between them are depicted as solid lines. Shaded boxes specify intact CSs, RCSs, YFV repeats and poly-AT, whereas partial conservation is depicted as open boxes. Schematically presented RNA secondary structures are depicted against the corresponding regions of the 3’UTR and are specified appropriately.
JEV, DENV and YFV subgroups (Fig. 2). However, the YFV repeats were not reproduced robustly in the MBFV subgroups, and therefore they may be unique to the YFV subgroup.

Was the poly-AT region originally present in all the MBFV? The nucleotide region between the YFV stop codon and the YFV-R1 is AT-enriched and is referred to as the poly-AT region. This region appears to have evolved as the result of multiple stuttering by the viral RNA polymerase around the TAA stop codon. West Nile virus (WNV) and Murray Valley encephalitis virus (MVEV) have the longest AT-enriched regions whereas, in the DENV and YFV subgroups, this region is much shorter, suggesting that AT-enrichment occurred only in the JEV subgroup. However, the equivalent region of YOKV aligns quite closely with JEV, although beyond the poly-AT region YOKV aligns closely with YFV, implying that the precursor lineage of YFV may have had a fully developed AT-enriched domain. In order to illustrate this, the YOKV 3’UTR sequence is placed next to both YFV and JEV in the alignment (Supplementary Fig. S2).

The preservation of the poly-AT region in JEV may reflect a selective advantage in the natural habitat of this virus. In structural and functional terms, the AT-enriched region could promote the formation of relaxed, i.e. unstable secondary structures. This is supported by the observation that the region of the 3’UTR proximal to the stop codon of all flaviviruses is extremely variable and in the laboratory is readily altered by spontaneous deletions (Mandl et al., 1998; Mutebi et al., 2004). Moreover, engineered deletion of this region did not prevent the production of infectious virus (Men et al., 1996; Mandl et al., 1998). Computer-predicted folding using Mfold (available at http://mfold.burnet.edu.au/) demonstrated the formation of only short AT- and GT-paired stem–loop structures (not shown). The advantage for the virus to preserve these new ‘immature’ sequences might be that they do not form long stable stems and might serve as spacers that separate the ORF and folded distal 3’UTR essential for the assembly of the replication complexes (RC). The folding of YFV repeats as two short stem–loop structures (Mutebi et al., 2004) demonstrated a similar possibility.

Superimposition of secondary RNA structures on the MBFV alignment. Another important observation from Supplementary Fig. S2 is the compatibility between the positions of RNA secondary structures predicted for each subgroup of the MBFV. Although these viruses are divergent, the sequences that form the 3’LSH, SL2, DB1 and DB2 for individual groups were located in equivalent regions of the 3’UTR for each of the MBFV subgroups (Supplementary Fig. S2). Firstly, this could be interpreted as evidence (although indirect) of the robustness of the alignment and, secondly, it implies that the secondary RNA structures in the most divergent MBFV lineages could have been formed by similar nucleotide sequences.

Each of these RNA structures exposes the Cs/RCs, the preservation of which might be required for an essential signalling function for RNA replication. For example, the 3’LSH exposes the pentanucleotide CACAG which is conserved between vector-borne flaviviruses (Charlier et al., 2002); an engineered mutation in this region abolishes or impairs virus replication (Gritsun et al., 2001; Khromykh et al., 2003; Tilgner et al., 2005). Similarly, DB1 and DB2 expose CS2 and RCS2 (Olsthoorn & Bol, 2001), the significance of which is not obvious since, despite their conservation, they are not essential for viability of experimentally engineered viruses, although mutant viruses had reduced infectivity titres (Men et al., 1996; Proutski et al., 1999; Bredenbeek et al., 2003).

The CS3 and RCS3 that are located in close proximity to the stop codon of JEV also form two identical short stem–loop structures (Fig. 2). The deletion of CS3 and RCS3 did not lead to the loss of virus infectivity, although reduced virus replication was observed (Khromykh & Westaway, 1997; Lo et al., 2003; Tilgner et al., 2005).

What is the role of the Cs/Rcs? Previously, it was proposed that the 3’UTR consists of a promoter region that is essential for virus viability and an enhancer region that is not essential for virus viability but increases the RNA replication rate (Proutski et al., 1999). The experimentally established boundary that determines whether or not the virus is viable (Men et al., 1996) could correspond to the boundary between the promoter and enhancer region (Supplementary Fig. S2). Thus, the minimal promoter function in MBFV could be provided by 3’LSH–SL2–CS1; the engineered virus with a completely deleted enhancer region is infectious, but only in mosquito (not in vertebrate) cells (Men et al., 1996). Accordingly, the virus replication enhancer includes all DRs (Cs/Rcs and YFV repeats) that probably facilitate virus replication in vertebrate cells.

Experiments that demonstrate a non-essential role of the DRs in virus viability and RNA synthesis (Men et al., 1996; Khromykh & Westaway, 1997; Bredenbeek et al., 2003; Lo et al., 2003; Tilgner et al., 2005) appear to contradict the concept of conservation of the Cs/Rcs (Hahn et al., 1987; Mutebi et al., 2004). However, laboratory experiments may not adequately reflect the essential requirements of DRs for flaviviruses in nature. Since all engineered MBFV exhibit reduced RNA replication rates, compared with wild-type viruses, it is possible that a high RNA replication rate might exert a selection pressure on MBFV. Mosquitoes feed on vertebrate hosts for no more than a few seconds; consequently, the frequency of viral transmission is increased if the virus rapidly reaches a high titre in the blood, before the immune system reduces viraemia. Indeed, the engineered deletions of Cs and Rcs decrease the replication rate of DENV4 to a greater extent in simian than in mosquito cells (Men et al., 1996). This corresponds with the proposition that the virus enhancer has evolved to meet
the requirements of the short viraemic period in the vertebrate host. In contrast, infected mosquitoes have a period of weeks to amplify the virus before the next feeding period, and therefore the enhancement function is less essential for virus replication in mosquitoes than in vertebrates.

In molecular terms, how could the enhancer accelerate virus replication? It is recognized that a wide variety of host proteins, with functions not directly related to cellular RNA synthesis, are frequently associated with replication of positive-strand RNA viruses (reviewed by Lai, 1998; Boguszewska-Chachulska & Haenni, 2005), including flaviviruses (Blackwell & Brinton, 1995, 1997; Ta & Vrati, 2000; De Nova Ocampo et al., 2002). These are often identified as proteins with RNA-binding and trafficking activities. Indeed, the individual steps of the flavivirus life cycle, i.e. translation, RNA synthesis and assembly, take place in a variety of cellular compartments (convoluted membranes and vesicular packets) that proliferate in cells in response to virus infection (Mackenzie et al., 2001; Westaway et al., 2002, 2003; Uchil & Satchidanandam, 2003). The DRs might interact with cellular/viral proteins that direct viral RNA from the sites of translation into the replication sites where assembly of viral RCs is most efficient. Alternatively, the Cs could be direct/indirect signals for cells to express factors essential for efficient viral RNA synthesis. In the absence of DRs, flavivirus replication could be initiated at low efficiency because of the leakage of viral RNA into the appropriate cell compartment or because of the absence of sufficient levels of cellular components that accelerate the assembly of the viral RC.

The particular signalling function of DRs has presumably evolved in different flaviviruses according to their adaptation to different invertebrate and vertebrate hosts. In other words, DRs as enhancers could be the means by which virus adapts to the molecular environment of a new host to maximize replication and therefore increase transmission efficiency. This would explain why some DRs are shared between MBFV subgroups, whilst others are different. It could also explain the apparent discrepancy between conservation of the Cs and RCSs and their apparent non-essential function in laboratory-maintained viruses.

The reason for preservation of the DRs as duplicated signals is less clear. It is possible that host/viral proteins involved in RNA synthesis are active as dimers and therefore the presence of a double signal would accelerate assembly of the RC. The effect of each individual element of the replication enhancer could be relatively moderate and insignificant for laboratory-maintained viruses, but the cumulative action of DRs might accelerate virus replication and this could be a vital requirement for efficient virus transmission and dissemination in the more fastidious natural environment.

If DRs do provide a signalling function, a plausible explanation can now be offered for the preservation of YFV repeats in W strains and their loss in CE and SA strains; their signal function as triple DRs may be essential only in the particular hosts (mosquitoes and vertebrate) in West Africa. Once virus escapes this environment, the need for a triple signal is abolished as postulated for CE and SA strains during their adaptation to new species.

**Why is CS1 conserved?** The exposed positions of the DRs in the secondary RNA structure of MBFV are compatible with a proposed signalling function for these elements and also provide a possible reason for why they are conserved. However, the reason for CS1 conservation is less clear. CS1 is recognized as the most conserved region within the MBFV, but the alignment (Supplementary Fig. S2) shows that this region is not absolutely conserved. Nucleotide differences in CS1 were revealed not only between the MBFV subgroups but also within closely related viruses such as MVEV and other viruses from the JEV subgroup (Supplementary Fig. S2). Secondly, the overlapping of two structural models, i.e. the linear (Proutski et al., 1997a, b, 1999) and the cyclization model (Kromykh et al., 2001; Thurner et al., 2004), within CS1 enabled us to conclude that CS1 may not represent one continuous domain. In fact, linear CS1 could be subdivided into two domains. The first was already described, i.e. 3′CYCL, which is absolutely conserved between all MBFV and has been proven experimentally to interact with 5′CYCL to initiate RNA replication (You & Padmanabhan, 1999; Ackermann & Padmanabhan, 2001; Charrel et al., 2001; Kromykh et al., 2001; You et al., 2001; Corver et al., 2003; Lo et al., 2003). However, for RNA–RNA interactions to occur, there is no requirement for absolute preservation of the sequence, and it was suggested earlier that 3′CYCL, as either a single-stranded or double-stranded sequence, might have a signalling function, not only for replication but also for virion assembly (Proutski et al., 1999).

Recently, it was also suggested that, in addition to interacting with 5′CYCL, 3′CYCL might interact with the 5′UTR (possibly mediated by cellular proteins) to initiate translation (Chiu et al., 2005). It is also noteworthy that the pseudoknot PK1, which is conserved in the JEV and DENV subgroups (Olsthoorn & Bol, 2001), overlaps with the CS1 sequence (Supplementary Fig. S2).

The second domain of the CS1, downstream of the 3′CYCL, is less conserved and has not been predicted to be involved in the formation of a well-defined secondary structure. However, sequences of several secondary structures overlap significantly within this domain. One of these is SL2, with a loop that exposes a pseudoknot-interacting sequence (Fig. 2). Additionally, in the cyclization model, some sequences of the CS1 beyond 3′CYCL are involved in the formation of a long stem (interacting with the 5′UTR and capsid gene) that is essential to initiate virus replication, although these sequences are not identical within the three MBFV subgroups (Supplementary Fig. S2).

Thus, CS1 probably does not represent an entire entity and its conservation might be explained by the overlapping of...
How did CS and RCS emerge? What were the molecular events that resulted in the appearance of 20- to 45-nucleotide repeats separated by longer (~50–130 nucleotides) non-repeated sequences? It is possible that the DRs within the 3'UTR of the MBFV were initially longer, but they were subsequently reduced to short DRs (CSs/RCSs and YFV repeats) by extensive mutagenesis. Therefore, these short DRs could represent remnants of longer tandem repeats that have gradually been deleted and mutated. However, our alignment did not reveal similar remnants outside the CSs, probably because the MBFV subgroups are so divergent that no traces remain.

Nevertheless, we recently demonstrated that formation of the 3'UTR of TBFV was mediated by multiple duplication of long regions of about 200 nucleotides (Gritsun & Gould, 2006b); the preservation of these long DRs might be explained by the slower rate of evolution of TBFV in comparison with MBFV (Gritsun et al., 1995; Zanotto et al., 1996). We also demonstrated that formation of the 3'UTR of NCFV resulted from the duplication of a 600-nucleotide-long sequence (Gritsun & Gould, 2006a). Duplication of a region of 216 nucleotides in the 3'UTR of South American YFV isolates effectively doubled the secondary RNA structures and recreated RCS2, not present in 'conventional' YFV isolates (Bryant et al., 2003). Finally, we have recently demonstrated the possible common origin of conserved DRs in the genomes of all flaviviruses, including the most divergent species. We believe that this can be explained by development of their 3'UTRs from a common precursor lineage (manuscript in preparation).

Conclusions

In this paper, we have described the construction and analysis of a robust comparative alignment between the 3'UTRs of three divergent MBFV subgroups, the JEV-, DENV- and YFV-related viruses. Conserved DRs (CSs/RCSs and YFV repeats) were revealed either intact or as remnants in each of the three MBFV subgroups, suggesting their common origin. We conclude that, at the molecular level, DRs function as replication enhancers selected under the constraints of the particular transmission and dissemination imposed by the hosts. Appropriate genetic modification of MBFV DRs and studies of their transmission efficiency would clarify whether or not DRs were preserved as essential elements for virus survival in the environment.

REFERENCES


