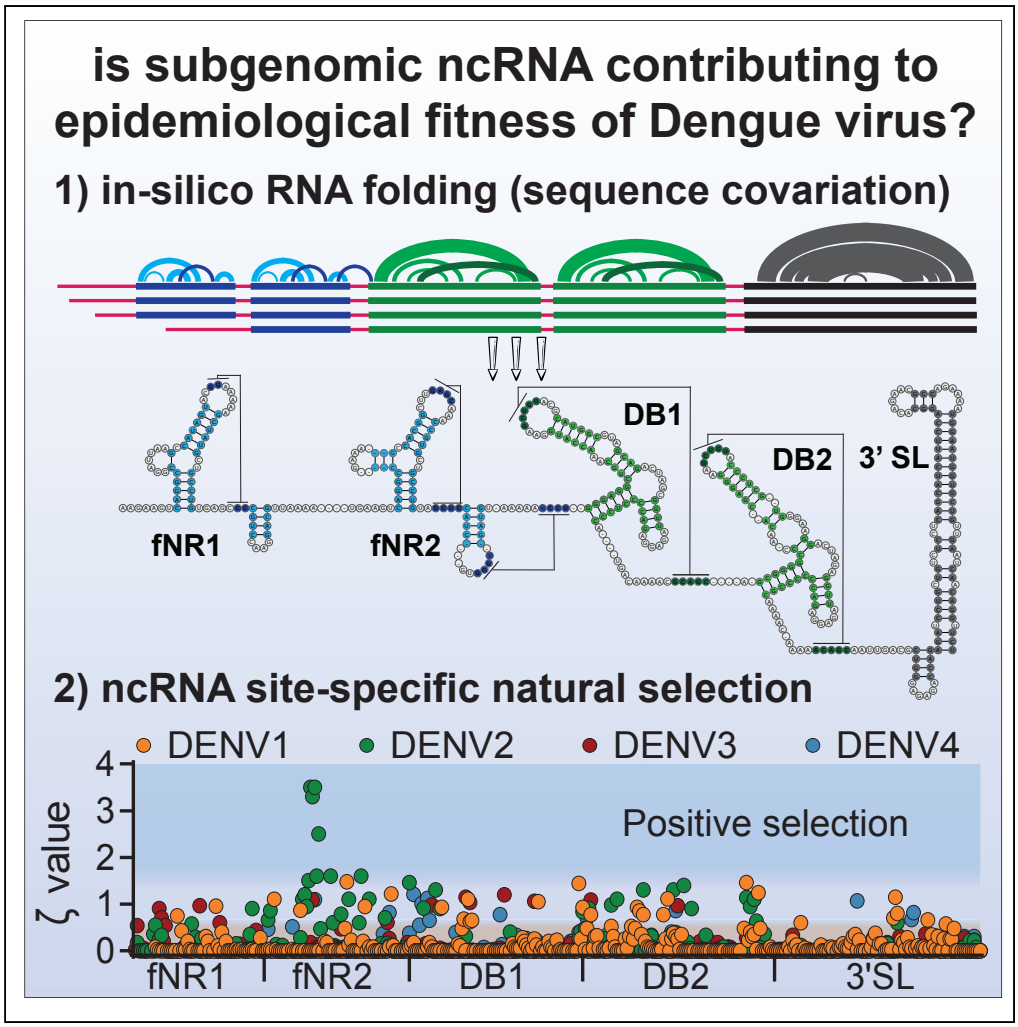
iScience

CellPress

Article

Evolution of Subgenomic RNA Shapes Dengue Virus Adaptation and Epidemiological Fitness



Esteban Finol, Eng Eong Ooi

esteban.finol@mol.biol.ethz.ch (E.F.)
engeong.ooi@duke-nus.edu.sg (E.E.O.)

HIGHLIGHTS

Dengue viruses (DENVs) preserve RNA elements in their 3′ untranslated region (UTR).

Quantification of natural selection revealed positive selection on DENV2 sfRNA

Flaviviral nuclease-resistant RNAs (fNR) in the 3′ UTRs contribute to DENV speciation

A highly evolving fNR structure appears to increase DENV2 epidemiological fitness

Finol & Ooi, iScience *16*, 94–105
June 28, 2019
https://doi.org/10.1016/j.isci.2019.05.019

iScience

CellPress

Article

Evolution of Subgenomic RNA Shapes Dengue Virus Adaptation and Epidemiological Fitness

Esteban Finol[1,2,3,5,6,*] and Eng Eong Ooi[1,3,4,*]

SUMMARY

Changes in dengue virus (DENV) genome affect viral fitness both clinically and epidemiologically. Even in the 3′ untranslated region (3′ UTR), mutations could affect subgenomic flaviviral RNA (sfRNA) production and its affinity for host proteins, which are necessary for successful viral replication. Indeed, we recently showed that mutations in DENV2 3′ UTR of epidemic strains increased sfRNA ability to bind host proteins and reduce interferon expression. However, whether 3′ UTR differences shape the overall DENV evolution remains incompletely understood. Herein, we combined RNA phylogeny with phylogenetics to gain insights on sfRNA evolution. We found that sfRNA structures are under purifying selection and highly conserved despite sequence divergence. Only the second flaviviral nuclease-resistant RNA (fNR2) structure of DENV2 sfRNA has undergone strong positive selection. Epidemiological reports suggest that substitutions in fNR2 may drive DENV2 epidemiological fitness, possibly through sfRNA-protein interactions. Collectively, our findings indicate that 3′ UTRs are important determinants of DENV fitness in human-mosquito cycles.

INTRODUCTION

Dengue virus (DENV) is the leading cause of mosquito-borne viral disease globally. An estimated 100 million cases of acute dengue occur annually, some of which develop into life-threatening severe dengue (Bhatt et al., 2013). DENV exists as four antigenically distinct but genetically related viruses (DENV1–4), all of which can cause the full spectrum of disease outcome. A tetravalent dengue vaccine has been licensed in several countries for use to prevent dengue. However, its protective efficacy varied across the four serotypes of DENV, and long-term protection was only observed in older children with at least one episode of prior DENV infection (Hadinegoro et al., 2015). Thus despite application of this vaccine and current approaches to vector control, DENV will likely continue to be a major public health challenge in the coming years.

Dengue is distributed throughout the tropics and is now encroaching into the subtropical regions of the world, causing frequent and recurrent epidemics (Messina et al., 2015). Although several of these epidemics were caused by fluctuations in the relative prevalence of the DENV serotypes in a background of low herd serotype-specific immunity, genetic differences in DENV also appear to play a distinct role in epidemic emergence (OhAinle et al., 2011). Indeed, we showed that the 3′ untranslated region (3′ UTR) of DENV genome contributes to the epidemiological fitness of DENV (Manokaran et al., 2015). Nucleotide substitutions in the 3′ UTR of DENV2 strains resulted in increased subgenomic flaviviral RNA (sfRNA) levels. This viral non-coding RNA (ncRNA) binds to TRIM25 protein to inhibit its deubiquitylation; without TRIM25 E3 ligase activity, RIG-I signaling for type I interferon (IFN) induction was repressed. Reduced type I IFN response, at least in part, contributed to the increased viral spread of these strains in Puerto Rico in 1994 (Manokaran et al., 2015). More recently, we have also shown that sfRNA from the same DENV2 isolated during the 1994 Puerto Rico outbreak also disrupts the antiviral response in the salivary gland of the *Aedes* mosquito vector (Pompon et al., 2017). Similarly, changes in the 3′ UTR sequence that resulted in increased sfRNA production were observed in a new DENV2 clade in 2005 that resulted in a dengue epidemic in Nicaragua (OhAinle et al., 2011; Manokaran et al., 2015). Likewise, nucleotide composition in the 3′ UTRs also differentiated dominant from weaker DENV strains in Myanmar, India, and Sri Lanka (Myat Thu et al., 2015; Dash et al., 2015; Silva et al., 2008; respectively), although the structural consequences and impact of those substitutions on viral fitness have yet to be experimentally defined.

DENV 3′ UTRs can be functionally segmented into three domains. The two more downstream domains possess RNA structures necessary for viral genome cyclization, viral RNA synthesis, translation, and replication (Alvarez

[1]Programme in Emerging Infectious Diseases, Duke-NUS Medical School, Singapore 169857, Singapore

[2]Swiss Tropical and Public Health Institute, University of Basel, Basel 4051, Switzerland

[3]Department of Microbiology and Immunology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117545, Singapore

[4]Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117549, Singapore

[5]Present address: Institute for Molecular Biology and Biophysics, ETH Zurich, Zürich 8093, Switzerland

[6]Lead Contact

*Correspondence: esteban.finol@mol.biol.ethz.ch (E.F.), engeong.ooi@duke-nus.edu.sg (E.E.O.)

https://doi.org/10.1016/j.isci.2019.05.019

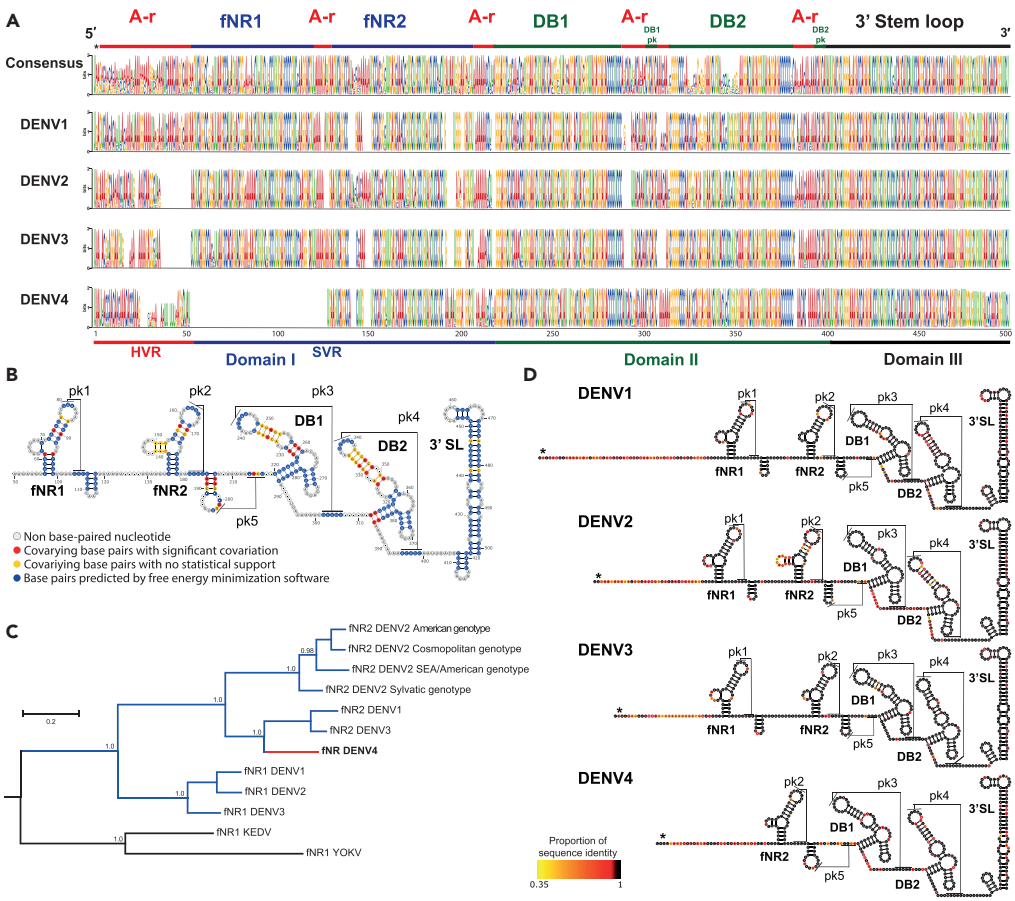


Figure 1. The 3′ UTR of Dengue Viruses Diverged by Deletion and Sequence Coevolution of Functional RNA Structures

Dengue viruses are phylogenetically related RNA viruses; their 3′ UTR sequences have diverged along evolution, and they now differ in sequence length and nucleotide composition. However, they kept functional RNA structures through sequence covariation. We observed and quantified RNA sequence covariation to predict secondary structures in the 3′UTRs and implemented Bayesian RNA phylogenetics to establish the phylogenetic relations among the RNA structures across dengue viruses.

(A) Alignment of DENV 3′ UTR sequence logos. Sequence logos for the 3′ UTR of all DENV were aligned based on the multiple sequence alignment of DENV 3′ UTRs sequences (consensus sequence logo). The three domains of DENV 3′ UTR are highlighted at the bottom of the figure, along with the hypervariable (HVR) and semi-variable (SVR) regions in domain I. Highly conserved sequences were used to demark the boundaries between three domains in the 3′ UTR of dengue viruses. A detailed view on domain-specific sequence logos is provided in Figures S1–S3. Five conserved stretches were found across DENV 3′ UTRs. They correspond to two flaviviral nuclease-resistant RNA (fNR) structures, two dumbbell (DB) structures and the terminal 3′ stem loop (3′ SL). They were spaced by adenylate-rich (A-r) segments. This figure also illustrates the location of extra or missing nucleotides that account for the different lengths across the 3′ UTR of DENV. The sequence logos also provide a glance on sequence conservation and nucleotide composition. These data are further described in Tables S1–S3. Nucleotides are color coded (blue, cytosine; green, uracil; yellow, guanine; red, adenine).

(B) Consensus model for the secondary structure of DENV 3′ UTRs. After applying the RNA phylogeny approach, we obtained the secondary interactions for the five conserved RNA structures. Preliminary secondary structures and pseudoknots were predicted through free energy minimization and further refined by covarying base pairs. The statistical support for covarying base pairs was estimated by G-statistics in Rscape software (Figure S5 provides the parameters for the implementation and detailed results). The first 50 non-base-paired nucleotides in the HVR were excluded from the figure.

(C) Phylogeny of fNR structures in the 3′ UTR of dengue viruses. As the sequence logos revealed (A), DENV4 3′ UTR bears only fNR structures; to determine whether this structure shares its most recent common ancestor with the fNR1 or fNR2 structures in other dengue viruses, Bayesian RNA phylogenetics was implemented under PHASE 3.0 software. It included all DENV fNR structure sequences (branches in blue) and a DENV4 NR (branch in red); the fNR structures from Kedougou

Figure 1. *Continued*
virus (KEDV) and Yokose virus (YOKV) were used as outgroup (branch in black). Posterior probabilities are only depicted on relevant nodes.
(D) Ribonucleotide sequence identity on predicted RNA secondary structures in the 3′ UTR of dengue viruses. Sequence identity is color coded according to the heatmap at the bottom of the figure. Highly conserved sites are highlighted in a scale from red to black (site conservation >95%).

et al., 2005, 2008; reviewed by Gebhard et al., 2011). These structures have been termed small hairpin and 3′ end stem loop (3′SL) in domain III and dumbbell (DB) 1 and 2 in domain II. Remarkably, a 30-nucleotide deletion (Δ30) in DB1 generated attenuated DENV1, 3, and 4 but not DENV2 strains (Men et al., 1996; Durbin et al., 2001). Chimeric live attenuated DENV vaccines were designed to bear a Δ30 DENV4 3′ UTR. They appear to be promising live vaccine candidates in clinical trials (Durbin et al., 2013). Vaccination with these vaccine candidates protected human volunteers against live DENV challenge infection (Kirkpatrick et al., 2016).

The proximal segment of the DENV 3′ UTR contains RNA structures that are resistant to host nuclease activity, such as that of 5′-3′ exoribonuclease 1 (Xrn1), resulting in the production of sfRNA during infection (Pijlman et al., 2008). These structures have been referred to with different names in the literature: "stem loops" (Shurtleff et al., 2001; Pijlman et al., 2008; Villordo et al., 2015) or "Xrn1-resistant RNA" (Chapman et al., 2014a). However, crystal structures of homologous RNA sequences from related flaviviruses, Murray Valley encephalitis virus (Chapman et al., 2014b), and Zika virus (Akiyama et al., 2016), revealed a three-way junction RNA folding, rather than an SL structure. More importantly, these RNA structures halt diverse exoribonucleases to produce sfRNA, not only Xrn1 exoribonuclease (MacFadden et al., 2018). Therefore we refer herein to these RNA structures as flaviviral nuclease-resistant RNA (fNR) structures, as initially termed by Pijlman et al. (2008), and to differentiate them from the unrelated nuclease-resistant RNA structures in dianthoviruses (Steckelberg et al., 2018).

As the sequence and hence the RNA structures in the 3′ UTR of the DENV genome appear to be an important determinant of epidemiological fitness, it is possible that this part of the genome contributes to DENV evolution. Here we report the results of a detailed bioinformatic analysis that included all publicly accessible 3′ UTR nucleotide sequences from DENV. We combined free energy minimization and sequence comparative analysis—also known as RNA phylogeny—to estimate secondary and tertiary RNA interactions in the 3′ UTR of the four DENV types. Our RNA phylogenetic and natural selection analyses provide an evolutionary framework for further exploration into the molecular, epidemiological, and clinical consequences of variations in the 3′ UTR of dengue viruses.

RESULTS

Sequence Identity and RNA Structures in the 3′ Untranslated Region of Dengue Viruses

A multiple nucleotide sequence alignment across many homologous ncRNA sequences can depict several features of their evolutionary trajectories. The ncRNA length, sequence identity, and nucleotide composition can unveil insertion or deletion events and conserved GC-rich functional RNA segments. Our analysis on the 3′ UTR of dengue viruses confirmed the existence of substantial differences in the nucleotide sequence composition and length across and within DENV serotypes as previously reported (Shurtleff et al., 2001; Proutski et al., 1999; Men et al., 1996). The 3′ UTR of DENV1 serotype is the longest (mode 465; range 436–475) followed by those of DENV2 (mode 454; range 444–469), DENV3 (Mode 443; range 429–455), and DENV4 (mode 387; range 387–407) (Table S1). Their distal segments (domains II and III) are highly conserved within each DENV type (Figure 1A, see a detailed view in Figures S1–S3). Although a small segment in domain II showed low sequence identity across DENV types (see the consensus alignment in Figure 1A), domain II—as well as the above-mentioned small segment—is well preserved within each DENV type. Comparably, known functional sequences in domains II and III also showed high level of conservation (>99.2%) (Figure S4). These sequences are involved in 3′ UTR to 5′ UTR long-range interactions that are essential for viral genome translation and replication in DENV and other flaviviruses (reviewed by Gebhard et al., 2011 and Brinton and Basu, 2015). In contrast, domain I of 3′ UTR exhibits high genetic variability, including multiple insertions, deletions, and point mutations within and across DENV types. Indeed, its most proximal region—the hypervariable region (HVR)—depicted a significantly poor nucleotide conservation (average identity <89% in all serotypes, $p < 0.001$) and significant adenine enrichment ($p < 0.05$) (Figure 1A, Tables S2 and S3). This finding suggests a lack of folded RNA structures as RNA viruses tend to accumulate adenine in non-base-paired and structurally flexible regions (van Hemert et al., 2013;

Keane et al., 2015). Compared with the HVR, domain I also possesses a semivariable region with a high level of nucleotide conservation (average identity ≥96% in all serotypes) and significantly higher GC content (Figure 1A and Table S2), suggesting the presence of conserved RNA structures across DENV serotypes. These conserved stretches are separated by small adenine-rich sequences, which may serve as spacers to facilitate the proper folding of functional RNA structures. In total, five adenine-rich stretches were found (Table S3); they represent some of the less conserved genomic segments in the 3′ UTR of DENV and appear to space highly conserved functional RNA structures in domains I and II.

To define the secondary RNA structures and tertiary interactions of the 3′ UTR, a "divide, learn, and conquer" approach was adopted (see Transparent Methods). It included (1) identification of conserved RNA structures within each DENV serotype, (2) prediction of preliminary RNA secondary structures from conserved short RNA segments using base-pairing probabilities and thermodynamic methods (Lorenz et al., 2011; Ren et al., 2005), and (3) validation, improvement, and building of consensus RNA structures for DENV 3′ UTR using RNA phylogeny. RNA phylogeny identifies evolutionarily conserved secondary and tertiary RNA structures through nucleotide sequence covariation (Jaeger et al., 1993). To further validate the covariation of base pairs, G-test statistics was also applied to determine whether these covariations occur at a rate higher than phylogenetically expected (Figure S5) (Rivas et al., 2017). The consensus DENV 3′ UTR secondary structures derived from RNA phylogeny (Figure 1B) resembled the DENV2 3′ UTR structure previously obtained by chemical probing (Chapman et al., 2014a), suggesting the validity of our bioinformatics approach. The predicted fNR structures in domain I are also compatible with the known crystal structures, although they substantially differ from the secondary structures suggested by Shurtleff et al. (2001), and to a lesser extent, from the one reported by Villordo et al. (2015).

fNR Structures and DENV Evolution

Interestingly, DENV1, 2, and 3, but not DENV4, bear two fNR structures (Figures 1C and 1D). Among the duplicated fNR structures, the first fNR structure (fNR1) exhibited greater conservation than fNR2, suggesting that fNR1 is preserved because of its nuclease-resistant function for sfRNA production in DENV1, 2, and 3. The downstream fNR (fNR2) structure has thus relatively less constraint to evolve and adapt its structure possibly for additional function. The only fNR structure in DENV4 appeared to be phylogenetically closer to the fNR2 than fNR1 structure of the other DENV serotypes, suggesting that the upstream stretch of nucleotides have been deleted (Figure 1C).

To determine whether the variability of sfRNA structures influences DENV evolution, we examined the selection pressure on each nucleotide position using a maximum likelihood (ML) phylogenetic method (see Transparent Methods). In this approach, the nucleotide substitution rate in each position of the sfRNA sequence was calculated and compared to the synonymous substitution rate in the coding region of each DENV serotype. This ratio models a ζ parameter. When a nucleotide position evolved neutrally, $\zeta \approx 1$ (i.e., equivalent substitutions rates). In contrast, $\zeta > 1$ or $\zeta < 1$ indicate that a given position in the sfRNA has, respectively, a higher or lower substitution rate than the synonymous substitution rate in the coding region of the genome from the same set of DENV strains. This approach thus provides an estimate on whether the higher substitution rate in a given nucleotide position of the ncRNA has contributed to increase the fitness of the bearing DENV strains, i.e., positive selection. If instead the ncRNA nucleotide position remains more conserved when compared with the neutral evolutionary rate of the coding region in DENV genome, it is regarded as the effect of purifying or negative selection. Figure 2A shows the ζ values for every position in the four DENV sfRNAs along with the 95% confidence intervals of the ζ parameter calculated for the coding region of the corresponding DENV genome (depicted as gray zone on the dot plots). Our results show that most nucleotide positions in DENV sfRNA have $\zeta < 1$, suggesting strong negative selection. This observation concurs with the predominant negative selection reported for the 3′ UTR of DENV1 by Wong and Nielsen (2004). It is also consistent with a reported finding that strong purifying selection characterizes the evolution of DENV genomes (Holmes, 2006 and Lequime et al., 2016). However, several nucleotide positions in the fNR and DB structures appear to follow a neutral evolutionary rate ($\zeta \cong 1$) when compared with the coding genome. When compared, these duplicated RNA elements also showed distinct evolutionary rates in some DENV types (Figure 2B), suggesting that DENV sfRNAs undergo unique evolutionary forces in every DENV type. We also observed a slightly higher substitution rate in DB2 than in DB1 in DENV2 lineages, consistent with the findings reported by de Borba et al. (2019), although in both cases the rate is significantly lower than the neutral substitution rate. Although purifying selection prevails on DB evolution, the DB2 structures appear to have diverged

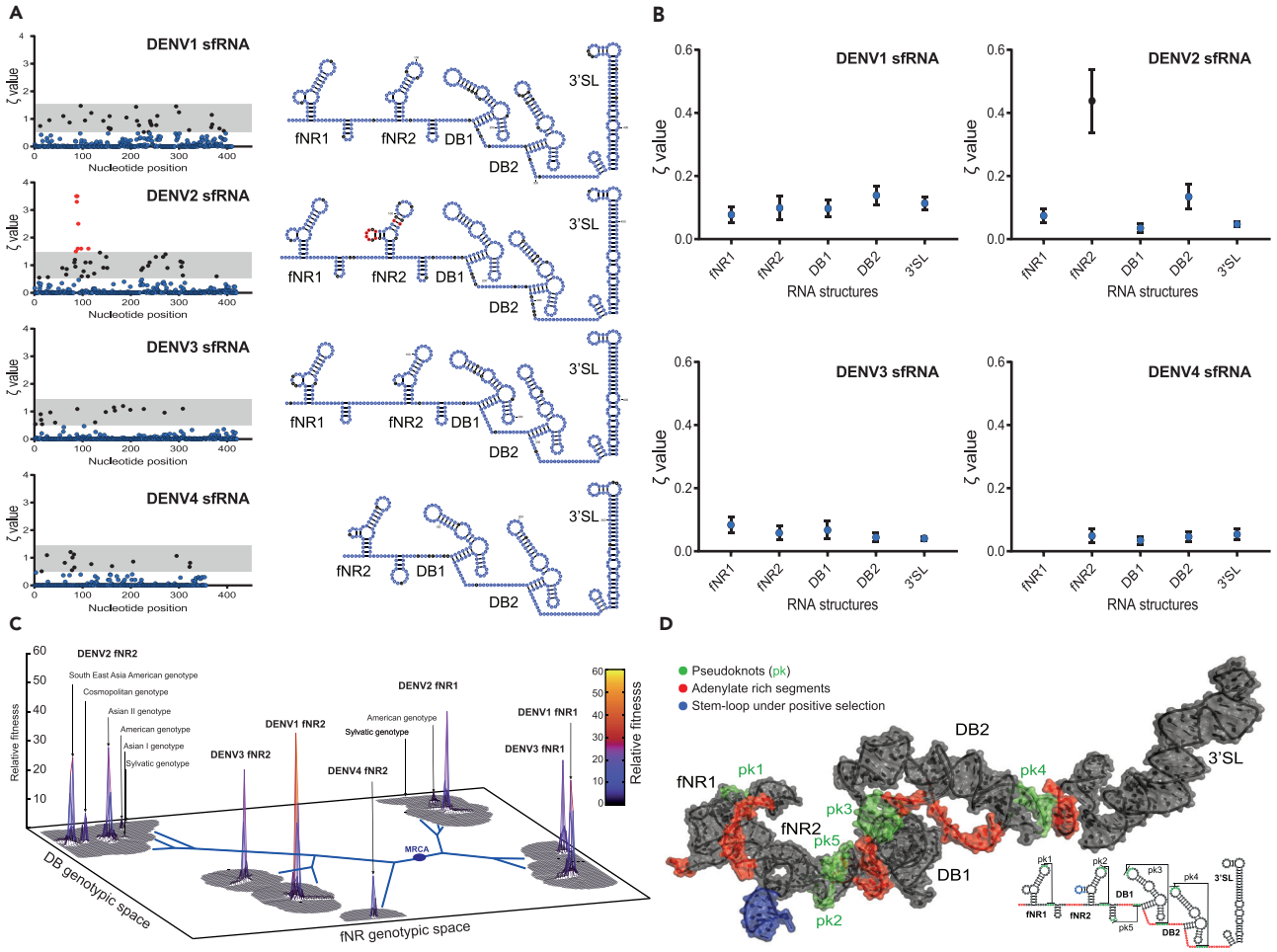


Figure 2. DENV2 sfRNA Possesses a Highly Evolving RNA Structure

(A) Site-specific quantification of natural selection on the sfRNA from dengue viruses. A maximum-likelihood method was applied to detect the action of natural selection in DENV sfRNAs in a site-by-site basis. A ζ parameter and its 95% confidence interval (CI) across the full genome determined whether a nucleotide position underwent negative selection (blue dots), positive selection (red dots), or neutral evolution (black dots) in the ncRNA sequence. On the left, dot plots depict the zeta values for all DENV sfRNAs. The 95% CI is shown as gray zone on the dot plots. The 95% CI slightly varied across DENV genomes. DENV1 = (0.513; 1.469), DENV2 = (0.547; 1.473), DENV3 = (0.532; 1.451), DENV4 = (0.506; 1.437). On the right side, the ribonucleotide positions in the DENV sfRNA secondary structures are color coded accordingly. Pseudoknots are not shown for the sake of simplicity.

(B) Quantification of natural selection on the RNA structures of DENV sfRNAs. The ζ parameter for every RNA structure and their standard errors are shown for every DENV type. The ζ parameter for every RNA structure is color coded based on whether they underwent negative selection (blue dots), positive selection (red dots), or neutral evolution (black dots).

(C) Fitness landscape of DENV sfRNA as determined by fNR and DB sequence abundance. A fitness landscape is a function that assigns to every genotype a numerical value proportional to its fitness. It involves a vast 2D genotypic space and a fitness value (z axis). The 2D genotypic space for the sfRNA fitness landscape was resolved in X axis by the fNR nucleotide sequences and in the Y axis by the DBs (full length Domain II) sequences, whereas their normalized combined abundance (copy numbers) across DENV 3′ UTR sequence alignments served as relative fitness value. It allowed us to elucidate the ability of these RNA structures to contribute to genome survival and to establish distinct evolutionary trajectories in sfRNA evolution.

(D) 3D simulation of DENV2 sfRNA. By combining predicted base pairing and comparative RNA modeling, we obtained an *in silico* 3D model of DENV2 sfRNA. Pseudoknots, adenylate-rich regions, and the highly evolving hairpins are highlighted in colors as in the secondary RNA structures at the bottom right of the figure.

across DENV types through some base pair covariations, explaining the low sequence identity in a small segment of domain II. Strikingly, some nucleotide positions in the fNR2 structure of DENV2 could have undergone positive selection with ζ > 1 (Figure 2A). This RNA structure also shows higher evolutionary rate than other RNA structures in DENV2 sfRNA (Figure 2B), suggesting that substitutions in fNR2 may confer competitive advantage for DENV2 strains. These results are uniformly consonant with the greater

dispersion and diversity of fitness peaks that appeared in our DENV sfRNA fitness landscape (Figure 2C). The roughness on the surface of the DENV sfRNA fitness landscape indicates emergence or existence of multiple fNR/DB fitness peaks across the four DENV types and their divergence after fNR duplication. Although the DB divergence across DENV types resolves the fitness peaks in the DB dimension, the fNR sequences explore a larger genotypic space and generate more fitness peak than the DBs. More importantly, the fNR2 genotypic space is populated by more dispersed and isolated peaks of fitness. It further confirmed the greater evolvability and divergence of fNR2 when compared with fNR1 sequences and the ability of fNR2 sequences to contribute to DENV type 2 divergence into different genotypes.

To further assess the role of DENV2 3′ UTR evolution, an ML phylogenetic tree was constructed using a nucleotide substitution model of evolution for ncRNA, based on the consensus secondary RNA structure of DENV2 sfRNA (see Transparent Methods). The phylogenetic tree (Figure 3A) segregated the DENV2 sfRNA sequences into six clades, consistent with the six genotypes that characterize DENV2 evolution (reviewed by Chen and Vasilakis, 2011).

Remarkably, the positively selected hairpin in the fNR2 structure differed in nucleotide composition and structure across DENV2 genotypes, despite being relatively conserved within genotypes (Figure 3A). Most genotypes are rich in adenine in this hairpin structure except the American genotype, which has mostly uracil. It is noteworthy that the American genotype has shown poor epidemiological fitness and has now been completely displaced by other DENV2 genotypes in many parts of the world. Likewise, analysis of DENVs derived from epidemiological studies (Table 1) also showed that four clade replacement episodes that resulted in greater or less than expected dengue incidence involved nucleotide substitutions in fNR2 structures (Figure 3B).

DISCUSSION

The identification of 3′ UTR structure and sfRNA production as having functional importance in determining viral fitness is of major interest in both experimental and epidemiological settings. The frequent emergence of DENV strains with insertions, deletions, and point mutations in their 3′ UTR (Zhou et al., 2006; Pankhong et al., 2009; de Castro et al., 2013; Dash et al., 2015) and the differences in nucleotide lengths underscores the need for improved understanding of this part of the DENV genome. Given that the sfRNA is an ncRNA, its influence on DENV fitness and evolution must be understood in the context of its RNA structures. RNA phylogeny provides a bioinformatic approach to glean insights to direct further mechanistic investigations. Furthermore, a phylogenetic based estimation of substitution rate using ncRNA model of nucleotide substitutions coupled with normalization by the substitution rate in the coding genome enabled us to (1) overcome the bias that dataset size can introduce in sequence identity (conservation) analysis, (2) avoid the misleading interpretation of "representative" sequences, and (3) exploit the growing sequencing databases to gain insight into the ncRNA evolution of a widely spread virus.

Available sequence data indicate that the DBs, the terminal 3′ SL, and long-range interacting sequences in domains II and III are highly conserved. Their role in flaviviral replication may enforce a strong purifying selection on these 3′ UTR domains. The integrity of DB structures also contributes to increased sfRNA production, as the 30-nucleotide deletion in DB1 decreases sfRNA production and increases type I IFN susceptibility of the live attenuated DENV vaccine candidates (Bustos-Arriaga et al., 2018). Likewise, fNR structures in the DENV 3′ UTR are relatively well conserved. All the nucleotide interactions that confer nuclease resistance are highly conserved. In addition, any nucleotide substitution in base-paired positions of the fNR or DB structures is often accompanied by compensatory mutation to maintain structural integrity. Although some nucleotide positions in these duplicated RNA elements seem to follow the neutral evolutionary rate in DENV and the evolvability of these RNA structures varied across DENV types (Figure 2B), as a rule of thumb, purifying selection remains the prevailing force in the evolution of RNA structures in DENV 3′ UTRs. Along this line of thought, the finding of a positive selection in the fNR2 structure of DENV 3′ UTR is thus intriguing. fNR2 mutations may emerge during mosquito infection and subsequently be selected in human infections, owing to their contribution to replicative fitness (Villordo et al., 2015; Filomatori et al., 2017). Certainly, serial passaging of DENV2 in *Aedes albopictus*-derived C6/36 cells resulted in multiple mutations in its fNR2 and production of different sfRNA species (Villordo et al., 2015). More detailed experimental investigation revealed that mutations in the positively selected hairpin did not impair DENV replicative fitness in mosquitoes or mammalian cell lines (Villordo et al., 2015). We therefore

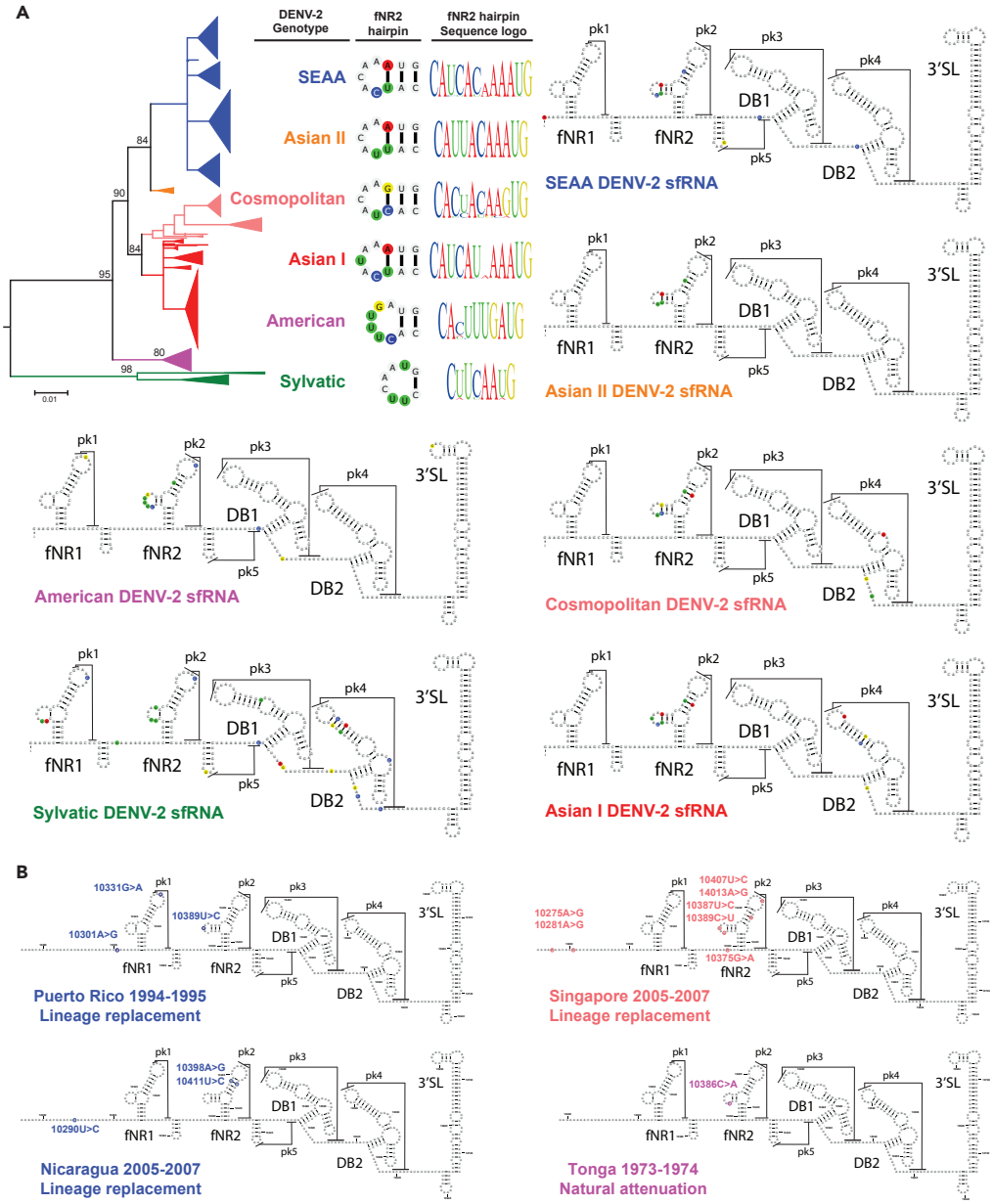


Figure 3. Nucleotide Substitutions in DENV2 sfRNA Are Associated with DENV2 Speciation and Increased Epidemic Potential

(A) Phylogenetics and nucleotide substitutions in the sfRNA of DENV2 strains. A maximum likelihood phylogenetic tree from DENV2 sfRNA was built using PHASE 3.0 software. This software applies an RNA structure-based approach to construct phylogenies of non-coding RNAs. The highly evolving hairpin in DENV2 sfRNA exhibited distinct nucleotide composition and structure across DENV2 genotypes. Hairpin secondary structures and sequence logos are shown next to the corresponding branch for the DENV2 genotypes in the phylogenetic tree. The consensus secondary structures for the sfRNA in all the DENV2 genotypes are also shown. Genotype-specific nucleotide substitutions are highlighted in colors. (B) Epidemic DENV2 strains underwent nucleotide substitution in the highly evolving NR2 of DENV2 sfRNA. Location of nucleotide substitutions are shown in dominant strains that have been involved in three DENV2 clade replacements and a natural attenuation event.

hypothesize that the mosquito-derived fNR2 variability has been selected in subsequent human infections in distinct geographical locations, possibly based on their ability to bind host proteins for the suppression of antiviral immune activation. Previous biochemical work has shown that mutations in the highly evolving

Events	Geographical Location	Year	Serotype/Genotype	Context	Reference
Clade replacement	Puerto Rico[a]	1994	DENV2 SEAA	Outbreak	McElroy et al. (2011)
	Nicaragua[a]	2005	DENV2 SEAA	Outbreak	OhAinle et al. (2011)
	Singapore[a]	2005	DENV2 Cosmopolitan	Outbreak	Lee et al. (2012)
	Thailand	1990s	DENV1	Outbreaks	Teoh et al. (2013)
	Brazil	1990s	DENV1 Genotype I	Outbreaks	Carneiro et al. (2012)
	Mexico	2006 1999	DENV1 Genotype III DENV2 SEAA	Several outbreaks	Carrillo-Valenzo et al. (2010)
	India[a]	2009–2011	DENV1 genotype III	Outbreaks	Dash et al. (2015)
	Malaysia	1987 1997 2004	DENV1 genotype I	Recurring outbreaks	Dupont-Rouzeyrol et al. (2014)
	Tonga Island[a]	1974	DENV2 American	Reduced severity	Steel et al. (2010)
Genotype replacement	Sri Lanka[a]	1989	DENV3 Genotype I > III	DHF emergence	Silva et al. (2008) Manakkadan et al. (2013)
	Myanmar[a]	1974–2002	DENV1 Genotype I > III	Several outbreaks	Myat Thu et al. (2015)
	Indian Subcontinent[a]	1971	DENV2 American > cosmopolitan	Several outbreaks	Kumar et al. (2010)
	Americas[a]	1983	DENV2 American > SEAA	DHF emergence	Mir et al. (2014)

Table 1. List of Epidemiological Events Associated with Increased DENV Epidemiological Fitness

An extensive literature revision on DENV epidemiology revealed at least 13 events associated with increased DENV epidemiological fitness.

[a]Nucleotide substitutions in the 3′ UTR were reported in nine of those epidemic DENV strains.

nucleotide positions do not compromise the nuclease-resistant activity of fNRs (Chapman et al., 2014a). Therefore a mechanism other than increased nuclease-resistant activity would better serve to explain the driving force for positive selection on this hairpin. As the *in silico* modeling of DENV sfRNA would suggest (Figure 2D), the positive selected hairpin appears to be fully solvent exposed and prone to RNA-protein interactions. Indeed, Bidet et al. (2014) showed that the fNR2 structure interacts with CAPRIN G3BP1 and G3BP2 proteins, mediating the sfRNA-induced repression of IFN-stimulated mRNAs in human liver-derived Huh7 cells. Moreover, mutations in the fNR2 structures produced higher replicative fitness in *Aedes albopictus* compared with the corresponding wild-type DENV2 (Filomatori et al., 2017). Collectively, these findings suggest a strong evolutionary pressure on DENV2 fNR2 structures. In addition, the consistent concordance between previously reported experimental data and our bioinformatics findings highlights the robustness of ML method developed by Wong and Nielsen (2004).

Viral genomes can undergo different mutations, including but not limited to nucleotide substitutions, insertion, deletions, rearrangements, and reassortments. Among them, duplications of genomic elements are unique as only duplications increase the genome's ability to withstand mutations, i.e., mutational robustness (Wagner, 2008). Immediately after duplication, the redundant elements exert the same function and experience relaxed selection during a brief period of their early history. Subsequently, mutations accumulate in one of the redundant elements, leading to its structural or functional elimination or to the divergence in the elements' function. Thus the duplication and subsequent divergence of genomic elements can increase mutational robustness and favor evolutionary innovation, allowing the corresponding fixation of duplicated elements in the genome. This phenomenon has been well documented for duplicated genes, which have withstood more nucleotide substitutions and diverged more than their single-copy counterparts in their early evolutionary history (Lynch and Conery, 2000). In a similar manner, duplication of RNA elements can open the genotypic space for evolutionary innovation and emergence of peaks in the fitness landscape. Given our bioinformatics findings and other available experimental evidence, we propose that

the duplicated fNR structures in domain I currently embody two functionally distinct RNA segments: the first NR structure is conserved to enable sfRNA production. In contrast, the downstream NR structure is relatively free to evolve and may be selected based on advantageous RNA-protein interactions in human or mosquito cells for increased fitness. This evolutionary model would agree with the reduced sfRNA production and transmission fitness that fNR1 mutations caused in DENV2 strains (Pompon et al., 2017), and the increased replicative, transmission, and epidemiologic fitness that fNR2 mutations conferred to some DENV2 genotypes and more specifically in some dominant DENV2 strains. Indeed, Cologna and Rico-Hesse (2003) cloned the 3′ UTR of the American genotype into a Southeast Asian -American (SEAA) DENV2 and found small viral plaques in Vero cells and slower growth kinetics in both mosquito and human cells (also Anderson and Rico-Hesse, 2006), which are phenotypes more congruent with the American than SEAA DENV2 genotype.

Our proposed model also explains the lack of positive selection in DENV4 as these viruses only possess one fNR structure. It, however, raises questions on why no positive selection was detected on the 3′ UTR of either DENV1 or 3. We offer several interpretations. First, if we assume that the distinct adaptation of the fNR2 in interspecies transmission—high sequence diversity in mosquito infection and sequence bottleneck in human infection—(Villordo et al., 2015; Filomatori et al., 2017) is happening in all DENVs, our bioinformatics data would indicate that a stronger purifying selection occurs in the 3′ UTR of DENV1 and 3 during human infection when compared with DENV2. This is especially true because most publicly available DENV full-genome sequence data were derived from clinical isolates; few full-genome sequences from infected mosquitoes have been reported. A second and interesting scenario would suggest that the proposed model for the evolution of DENV2 sfRNA and its distinct fNR adaptation do not occur in other DENV types. This second postulate would help us understand why previous studies using RNA sequencing observed mutational hotspot in the 3′ UTR of DENV2 but not DENV1 after replication in mosquitoes (Sessions et al., 2015; Sim et al., 2015). Experimental studies will be needed to test the validity of these postulates.

Notwithstanding the need for mechanistic validation, we suspect that fNR duplication has contributed to shape the overall divergence of dengue viruses. Although the duplication of genomic elements can provide mutational robustness and evolvability, the stochastic deletion of such elements may play a significant role in the passive origin of new species (Lynch and Conery, 2000). Thus if the fNR duplication occurred early in DENV evolutionary history, as the RNA phylogenetic analysis suggested, it is likely that the later fNR1 deletion in DENV4 imposed an evolutionary constraint in DENV4 lineage, limiting its adaptability to infect new "urbanized" hosts and forcing a sympatric speciation and its greater divergence. It concurs with weaker DENV4 transmission and replicative and epidemiological fitness when compared with other DENV types. Indeed DENV4 has shown the following characteristics: lower human-to-mosquito transmission rates (Duong et al., 2015), lower infectivity in both field (Nguyet et al., 2013) and experimental settings (Gubler and Rosen, 1976; Gubler et al., 1979; Mitchell et al., 1987), lower infectivity and shorter viremia in non-human primates (Kraiselburd et al., 1985; Althouse et al., 2014), lower and shorter viremia levels in human infections (Gubler et al., 1981; Nguyet et al., 2013), and slower global spread (Messina et al., 2014) among the different types of DENVs. This constraint might have been overcome through antibody-dependent enhancement in primates or competitive advantage in vector DENV co-infections in the current allopatric DENV distribution (Halstead, 2014; Vazeille et al., 2016). It would also help to explain why DENV4 has shown reduced epidemic potential during its global spread in the last decades and why only an additional 30-nucleotide deletion in the 3′ UTR is required to generate a complete attenuated phenotype in DENV4 and in recombinant DENV1 to 3 strains bearing a Δ30rDENV4 -3′ UTR (Durbin et al., 2001 and 2013).

Collectively, our findings suggest that 3′ UTR evolution and sfRNA production are important determinants of DENV adaptation, survival, and epidemiological fitness.

Limitation of the Study

This study has benefitted from the wealth of information embedded in the large number of DENV genomic sequences that have been submitted to NCBI GenBank in the last four decades. However, the viral sample collection, processing, and sequencing methodologies might have introduced some bias in the raw data for this study. For instance, most of the currently available DENV genomes were obtained from human serum samples and very few of them were isolated from mosquitoes, as indicated in the Transparent Methods section. A larger sampling of mosquitoes could provide a more nuanced picture in terms of

sequence identity, variability, and substitution rates, when compared with what the authors have observed in this study. Finally, this article is primarily hypothesis generating by adopting an evolutionary interpretation of the bioinformatic data. Experimental studies to systematically test these hypotheses will be needed to validate the notions proposed herein.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2019.05.019.

ACKNOWLEDGMENTS

This work was supported by the National Medical Research Council of Singapore (NMRC/CIRG/1462/2016). The authors would also like to thank Drs. Vijaykrishnan Dhanasekaran and Mariano Garcia-Blanco for their invaluable advice throughout this work. The authors also thank Eneida Hatcher and NCBI GenBank team members for handling and reviewing the suggested DENV 3′ UTR annotation. This annotation is now available in the four DENV reference sequences (GenBank: NC_001477, NC_001475, NC_002640, and NC_001474).

AUTHOR CONTRIBUTIONS

Conceptualization, E.F. and E.E.O.; Methodology, E.F; Software, E.F.; Investigation, E.F. and E.E.O; Formal Analysis, E.F.; Data Curation, E.F.; Writing – Original Draft, E.F.; Writing – Review & Editing, E.F. and E.E.O.; Visualization, E.F.; Funding Acquisition, E.E.O.; Resources, E.F. and E.E.O.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.

Received: March 26, 2018
Revised: February 2, 2019
Accepted: May 13, 2019
Published: June 28, 2019

REFERENCES

Akiyama, B.M., Laurence, H.M., Massey, A.R., Costantino, D.A., Xie, X., Yang, Y., Shi, P.Y., Nix, J.C., Beckham, J.D., and Kieft, J.S. (2016). Zika virus produces noncoding RNAs using a multi–pseudoknot structure that confounds a cellular exonuclease. Science *354*, 1148–1152.

Althouse, B.M., Durbin, A.P., Hanley, K.A., Halstead, S.B., Weaver, S.C., and Cummings, D.A. (2014). Viral kinetics of primary dengue virus infection in non-human primates: a systematic review and individual pooled analysis. Virology *452–453*, 237–246.

Alvarez, D.E., De Lella Ezcurra, A.L., Fucito, S., and Gamarnik, A.V. (2005). Role of RNA structures present at the 3′UTR of dengue virus on translation, RNA synthesis, and viral replication. Virology *339*, 200–212.

Alvarez, D.E., Filomatori, C.V., and Gamarnik, A.V. (2008). Functional analysis of dengue virus cyclization sequences located at the 5′ and 3′UTRs. Virology *375*, 223–235.

Anderson, J.R., and Rico–Hesse, R. (2006). Aedes aegypti vectorial capacity is determined by the infecting genotype of dengue virus. Am. J. Trop. Med. Hyg. *75*, 886–892.

Bhatt, S., Gething, P.W., Brady, O.J., Messina, J.P., Farlow, A.W., Moyes, C.L., Drake, J.M., Brownstein, J.S., Hoen, A.G., Sankoh, O., et al. (2013). The global distribution and burden of dengue. Nature *496*, 504–507.

Bidet, K., Dadlani, D., and Garcia–Blanco, M.A. (2014). G3BP1, G3BP2 and CAPRIN1 are required for translation of interferon stimulated mRNAs and are targeted by a dengue virus non–coding RNA. PLoS Pathog. *10*, e1004242.

Brinton, M.A., and Basu, M. (2015). Functions of the 3′ and 5′ genome RNA regions of members of the genus Flavivirus. Virus Res. *206*, 108–119.

Bustos-Arriaga, J., Gromowski, G.D., Tsetsarkin, K.A., Firestone, C.Y., Castro-Jiménez, T., Pletnev, A.G., Cedillo-Barrón, L., and Whitehead, S.S. (2018). Decreased accumulation of subgenomic RNA in human cells infected with vaccine candidate DEN4Δ30 increases viral susceptibility to type I interferon. Vaccine *36*, 3460–3467.

Carneiro, A.R., Cruz, A.C., Vallinoto, M., Melo Dde, V., Ramos, R.T., Medeiros, D.B., Silva, E.V., and Vasconcelos, P.F. (2012). Molecular characterisation of dengue virus type 1 reveals lineage replacement during circulation in Brazilian territory. Mem. Inst. Oswaldo Cruz *107*, 805–812.

Carrillo-Valenzo, E., Danis-Lozano, R., Velasco-Hernández, J.X., Sánchez-Burgos, G., Alpuche, C., López, I., Rosales, C., Baronti, C., de Lamballerie, X., Holmes, E.C., and Ramos-Castañeda, J. (2010). Evolution of dengue virus in Mexico is characterized by frequent lineage replacement. Arch. Virol. *155*, 1401–1412.

Chapman, E.G., Moon, S.L., Wilusz, J., and Kieft, J.S. (2014a). RNA structures that resist degradation by Xrn1 produce a pathogenic Dengue virus RNA. Elife *3*, e01892.

Chapman, E.G., Costantino, D.A., Rabe, J.L., Moon, S.L., Wilusz, J., Nix, J.C., and Kieft, J.S. (2014b). The structural basis of pathogenic subgenomic flavivirus RNA (sfRNA) production. Science *344*, 307–310.

Chen, R., and Vasilakis, N. (2011). Dengue–quo tu et quo vadis? Viruses *3*, 1562–1608.

Cologna, R., and Rico-Hesse, R. (2003). American genotype structures decrease dengue virus output from human monocytes and dendritic cells. J. Virol. *77*, 3929–3938.

Dash, P.K., Sharma, S., Soni, M., Agarwal, A., Sahni, A.K., and Parida, M. (2015). Complete genome sequencing and evolutionary phylogeography analysis of Indian isolates of Dengue virus type 1. Virus Res. *195*, 124–134.

de Borba, L., Villordo, S.M., Marsico, F.L., Carballeda, J.M., Filomatori, C.V., Gebhard, L.G., Pallarés, H.M., Lequime, S., Lambrechts, L., Sánchez Vargas, I., et al. (2019). RNA structure duplication in the dengue virus 3′ UTR: redundancy or host specificity? MBio *10*, e02506–18.

de Castro, M.G., de Nogueira, F.B., Nogueira, R.M., Lourenço-de-Oliveira, R., and dos Santos, F.B. (2013). Genetic variation in the 3′ untranslated region of dengue virus serotype 3 strains isolated from mosquitoes and humans in Brazil. Virol. J. *10*, 3.

Dupont-Rouzeyrol, M., Aubry, M., O'Connor, O., Roche, C., Gourinat, A.C., Guigon, A., Pyke, A., Grangeon, J.P., Nilles, E., Chanteau, S., et al. (2014). Epidemiological and molecular features of dengue virus type-1 in New Caledonia, South Pacific, 2001-2013. Virol. J. *11*, 61.

Duong, V., Lambrechts, L., Paul, R.E., Ly, S., Lay, R.S., Long, K.C., Huy, R., Tarantola, A., Scott, T.W., Sakuntabhai, A., and Buchy, P. (2015). Asymptomatic humans transmit dengue virus to mosquitoes. Proc. Natl. Acad. Sci. U S A *112*, 14688–14693.

Durbin, A.P., Karron, R.A., Sun, W., Vaughn, D.W., Reynolds, M.J., Perreault, J.R., Thumar, B., Men, R., Lai, C.J., Elkins, W.R., et al. (2001). Attenuation and immunogenicity in humans of a live dengue virus type–4 vaccine candidate with a 30 nucleotide deletion in its 3′-untranslated region. Am. J. Trop. Med. Hyg. *65*, 405–413.

Durbin, A.P., Kirkpatrick, B.D., Pierce, K.K., Elwood, D., Larsson, C.J., Lindow, J.C., Tibery, C., Sabundayo, B.P., Shaffer, D., Talaat, K.R., et al. (2013). Single dose of any of four different live attenuated tetravalent dengue vaccines is safe and immunogenic in flavivirus–naive adults: a randomized, double–blind clinical trial. J. Infect. Dis. *207*, 957–965.

Filomatori, C.V., Carballeda, J.M., Villordo, S.M., Aguirre, S., Pallarés, H.M., Maestre, A.M., Sánchez-Vargas, I., Blair, C.D., Fabri, C., Morales, M.A., et al. (2017). Dengue virus genomic variation associated with mosquito adaptation defines the pattern of viral non-coding RNAs and fitness in human cells. PLoS Pathog. *13*, e1006265.

Gebhard, L.G., Filomatori, C.V., and Gamarnik, A.V. (2011). Functional RNA elements in the dengue virus genome. Viruses *3*, 1739–1756.

Gubler, D.J., Nalim, S., Tan, R., Saipan, H., and Sulianti Saroso, J. (1979). Variation in susceptibility to oral infection with dengue viruses among geographic strains of Aedes aegypti. Am. J. Trop. Med. Hyg. *28*, 1045–1052.

Gubler, D.J., and Rosen, L. (1976). Variation among geographic strains of Aedes albopictus in susceptibility to infection with dengue viruses. Am. J. Trop. Med. Hyg. *25*, 318–325.

Gubler, D.J., Suharyono, W., Tan, R., Abidin, M., and Sie, A. (1981). Viraemia in patients with naturally acquired dengue infection. Bull. World Health Organ. *59*, 623–630.

Hadinegoro, S.R., Arredondo–García, J.L., Capeding, M.R., Deseda, C., Chotpitayasunondh, T., Dietze, R., Muhammad Ismail, H.I., Reynales, H., Limkittikul, K., Rivera–Medina, D.M., et al. (2015). Efficacy and long-term safety of a dengue vaccine in regions of endemic disease. N. Engl. J. Med. *373*, 1195–1206.

Halstead, S.B. (2014). Dengue antibody-dependent enhancement: knowns and unknowns. Microbiol. Spectr. *2*, https://doi.org/10.1128/microbiolspec.AID-0022-2014.

Holmes, E.C. (2006). The evolutionary biology of dengue virus. Novartis Found Symp. *277*, 177–187.

Jaeger, J.A., SantaLucia, J., and Tinoco, I. (1993). Determination of RNA structures and thermodynamics. Annu. Rev. Biochem. *62*, 255–287.

Keane, S.C., Heng, X., Lu, K., Kharytonchyk, S., Ramakrishnan, V., Carter, G., Barton, S., Hosic, A., Florwick, A., Santos, J., et al. (2015). RNAstructure. Structure of the HIV-1 RNA packaging signal. Science *348*, 917–921.

Kirkpatrick, B.D., Whitehead, S.S., Pierce, K.K., Tibery, C.M., Grier, P.L., Hynes, N.A., Larsson, C.J., Sabundayo, B.P., Talaat, K.R., Janiak, A., et al. (2016). The live attenuated dengue vaccine TV003 elicits complete protection against dengue in a human challenge model. Sci. Transl. Med. *8*, 330–336.

Kraiselburd, E., Gubler, D.J., and Kessler, M.J. (1985). Quantity of dengue virus required to infect rhesus monkeys. Trans. R. Soc. Trop. Med. Hyg. *79*, 248–251.

Kumar, S.R., Patil, J.A., Cecilia, D., Cherian, S.S., Barde, P.V., Walimbe, A.M., Yadav, P.D., Yergolkar, P.N., Shah, P.S., Padbidri, V.S., et al. (2010). Evolution, dispersal and replacement of American genotype dengue type 2 viruses in India (1956-2005): selection pressure and molecular clock analyses. J. Gen. Virol. *91*, 707–720.

Lee, K.S., Lo, S., Tan, S.S., Chua, R., Tan, L.K., Xu, H., and Ng, L.C. (2012). Dengue virus surveillance in Singapore reveals high viral diversity through multiple introductions and in situ evolution. Infect. Genet. Evol. *12*, 77–85.

Lequime, S., Fontaine, A., Ar Gouilh, M., Moltini-Conclois, I., and Lambrechts, L. (2016). Genetic drift, purifying selection and vector genotype shape dengue virus intra-host genetic diversity in mosquitoes. PLoS Genet. *12*, e1006111.

Lynch, M., and Conery, J.S. (2000). The evolutionary fate and consequences of duplicate genes. Science *290*, 1151–1155.

Lorenz, R., Bernhart, S.H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol. Biol. *6*, 26.

MacFadden, A., O'Donoghue, Z., Silva, P.A.G.C., Chapman, E.G., Olsthoorn, R.C., Sterken, M.G., Pijlman, G.P., Bredenbeek, P.J., and Kieft, J.S. (2018). Mechanism and structural diversity of exoribonuclease-resistant RNA structures in flaviviral RNAs. Nat. Commun. *9*, 119.

Manakkadan, A., Joseph, I., Prasanna, R.R., Kunju, R.I., Kailas, L., and Sreekumar, E. (2013). Lineage shift in Indian strains of Dengue virus serotype-3 (Genotype III), evidenced by detection of lineage IV strains in clinical cases from Kerala. Virol. J. *10*, 37.

Manokaran, G., Finol, E., Wang, C., Gunaratne, J., Bahl, J., Ong, E.Z., Tan, H.C., Sessions, O.M., Ward, A.M., Gubler, D.J., Harris, E., Garcia-Blanco, M.A., and Ooi, E.E. (2015). Dengue subgenomic RNA binds TRIM25 to inhibit interferon expression for epidemiological fitness. Science *350*, 217–221.

McElroy, K.L., Santiago, G.A., Lennon, N.J., Birren, B.W., Henn, M.R., and Muñoz-Jordán, J.L. (2011). Endurance, refuge, and reemergence of dengue virus type 2, Puerto Rico, 1986-2007. Emerg. Infect. Dis. *17*, 64–71.

Men, R., Bray, M., Clark, D., Chanock, R.M., and 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*, 3930–3937.

Messina, J.P., Brady, O.J., Pigott, D.M., Golding, N., Kraemer, M.U., Scott, T.W., Wint, G.R., Smith, D.L., and Hay, S.I. (2015). The many projected futures of dengue. Nat. Rev. Microbiol. *13*, 230–239.

Messina, J.P., Brady, O.J., Scott, T.W., Zou, C., Pigott, D.M., Duda, K.A., Bhatt, S., Katzelnick, L., Howes, R.E., Battle, K.E., et al. (2014). Global spread of dengue virus types: mapping the 70 year history. Trends Microbiol. *22*, 138–146.

Mir, D., Romero, H., Fagundes de Carvalho, L.M., and Bello, G. (2014). Spatiotemporal dynamics of DENV-2 Asian-American genotype lineages in the Americas. PLoS One *9*, e98519.

Mitchell, C.J., Miller, B.R., and Gubler, D.J. (1987). Vector competence of Aedes albopictus from Houston, Texas, for dengue serotypes 1 to 4, yellow fever and Ross River viruses. J. Am. Mosq. Control Assoc. *3*, 460–465.

Myat Thu, H., Lowry, K., Jiang, L., Hlaing, T., Holmes, E.C., and Aaskov, J. (2015). Lineage extinction and replacement in dengue type 1 virus populations are due to stochastic events rather than to natural selection. Virology *336*, 163–172.

Nguyet, M.N., Duong, T.H., Trung, V.T., Nguyen, T.H., Tran, C.N., Long, V.T., Dui le, T., Nguyen, H.L., Farrar, J.J., Holmes, E.C., et al. (2013). Host and viral features of human dengue cases shape the population of infected and infectious Aedes aegypti mosquitoes. Proc. Natl. Acad. Sci. U S A *110*, 9072–9077.

OhAinle, M., Balmaseda, A., Macalalad, A.R., Tellez, Y., Zody, M.C., Saborío, S., Nuñez, A., Lennon, N.J., Birren, B.W., Gordon, A., et al. (2011). Dynamics of dengue disease severity determined by the interplay between viral genetics and serotype-specific immunity. Sci. Transl. Med. *3*, 114–128.

Pankhong, P., Weiner, D.B., Ramanathan, M.P., Nisalak, A., Kalayanarooj, S., Nimmannitya, S., and Attatippaholkun, W. (2009). Molecular genetic relationship of the 3′ untranslated region

among Thai dengue-3 virus, Bangkok isolates, during 1973-2000. DNA Cell Biol. *28*, 481–491.

Pijlman, G.P., Funk, A., Kondratieva, N., Leung, J., Torres, S., van der Aa, L., Liu, W.J., Palmenberg, A.C., Shi, P.Y., Hall, R.A., and Khromykh, A.A. (2008). A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. Cell Host Microbe *4*, 579–591.

Pompon, J., Manuel, M., Ng, G.K., Wong, B., Shan, C., Manokaran, G., Soto-Acosta, R., Bradrick, S.S., Ooi, E.E., Missé, D., et al. (2017). Dengue subgenomic flaviviral RNA disrupts immunity in mosquito salivary glands to increase virus transmission. PLoS Pathog. *13*, e1006535.

Proutski, V., Gritsun, T.S., Gould, E.A., and Holmes, E.C. (1999). Biological consequences of deletions within the 3′-untranslated region of flaviviruses may be due to rearrangements of RNA secondary structure. Virus Res. *64*, 107–123.

Ren, J., Rastegari, B., Condon, A., and Hoos, H.H. (2005). HotKnots: heuristic prediction of RNA secondary structures including pseudoknots. RNA *11*, 1494–1504.

Rivas, E., Clements, J., and Eddy, S.R. (2017). A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. Nat. Methods *14*, 45–48.

Sessions, O.M., Wilm, A., Kamaraj, U.S., Choy, M.M., Chow, A., Chong, Y., Ong, X.M., Nagarajan, N., Cook, A.R., and Ooi, E.E. (2015). Analysis of dengue virus genetic diversity during human and mosquito infection reveals genetic constraints. PLoS Negl. Trop. Dis. *9*, e0004044.

Silva, R.L., de Silva, A.M., Harris, E., and MacDonald, G.H. (2008). Genetic analysis of Dengue 3 virus subtype III 5′ and 3′ non-coding regions. Virus Res. *135*, 320–325.

Shurtleff, A.C., Beasley, D.W., Chen, J.J., Ni, H., Suderman, M.T., Wang, H., Xu, R., Wang, E., Weaver, S.C., Watts, D.M., et al. (2001). Genetic variation in the 3′ non–coding region of dengue viruses. Virology *281*, 75–87.

Sim, S., Aw, P.P., Wilm, A., Teoh, G., Hue, K.D., Nguyen, N.M., Nagarajan, N., Simmons, C.P., and Hibberd, M.L. (2015). Tracking dengue virus intra-host genetic diversity during human-to-mosquito transmission. PLoS Negl. Trop. Dis. *9*, e0004052.

Steckelberg, A.L., Akiyama, B.M., Costantino, D.A., Sit, T.L., Nix, J.C., and Kieft, J.S. (2018). A folded viral noncoding RNA blocks host cell exoribonucleases through a conformationally dynamic RNA structure. Proc. Natl. Acad. Sci. U S A *115*, 6404–6409.

Steel, A., Gubler, D.J., and Bennett, S.N. (2010). Natural attenuation of dengue virus type-2 after a series of island outbreaks: a retrospective phylogenetic study of events in the South Pacific three decades ago. Virology *405*, 505–512.

Teoh, B.T., Sam, S.S., Tan, K.K., Johari, J., Shu, M.H., Danlami, M.B., Abd-Jamil, J., MatRahim, N., Mahadi, N.M., and AbuBakar, S. (2013). Dengue virus type 1 clade replacement in recurring homotypic outbreaks. BMC Evol. Biol. *13*, 213.

van Hemert, F.J., van der Kuyl, A.C., and Berkhout, B. (2013). The A-nucleotide preference of HIV-1 in the context of its structured RNA genome. RNA Biol. *10*, 211–215.

Vazeille, M., Gaborit, P., Mousson, L., Girod, R., and Failloux, A.B. (2016). Competitive advantage of a dengue 4 virus when co-infecting the mosquito Aedes aegypti with a dengue 1 virus. BMC Infect. Dis. *16*, 318.

Villordo, S.M., Filomatori, C.V., Sánchez-Vargas, I., Blair, C.D., and Gamarnik, A.V. (2015). Dengue virus RNA structure specialization facilitates host adaptation. PLoS Pathog. *11*, e1004604.

Wagner, A. (2008). Gene duplications, robustness and evolutionary innovations. Bioessays *30*, 367–373.

Wong, W.S.W., and Nielsen, R. (2004). Detecting selection in noncoding regions of nucleotide sequences. Genetics *167*, 949–958.

Zhou, Y., Mammen, M.P., Jr., Klungthong, C., Chinnawirotpisan, P., Vaughn, D.W., Nimmannitya, S., Kalayanarooj, S., Holmes, E.C., and Zhang, C. (2006). Comparative analysis reveals no consistent association between the secondary structure of the 3′-untranslated region of dengue viruses and disease syndrome. J. Gen. Virol. *87*, 2595–2603.

ISCI, Volume *16*

Supplemental Information

Evolution of Subgenomic RNA Shapes Dengue Virus Adaptation and Epidemiological Fitness

Esteban Finol and Eng Eong Ooi

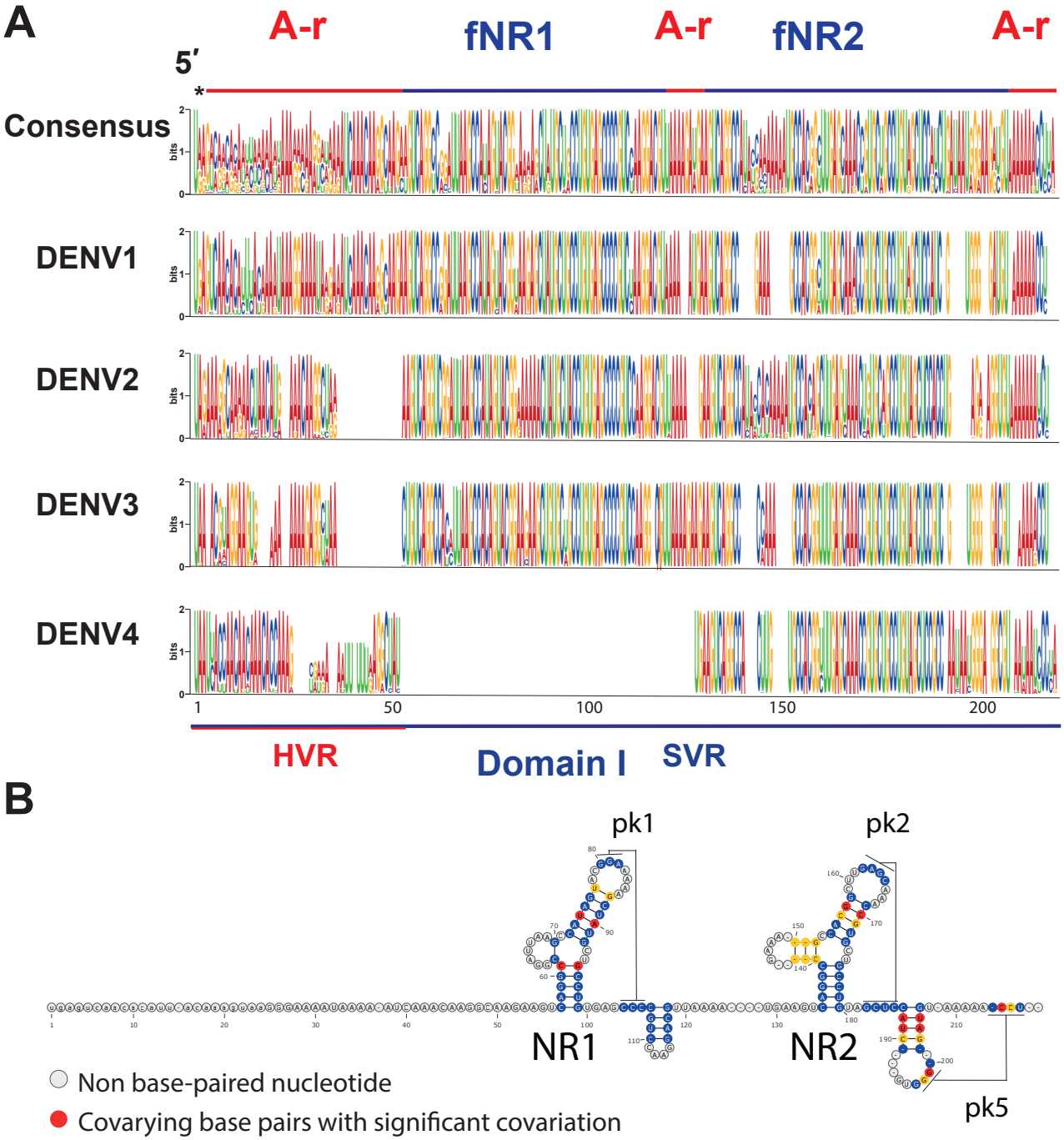
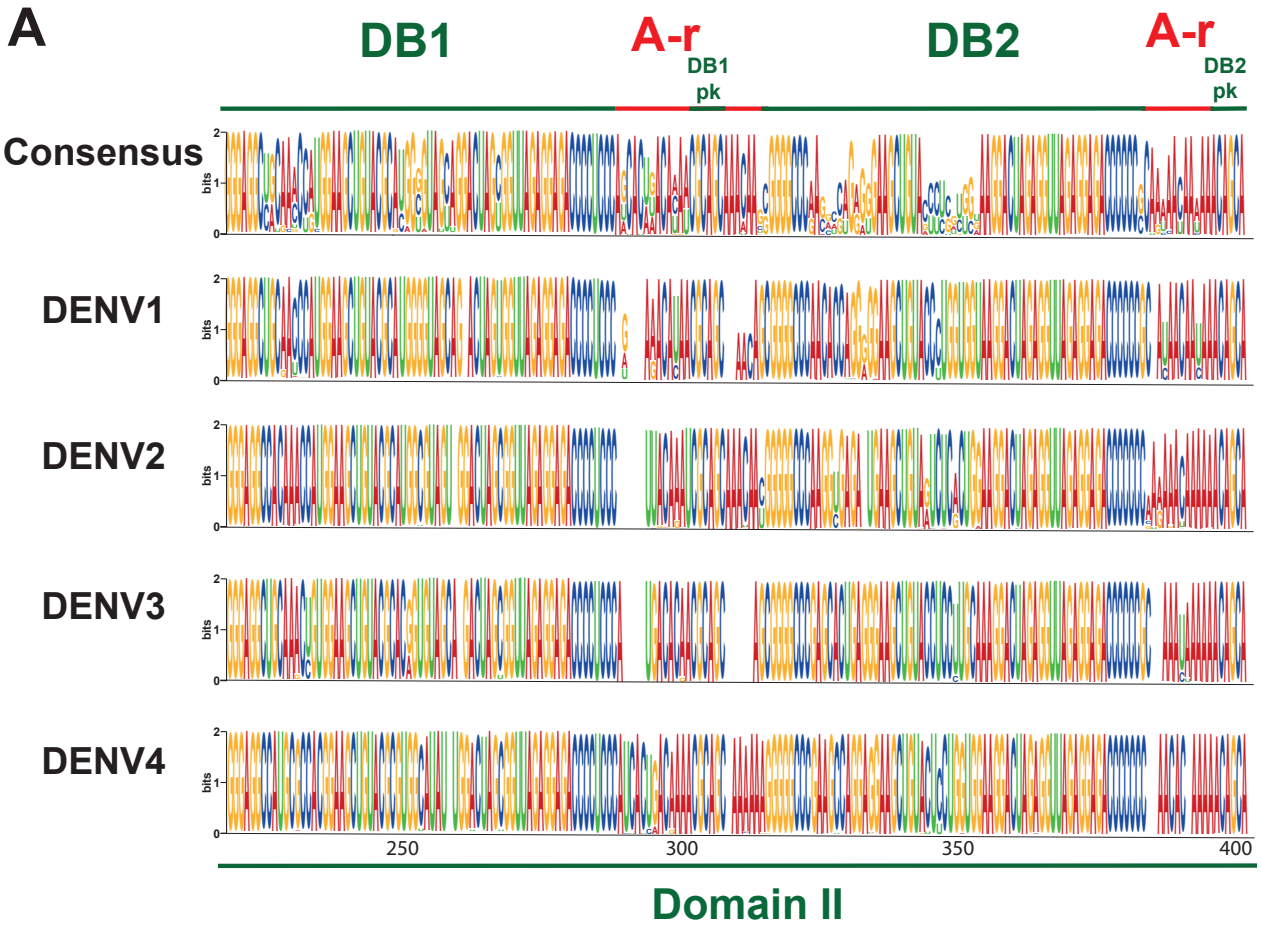


Figure S1, related to Figure 1A and Figure 1B. Detailed view of sequence conservation and secondary structures in domain I of DENV 3'UTRs.

(A) Alignment of Domain I sequence logos. Sequence logos for the 3'UTR of all DENV were aligned based on the multiple sequence alignment of DENV 3'UTRs sequences (consensus sequence logo). The highly variable (HVR) and semi-variable (SVR) regions in domain I are highlighted at the bottom of the figure. Nucleotides are color-coded (Blue = Cytosine, Green = Uracil, Yellow = Guanine, Red = Adenine).

(B) Consensus model for the secondary structure of Domain I in DENV 3'UTRs. The statistical support for covarying base pairs was estimated by G-statistics in Rscape software.



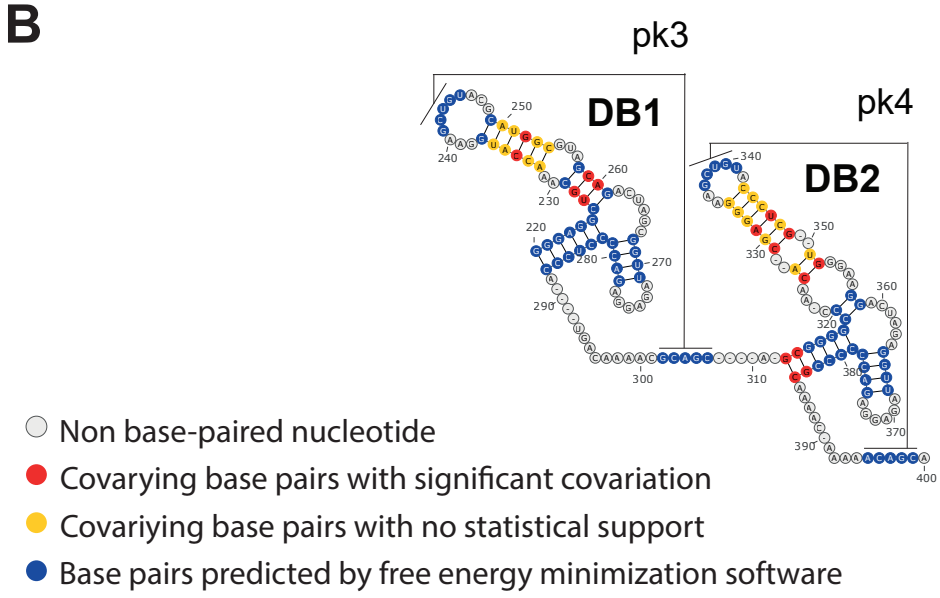
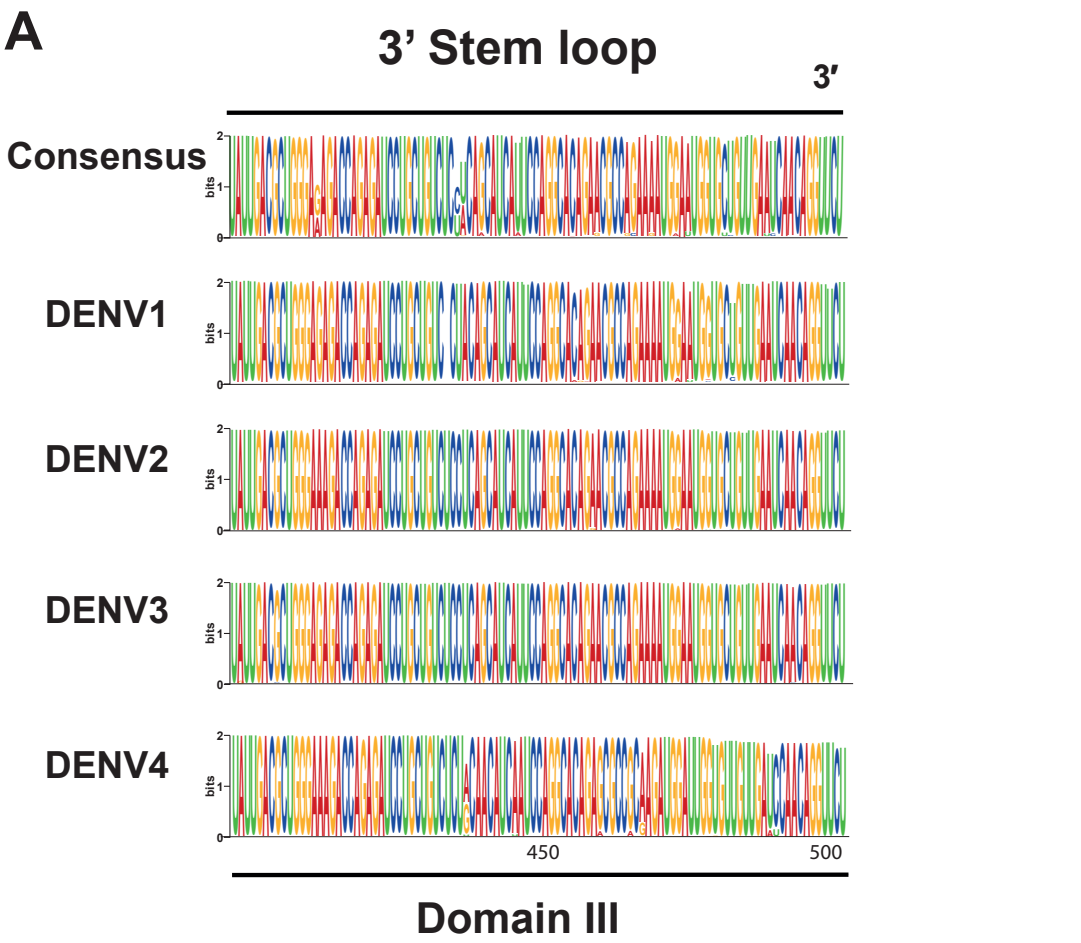


Figure S2, related to Figure 1A and Figure 1B. Detailed view of sequence conservation and secondary structures in domain II of DENV 3'UTRs.

(A) Alignment of Domain II sequence logos. Sequence logos for the 3'UTR of all DENV were aligned based on the multiple sequence alignment of DENV 3'UTRs sequences (consensus sequence logo). Nucleotides are color-coded (Blue = Cytosine, Green = Uracil, Yellow = Guanine, Red = Adenine).

(B)Consensus model for the secondary structure of Domain II in DENV 3'UTRs. The statistical support for covarying base pairs was estimated by G-statistics in Rscape software.



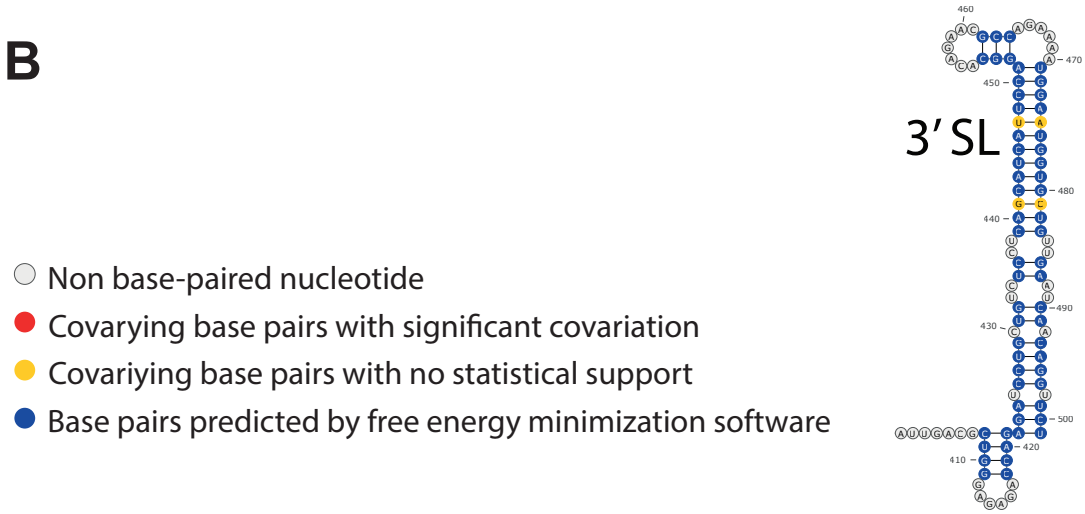
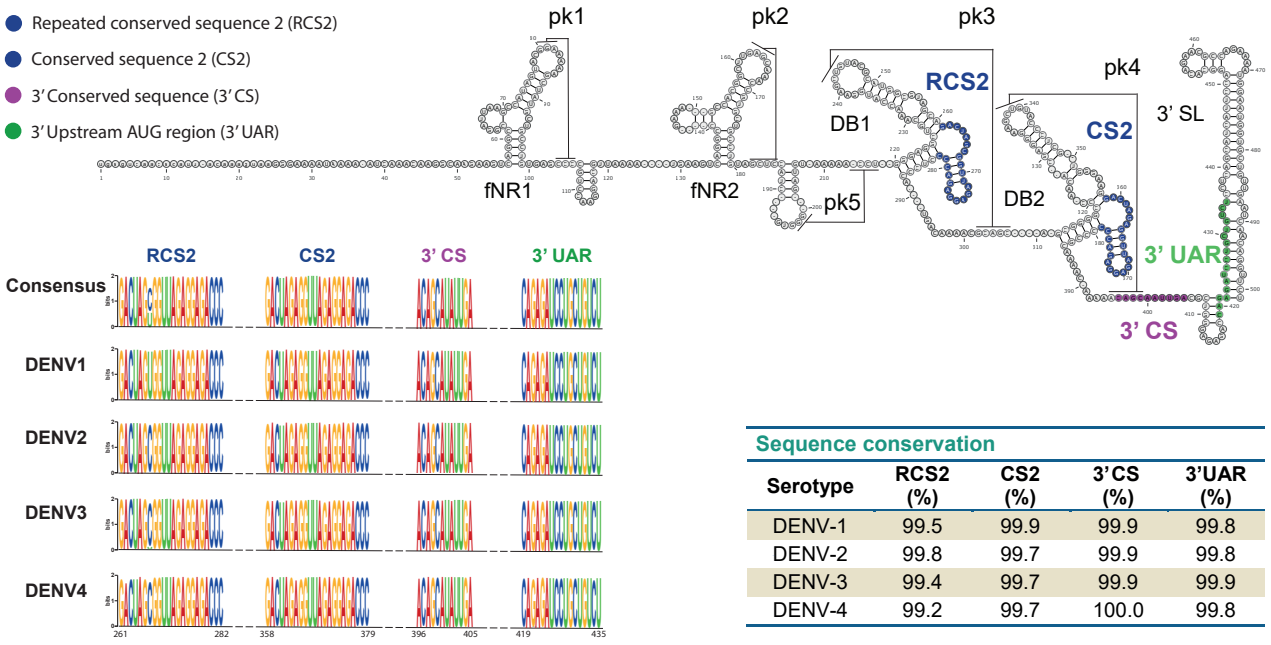


Figure S3, related to Figure 1A and Figure 1B. Detailed view of sequence conservation and secondary structures in domain III of DENV 3’UTRs.

(A) Alignment of Domain III sequence logos. Sequence logos for the 3’UTR of all DENV were aligned based on the multiple sequence alignment of DENV 3’UTRs sequences (consensus sequence logo). Nucleotides are color-coded (Blue = Cytosine, Green = Uracil, Yellow = Guanine, Red = Adenine).

(B) Consensus model for the secondary structure of Domain III in DENV 3’UTRs. The statistical support for covarying base pairs was estimated by G-statistics in Rscape software.



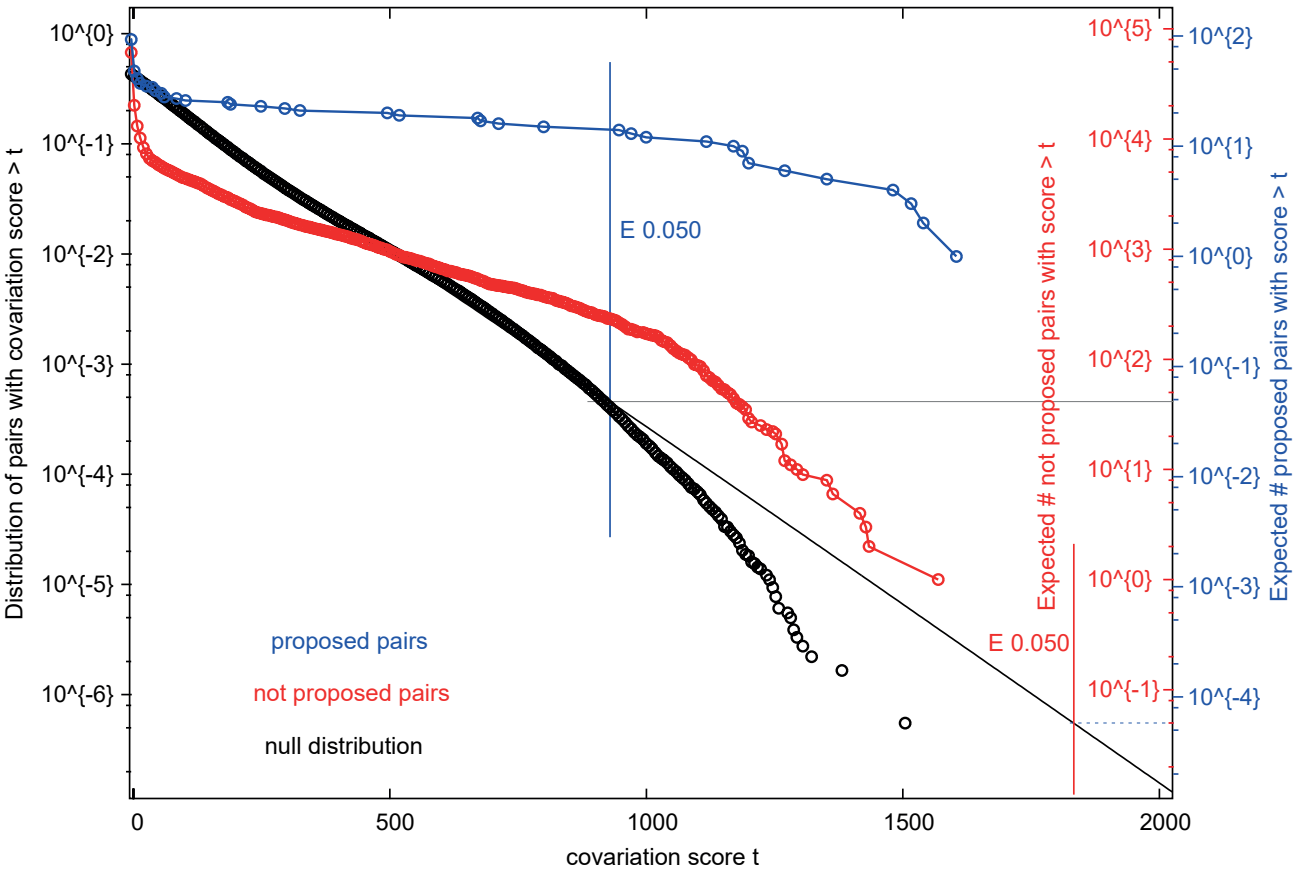
Sequence conservation

Serotype	RCS2 (%)	CS2 (%)	3′CS (%)	3′UAR (%)
DENV-1	99.5	99.9	99.9	99.8
DENV-2	99.8	99.7	99.9	99.8
DENV-3	99.4	99.7	99.9	99.9
DENV-4	99.2	99.7	100.0	99.8

Figure S4, related to Figure 1A. High sequence conservation in some segments of Domain II and III in the 3′ UTR of dengue viruses.

The sequences have been involved in genome cyclization, viral replication and translation.On the top panel, the conserved sequences are mapped on the secondary structure of DENV 3′UTRs. Sequence logos and the sequence conservation of these segments are shown at the bottom of the figure, left and right hand sides respectively.

A



B

R-scape analysis:

Input file:

- Multiple sequence alignment:
  - Number of sequences: 1825
  - Length: 501 nucleotides
  - Average identity: 83.17%
- RNA secondary structure:
  - Number of base pairs: 126

R-scape test:

- Covariation statistical method:
  APC-corrected G-Test statistic
- E-value threshold: 0.05

R-scape output:

- Number of base pairs after filter: 105
- Covarying base pairs: 14
- Covarying non base pairs: 0
- Range of scores: [-4.13 ; 1604.74]
- Sensitivity: 13.33
- Positive predictive value: 100

C

Base pairs with significant covariations

Left position	Right position	score	E-value
61	95	973.31	0.0340113
74	90	1192.42	0.00682436
157	170	1486.24	0.00076987
188	205	1604.74	0.000307065
189	204	1355.32	0.00201342
199	215	1540.87	0.00049705
227	260	1204.55	0.00625533
228	259	1189.19	0.00682436
235	251	1275.27	0.00370847
313	384	1003.30	0.027391
314	383	1117.62	0.0115008
325	352	1174.07	0.00777609
329	348	947.33	0.0404361
331	346	1517.10	0.000592138

Figure S5, related to figure 1B. G-statistics were applied to identify significant covariation of ribonucleotide base pairs in the 3' UTR of dengue viruses.

(A) Covariation scores survival functions. Two survival functions of scores are obtained upon G-test statistic implementation in R-scape software. In red, we drew the survival function of scores for all possible pairs in the input alignment excluding those proposed as base pairs. In blue, we plotted the survival function of scores for proposed base pairs in the input alignment. The survival function for the null alignments is depicted in black. The black line corresponds to fit to a truncated Gamma distribution of the tail of the null distribution.
(B) Summary of input alignment statistics, R-scape test parameters and R-scape output statistics.
(C) List of proposed base pairs with significant covariation. Nucleotide positions correspond to the position in the 2D model in figure 1B. Score and E-value are also shown.

Table S1A. Sequence length and Identity in the DENV 3'UTR alignments.

Serotype	Sequences	Length		Identical sites		Average identity
	n	Mode	Range	n	(%)	(%)
DENV-1	1486	465	[436–475]	169	35.4	96.5
DENV-2	1073	454	[444–469]	158	33.7	96.0
DENV-3	831	443	[429–455]	238	52.3	97.1
DENV-4	154	387	[387–407]	298	73.2	96.9

Table S1B. Nucleotide composition and Identity in the three domains of DENV 3'UTRs.

3' UTR Region	Serotype	Sequences	Length		Nucleotide composition (%)					Identical sites		Average identity
		n	Mode	[Range]	A	G	C	U	GC	n	(%)	(%)
Domain I	DENV-1	1486	196	[167–198]	35.0	21.5	25.3	18.2	46.8	82	(41.4)	**95.1**
	DENV-2	1073	183	[173–184]	36.7	21.9	22.8	18.6	43.7	66	(35.1)	**93.4**
	DENV-3	831	175	[161–180]	30.4	24.6	27.8	17.2	52.4	91	(50.0)	**94.8**
	DENV-4	154	113	[113–131]	31.3	26.9	21.4	20.4	48.3	76	(58.0)	**93.4**
HVR	DENV-1	1486	51	[22–51]	**55.8**	13.6	14.4	16.2	28.0	1	**(2.0)**	**88.1**
	DENV-2	1073	34	[25–36]	**55.4**	14.5	18.4	11.6	33.9	2	**(5.6)**	**82.7**
	DENV-3	831	28	[17–34]	**52.5**	6.1	26.9	14.5	33.0	2	**(5.9)**	**77.1**
	DENV-4	154	30	[30–48]	**50.4**	25.6	7.6	16.4	33.2	6	**(12.5)**	**84.5**
SVR	DENV-1	1486	145	[144–147]	27.6	24.4	29.1	18.9	**53.5**	81	(55.1)	97.6
	DENV-2	1073	151	[148–154]	28.0	23.5	28.8	20.2	**51.3**	65	(42.2)	95.7
	DENV-3	831	147	[147–148]	26.5	27.9	28.0	17.6	**55.9**	89	(60.1)	98.1
	DENV-4	154	83	[83]	24.2	27.3	26.7	21.8	**54.0**	70	(84.3)	96.8
Domain II	DENV-1	1356	167	[167–173]	29.0	25.7	32.0	13.3	**57.7**	52	(29.9)	97.8
	DENV-2	1071	169	[169–171]	31.6	24.7	29.9	13.8	**54.6**	75	(43.4)	97.6
	DENV-3	822	166	[166–169]	28.7	26.8	31.5	13.0	**58.3**	94	(55.6)	98.8
	DENV-4	153	172	[172–174]	29.7	26.6	31.3	12.4	**57.9**	138	(58.0)	98.4
Domain III	DENV-1	1140	102	[102–105]	30.1	24.9	25.1	19.9	50.0	35	(33.3)	99.2
	DENV-2	1008	102	[102–103]	30.7	24.8	24.1	20.4	48.9	30	(28.8)	99.5
	DENV-3	715	102	[102–104]	29.1	24.7	25.4	20.8	50.1	53	(51.0)	99.6
	DENV-4	151	102	[102]	28.8	24.6	25.9	20.7	50.5	84	(82.4)	98.5

Table S1C. Nucleotide composition and identity of Adenylate-rich spacers in DENV 3'UTRs.

Spacer	Serotype	Sequences	Length		Nucleotide composition (%)					Identical sites		Average identity
		n	Mode	[Range]	A	G	C	U	CG	n	(%)	(%)
HVR	DENV-1	1486	51	[22–51]	55.8	13.6	14.4	16.2	28.0	1	(2.0)	88.1
	DENV-2	1073	34	[25–36]	55.4	14.5	18.4	11.6	33.9	2	(5.6)	82.7
	DENV-3	831	28	[17–34]	52.5	6.1	26.9	14.5	33.0	2	(5.)	77.1
	DENV-4	154	30	[30–48]	50.4	25.6	7.6	16.4	33.2	6	(12.5)	84.5
fNR1–fNR2	DENV-1	1486	7	[7–8]	71.6	14.4	0.2	14.1	14.4	2	(25.0)	99.0
	DENV-2	1073	7	[7]	72.5	14.4	0.1	13.0	14.5	4	(57.1)	97.4
	DENV-3	831	10	[10]	70.0	20.0	0.0	10.0	20.0	8	(80.0)	99.9
	DENV-4	---	---	---	---	---	---	---	---	---	---	---
fNR2–DB1	DENV-1	1486	8	[8–9]	62.6	0.4	27.7	9.3	28.1	4	(44.4)	94.4
	DENV-2	1073	9	[8–11]	66.4	0.1	16.4	17.1	16.5	2	(18.2)	91.1
	DENV-3	831	7	[7–8]	54.0	3.2	28.6	14.3	31.8	4	(50.0)	94.6
	DENV-4	154	11	[10–11]	46.5	0.6	28.8	24.1	29.4	5	(45.5)	84.6
DB1	DENV-1	1356	9	[9–13]	62.4	10.3	14.9	12.4	25.1	1	(7.7)	83.5
	DENV-2	1071	14	[13–15]	56.7	17.1	17.9	24.2	19.1	1	(6.7)	90.5
	DENV-3	822	9	[9]	54.9	11.8	22.2	11.1	34.0	4	(44.4)	98.4
	DENV-4	153	16	[16–18]	63.2	6.0	18.8	12.0	24.8	7	(38.9)	92.0
DB2	DENV-1	1356	9	[8-10]	69.8	0.2	14.5	15.5	14.7	1	(10)	91.7
	DENV-2	1071	10	[10–12]	85.4	4.2	9.3	1.1	13.5	0	(0)	89.4
	DENV-3	822	9	[8–9]	87.3	0.1	2.1	10.5	2.2	1	(11.1)	95.4
	DENV-4	153	9	[8–9]	77.7	0.0	27.3	0.0	22.3	8	(88.9)	99.7

Table S1, related to Figure 1A. A detailed analysis of RNA sequence alignments exposed the variability within and across DENV 3'UTRs.

(A) Statistical analysis on the multiple sequence alignments from the 3' UTR of dengue viruses. This table summarizes the sequence conservation analysis on this segment of DENV genome. It includes number of sequences, the mode and range of sequence length, the absolute and relative number of identical sites (100% conserved sites) and the average identity.

(B) Nucleotide composition and sequence conservation in distinct segments of DENV 3'UTRs. This table summarizes the conservation analysis of the different domains in the 3' UTR of DENV. The comparison across the three domains within each dengue virus type revealed a significantly ($p<0.05$) higher CG content in domain II and a lower average identity in Domain I ($p<0.05$) (highlighted in red). Further analysis also revealed significant differences (highlighted in blue) between the Highly variable region (HVR) and the semivariable region (SVR) in Domain I.

(C) Nucleotide composition and sequence conservation of spacer sequences in DENV 3'UTRs. This table summarizes the conservation analysis of the different Spacer sequences. Adenine enrichment of the spacers is significant in all dengue virus types ($p<0.05$).

TRANSPARENT METHODS

Sequence conservation analysis

Complete DENV genome nucleotide sequences were downloaded directly from the GeneBank database. The search included the keywords "Dengue virus type X" (X=1–4). In total, 1486, 1073, 831 and 154 sequences were included in the analysis of the 3' UTR of DENV-1, DENV-2, DENV-3 and DENV-4, respectively. Most of the sequence dataset was reported to be isolated from human serum samples. Only 25 mosquito-isolated DENV sequences were available (4, 17 and 4 from DENV1, 2 and 3 respectively). All DENV mutants, laboratory adapted strains, replicons, vaccine candidate strains, serially passage strains, and duplicated sequences were previously excluded. We built multiple sequence alignments for each serotype using MAFFT (Multiple sequence alignment using Fast Furier Transformation) software (Katoh and Standley, 2013). The sequence alignments were limited to the 3'UTR, starting from the Stop codon in NS5. We used Geneious platform to calculate nucleotide composition, sequence length (mode and range), average identity (i.e. average nucleotide conservation in the alignment) and number of identical sites (100 percent conserved positions) in the 3' UTR of DENV (Kearse et al., 2012). To visualize the nucleotide composition pattern and conservation in the 3' UTR of each serotype, we generated sequence logos from the 3'UTR alignments using Weblogo server (Crooks et al., 2004) **(see Figure 1A)**. We used standard colors to represent each type of nucleotide in the alignment (Blue = Cytosine, Green = Uracil, Yellow = Guanine, Red = Adenine).

To further characterize the nucleotide conservation, composition and distribution in the 3'UTR of dengue viruses, we identified some conserved stretches that mapped to the start and end of the Dumbbell (DB) structures in the Domain II, as described by Shurtleff et al. (2001) and Gerhald et al. (2011). We used them to establish a clear border between the different domains in the 3' UTR of DENV and to perform subsequent sequence analyses in these domains. The statistical analysis of the nucleotide conservation and composition analysis was performed using STATA software (StataCorp, 2015). We used Analysis of Variance (ANOVA) with Bonferroni correction and Chi-square to test hypotheses from absolute values (average identity) and relative frequencies (nucleotide composition). Due to substantial differences across the sample size in the four data sets **(see Table S1A)**, all statistical comparisons were performed to test hypotheses within each serotype**.**

Sequence comparative analysis and RNA structure determination

To obtain secondary RNA structures and tertiary interactions, we applied a 'divide, learn and conquer' approach. It combines (1) an insightful 3' UTR sequence conservation analysis within each flavivirus and within flavivirus groups to identify the presence of conserved RNA structures, (2) the power of RNA structure prediction software to solve preliminary RNA secondary structures from short RNA segments and (3) the robustness of a sequence comparative analysis – or RNA phylogeny – to validate, improve and build a consensus RNA structure for the 3' UTR and sfRNA of flaviviruses (Jaeger et al., 1993). The strength of the RNA phylogeny approach relies upon the identification of evolutionary conserved functional RNA structures whose nucleotide sequences changed overtime but kept the RNA secondary and tertiary structures. Hence, it was possible to identify conserved functional RNA structures through sequence conservation analysis, to predict preliminary RNA structures using base-pairing probabilities and thermodynamic methods on the conserved stretches using RNAfold and HotKnots software (Lorenz et al., 2011 and Ren et al., 2005) and to validate secondary and tertiary interaction by identifying co-variations in the RNA nucleotide sequences, exploiting the growing sequencing dataset and high nucleotide substitution rate in RNA viruses. G-test statistics was implemented to further test whether observed RNA covariations occurred above phylogenetic expectation (Rivas et al., 2017) **(see Figure S5)**. We drew secondary RNA structures and pseudoknots using VARNA software (Darty et al., 2009).

Detecting natural selection in DENV sfRNA

To determine whether the RNA structures in the sfRNA play a role in the evolution of DENV, we explored natural selection pressure in a site-by-site basis in the sfRNA structure using a maximum-likelihood (ML) method (Wong and Nielsen, 2004) **(see Figure 2A)**. We modeled the evolution of coding and non-coding regions and assumed a constant neutral (synonymous) nucleotide substitution rate in both regions in each serotype viral genome. We modeled the evolution in the open reading frame of DENV genome and determined its synonymous substitution rate, using a model of codon evolution (General Time Reversible, GTR+Γ) that has been generally applied to study the coding region of DENV genome (Weaver & Vasilakis, 2009). On the other hand, we calculated the nucleotide substitution rate in the sfRNA sequence in site-by-site basis. We used PHASE 3.0 software and a combined model of non-conding RNA evolution (Loop model: Hanley and Knott Regression, HKR+Γ and Stem model: 16D) based on the RNA secondary structure for the sfRNA from each serotype (Allen and Whelan. 2014). We normalized the nucleotide substitution rate in each position of the sfRNA sequence by the synonymous substitution rate in the coding region of each DENV serotype and estimated a ζ parameter. Thus, a nucleotide position that exhibited a similar nucleotide substitution rate to the synonymous substitution rate ($\zeta \approx 1$) was assumed to be under a neutral evolution, whereas when ζ was found to be significantly higher or lower than 1 in a given position in the sfRNA sequence we assumed that it has experienced the action of positive or negative selection, respectively. To provide statistical significance to ζ parameter ratios, we calculated a 95% confidence interval (CI) for ζ parameter across each DENV serotype genome. If the ζ value of a given position was within the 95%CI, we confirmed neutral evolution. If the ζ value was above or below the 95%CI, we reported a significant positive and negative selection, respectively.

DENV2 sfRNA Phylogenetics tree

We constructed a phylogentics tree for the sfRNA of DENV-2 from an alignment of 356 unique and representative 3'UTR DENV-2 sequences. We used PHASE 3.0 package (Allen and Whelan, 2014) to build a maximum likelihood phylogenetic tree using the same composed model of nucleotide substitution (Loop model: HKR+Γ and Stem model: 16D) based on the predicted RNA secondary structures in the sfRNA of DENV-2. The statistical support for the topology of the tree was determined by 1000 bootstrap replications.

DENV-2 sfRNA 3D modeling

We modeled the 3D RNA structure of DENV-2 sfRNA using RNA composer (Popenda et al. 2012). We used for the input file all the secondary and tertiary interactions that we obtained from the RNA phylogeny approach. The modeling of DENV-2 fNR structures was optimized through comparative RNA modeling, using ZIKV fNR crystal structure as template (5TPY). This was performed using ModeRNA software (Piatkowski et al., 2016). The local geometry in preliminary models were refined through energy minimization using the AMBER force field in the Molecular Modelling toolkit (Hinsen, 2000). The final simulations were inspected for steric clashes using the *find-clashes* function in ModeRNA. Although the de novo and comparative modeling of structured RNA elements in the sfRNA was highly reproducible, the presence of non-base-paired spacers rendered the orientation of structured elements in relation to each other poorly reproducible.

Therefore, only a representative model is shown in Figure 2D. The final sfRNA model was visualized, colored and labeled using pyMOL software.

Construction of a sfRNA fitness landscape

A fitness landscape is a XYZ plot that depicts in the XY plane the genotypic space and every point in the genotypic space projects a fitness value to the Z plane. For the sfRNA fitness landscape, the extend of the genotypic space was defined in the X and Y planes by the product of all possible fNR and DB sequences. The number of possible sequences corresponded to sequence length to the power of 5, accounting for the possibility of having U, C, A, G or a missing nucleotide in every position. Thus, the overall genotypic space covers 1.3425^{14} points (fNR length = 75 nucleotides, BD length = 179 nucleotides, genotypic space = 75^5x179^5). Domain III of DENV 3'UTR was not integrated into the 2D genotypic space. Its length (102 nucleotides, see Table S2) would substantially increase the size of the genotypic space to be computed and analyzed (1.3425^{14}x102^5) and it will not contribute to resolve the 2D space due the high sequence identity in this part of DENV 3'UTR (>99%, see Table S2). The fitness value was obtained from the relative combined abundance of fNR and DBs sequences among all DENV full genome sequences in NCBI database. Thus, if a combination of fNR-DB sequences conferred high fitness to the bearing strain, they would pass through subsequent viral generations in the mosquito- human interspecies transmission, they would be sampled and sequenced. The sequences would them appear in the NCBI database. Unfortunately, some strains have been more frequently sampled and sequenced than other strains and the NCBI database holds a sequencing bias. To adjust the fitness value and reduce the impact of the sequencing bias, a phylogenetic based normalization was implemented. Every clade was weighted according to it proportional contribution to the total number of nucleotide sequences from every DENV type, this weighting was then transferred to every sequence in the clade. Thus, if a clade contains 40% of all the DENV sequences – an over-sampled clade, every sequence in this clade would no longer count as 1 in the sum for the Z value, it will rather count as 0.6. If the clade represents 0.1% of the all sequences, every sequence in the clade will add 0.99 to the Z Value. FASTAptamer-cluster software was used to cluster sequences into sequence families based on a fixed Levenshtein edit distance (Alam et al., 2015). The fitness landscape was plotted using gnuplot graphing utility in LINUX Operating System. An XYZ plot was generated by the *splot* command and the surface plot was obtained after a gaussian approximation to the XYZ raw data by the *gauss* kernel under *dgrid3d* command. Figure 2C shows only a small section of the overall fNR-DB genotypic space (3.38^{-9} %, 453882 XY points), where a Z value could be assigned. The rest of the genotypic space is completely deprived of fitness peaks and is not shown.

Data and Software Availability

The raw DENV genomic data, phylogenetic data and secondary structure models for DENV 3' UTRs are available as a Mendeley dataset. The annotation of DENV 3'UTR RNA elements is also available in the four DENV reference sequences (GeneBank: NC_001477, NC_001475, NC_002640 and NC_001474).

SUPPLEMENTAL REFERENCES

Alam K.K., Chang J.L., Burke D.H., FASTAptamer: A Bioinformatic Toolkit for High-throughput Sequence Analysis of Combinatorial Selections, Mol Ther Nucleic Acids. **4**, 2015, e230. doi: 10.1038/mtna.2015.4.

Allen J.E., Whelan S., Assessing the state of substitution models describing noncoding RNA evolution, Genome Biol Evol. **6**, 2014, 65–75. doi: 10.1093/gbe/evt206.

Crooks G.E., Hon G., Chandonia J.M., Brenner S.E., WebLogo: a sequence logo generator, Genome Res. **14**, 2004, 1188 –1190. doi: 10.1101/gr.849004.

Darty K., Denise A., Ponty Y., VARNA: Interactive drawing and editing of the RNA secondary structure, Bioinformatics. **25**, 2009, 1974–1975. doi: 10.1093/bioinformatics/btp250.

Katoh K. and Standley D.M., MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol Biol Evol. **30**, 2013, 772–780. doi: 10.1093/molbev/mst010.

Hinsen K., The molecular modeling toolkit: A new approach to molecular simulations, J. Comput. Chem. **21**, 2000, 79–85. doi: 10.1002/(SICI)1096-987X(20000130)21:2<79:AID-JCC1>3.0.CO;2-B.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A., Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, Bioinformatics **28**, 2012, 1647–1649. doi: 10.1093/bioinformatics/bts199.

Piatkowski P., Kasprzak J.M., Kumar D., Magnus M., Chojnowski G., Bujnicki J.M., RNA 3D Structure Modeling by Combination of Template-Based Method ModeRNA, Template-Free Folding with SimRNA, and Refinement with QRNAS, Methods Mol Biol. **1490**, 2016, 217–235. doi: 10.1007/978-1-4939-6433-8_14.

Popenda M., Szachniuk M., Antczak M., Purzycka KJ., Lukasiak P., Bartol N., Blazewicz J., Adamiak R.W., Automated 3D structure composition for large RNAs, Nucleic Acids Res. 40, 2012, e112. doi: 10.1093/nar/gks339.

StataCorp. Stata Statistical Software: Release 14, 2015, College Station, TX: StataCorp LP.

Weaver S.C. and Vasilakis N., Molecular evolution of dengue viruses: contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease, Infect. Genet. Evol. 9, 2009, 523–540.