Title:

Evolution of subgenomic RNA shapes dengue virus adaptation and epidemiological fitness

Authors:

Esteban Finol[1,2,3,4] & Eng Eong Ooi[1,3,5]

Corresponding author and Lead contact: Esteban Finol (esteban.finol@mol.biol.ethz.ch)

Co-corresponding author: Eng Eong Ooi (engeong.ooi@duke-nus.edu.sg)

Affiliations:

[1] Emerging infectious diseases programme, Duke-NUS Medical School, Singapore.

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

[3] Department of Microbiology and Immunology, Yong Loo Lin School of Medicine, National University of Singapore, 119077 Singapore.

[4] Current affiliation: Institute for Molecular Biology and Biophysics, ETH Zurich, Switzerland.

[5] Saw Swee Hock School of Public Health, National University of Singapore, 119077 Singapore.

Highlights

- Dengue viruses (DENV) preserve RNA elements in their 3' untranslated region (UTR).
- Site-specific quantification of natural selection revealed positive selection on DENV2 sfRNA.
- Flaviviral nuclease-resistant RNA (fNR) structures in DENV 3'UTRs contribute to DENV speciation.
- A highly evolving fNR structure appears to increase DENV-2 epidemiological fitness.

Abstract

Genetic changes in the dengue virus (DENV) genome affects viral fitness both clinically and epidemiologically. Even in the 3' untranslated region (3'UTR), mutations could impact the formation of subgenomic flaviviral RNA (sfRNA) and the specificity of sfRNA in inhibiting host proteins necessary for successful viral replication. Indeed, we have recently shown that mutations in the 3'UTR of DENV2 affected its ability to inhibit TRIM25 E3 ligase activity to reduce interferon (IFN) expression, which potentially contributed to the emergence of a new viral clade during the 1994 dengue epidemic in Puerto Rico. However, whether differences in 3'UTRs shaped DENV evolution on a larger scale 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. Interestingly, only the second flaviviral Nuclease-resistant RNA (fNR2) structure of DENV-2 has undergone strong positive selection. Epidemiological reports also suggest that nucleotide substitutions in fNR2 may drive DENV-2 epidemiological fitness, possibly through sfRNA-protein interactions. Collectively, our findings indicate that 3'UTRs are important determinants of DENV fitness in human-mosquito cycles.

Keywords:

dengue virus; sfRNA; epidemiological fitness; ncRNA evolution

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 to 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., 2016). Whereas 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 appears 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, both through its interaction with human (Manokaran et al., 2015). Nucleotide substitutions in the 3'UTR of DENV2 strains resulted in increased sfRNA levels. This viral non-coding RNA binds to TRIM25 protein to inhibit its deubiquitylation (Manokaran et al., 2015); 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. 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 was 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., 2005; 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 et al., 2005 and 2008). These structures have been termed small hairpin (sHP) 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 DENV-1, 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 subgenomic flavivirus RNA (sfRNA) during infection (Piljman et al., 2008). These structures have been referred to with different names in the literature: 'stem-loops' (Shurtleff et al., 2001, Piljman 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 (MVEV) (Chapman et al., 2014B) and Zika virus (ZIKV) (Akiyama et al., 2016), revealed a three-way junction RNA folding, rather than a SL structure. More importantly, these RNA structures halt diverse exoribonucleases to produce sfRNA, not only Xrn1 exoribonuclease (MacFadden A et al., 2018). Therefore, we refer herein to these RNA structures as flaviviral nuclease-resistant RNA (fNR) structures, as initially termed by Piljman et al. (2008), and to differentiate them from the unrelated nuclease-resistant RNA structures in dianthoviruses (Steckelberg AL 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 non-coding RNA sequences can depict several features of their evolutionary trajectories. The ncRNA length, sequence identity and nucleotide composition can unveil insertion/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 DENV2 (Mode 454; range 444–469), DENV3 (Mode 443; range 429–455), and DENV4 (Mode 387; range 387–407) **(Table S1A)**. Their distal segments (domains II and III) are highly conserved **(Figure 1A)**. In contrast, domain I of 3'UTR exhibits high genetic variability, including multiple insertions, deletions and point mutations. 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 ($p<0.05$) **(Table S1B)**. 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). In comparison with the HVR, domain I also possesses a semi variable region with a high level of nucleotide conservation (Average Identity ≥ 96% in all serotypes) and significantly higher GC content **(Figure 1A and Table S1B),** suggesting the presence of conserved RNA structures across DENV serotypes. These conserved stretches are separated by small adenine-rich sequences, which may serve as spacer to facilitate the proper folding of functional RNA structures.

To define the secondary RNA structures and tertiary interactions of the 3'UTR, a 'divide, learn and conquer' approach was adopted. 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 were also applied to determine whether these covariations occur at a rate higher than phylogenetically expected **(Figure S1)** (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 exerts the main nuclease-resistant function for sfRNA production and is therefore necessarily highly conserved. 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 related 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 fNR structures influences DENV evolution, we examined the selection pressure on each nucleotide positions using a maximum-likelihood (ML) phylogenetic method (Wong and Nielsen, 2004). 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 as compared to 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 NR2 structure of DENV2 could have undergone positive selection with $\zeta>1$ **(Figures 2A & 2B)**, suggesting that substitutions in fNR2 may confer competitive advantage for DENV-2 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 and/or existence of multiple fNR/DB fitness peaks across the four DENV types and their divergence after fNR duplication. More importantly, the fNR2 genotypic space is populated by more dispersed and isolated peaks of fitness. It further confirmed the greater evolvability and divergences of fNR2 as compared to 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, a ML phylogenetic tree was constructed using a nucleotide substitution model of evolution for non-coding RNA, based on the consensus secondary RNA structure of DENV2 sfRNA. 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 (Zhuo 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 a non-coding RNA, 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 non-coding RNA model of nucleotide substitutions coupled with normalizing 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 non-coding RNA evolution of a widely spread virus.

Available sequence data indicate that the DBs and the terminal 3' SL (Domain 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 increase sfRNA production, since the 30 bases nucleotide deletion in DB1 decreases sfRNA production and increases type I IFN susceptibility of the live attenuated DENV vaccine candidates (Bustos-Arriaga J et al., 2018). Likewise, fNR structures in the DENV 3'UTR are relatively well conserved. Any nucleotide substitution in base-paired positions of the fNR structures is often accompanied by compensatory mutation to maintain structural integrity. Along this line of thought, the finding of a positive selection in the NR2 structure of DENV 3'UTR is thus intriguing. fNR2 mutations may emerge during mosquito infection and subsequently be selected in human infections, due to their contribution to replicative fitness (Villordo et al., 2015; Filomatori et al., 2017). Indeed, 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). These mutations could have then been positively selected in subsequent human infections in distinct geographical locations, possibly based on their ability to bind host proteins for the suppression antiviral immune activation. Indeed, Bidet et al. (2014) showed that fNR structure interacts with CAPRIN G3BP1 and G3BP2 proteins, mediating the sfRNA-induced repression of interferon-stimulated mRNAs in human liver-derived Huh7 cells. Moreover, mutations in the fNR2 structures produced higher replicative fitness in *Aedes albopictus* compared to the corresponding wild type DENV2 (Filomatori et al., 2017). Collectively, these findings suggest a strong evolutionary pressure on DENV2 fNR2 structures. Additionally, the consistent concordance between previously reported experimental data and our bioinformatics findings highlights the robustness of ML method developed by Wong and Nielsen (2004).

Given our RNA phylogeny findings and other available experimental evidence, we propose that the two fNR structures in domain I 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 SE Asian/American DENV2 and found small viral plaques in Vero cells and slower growth kinetics in both mosquito and human cells, which are phenotypes more congruent with the American than SE Asian/American 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. Firstly, 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 and 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 as compared to DENV2. 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. If the fNR duplication occurred early during DENV evolution – 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. This constraint might have been overcome through antibody-dependent enhancement in primates and/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 nucleotides deletion in the 3'UTR is required to generate a complete attenuated phenotype in DENV4 as well as 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.

Acknowledgements

This work was supported by the Duke-NUS Signature Research Program funded by the Ministry of Health, Singapore and the Singapore National Medical Research Council. The authors would also like to thank Drs Vijaykrishnan Dhanasekaran and Mariano Garcia-Blanco for their invaluable advice throughout this work.

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. 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 (Crook et al., 2004) **(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, 2009). 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, all statistical comparisons were performed to test hypotheses within each serotype **(Table S1).**

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) **(Figure S1)**. 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 & Nielsen. 2004) **(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 JE, Whelan S. 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 JE, Whelan S. 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 K. 2000). The final simulations were inspected for steric clashes using the *find-clashes* function in ModeRNA. The final sfRNA model was visualized, colored and labeled using pyMOL software **(Figure 2B)**.

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 = $75^5 x 179^5$). Domain III of DENV 3'UTR was not integrated into the 2D genotypic space. Its length (102 nucleotides, Table S1) would substantially increase the size of the genotypic space ($1.3425^{14} x 102^5$) and it will not contribute to resolve the 2D space due the high sequence identity in this part of DENV 3'UTR (>99%, Table S1). 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 KK, 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.

Figure legends

Figure 1. The 3' UTR of dengue viruses diverged through deletion and sequence coevolution of functional RNA structures.

Dengue viruses are phylogenetically related RNA viruses, their 3'UTR sequences have diverged along evolution, 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). Highly conserved sequences were used to demark the boundaries between three domains in the 3'UTR of dengue viruses. Nucleotides are color-coded (Blue = Cytosine, Green = Uracil, Yellow = Guanine, Red = Adenine). 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 3' UTR of DENV. The sequence logos also provide a glance on sequence conservation and nucleotide composition. These data are further described in Table S1.

(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 S1 provides the parameters for the implementation and detailed results).

(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 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 accordingly to the heat map at the bottom of the figure. Highly conserved sites are highlighted in a scale from red to black (Site conservation > 95%).

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 zeta parameter and its 95% CI interval across the full genome determined whether a nucleotide position underwent negative (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) 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 harpin are highlighted in colors as in the secondary RNA structures at the bottom right of the figure.

(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.

Figure 3. Nucleotide substitution in DENV2 sfRNA are associated to 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. Harpin secondary structures and sequence logos are shown next to the corresponding branch for the DENV2 genotypes in the phylogentic tree. The consensus secondary structures for the sfRNA in all the DENV2 genotypes is 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 DENV-2 clade replacements and a natural attenuation event.

Table 1. List of epidemiological events associated to increased DENV epidemiological fitness.

An extensive literature revision on DENV epidemiology revealed at least 13 events associated with increased DENV epidemiological fitness. Nucleotide substitutions in the 3'UTR were reported in nine of those epidemic DENV strains (*).

Figure S1. G-statistics was 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 S1. 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 3'UTR DENVs. 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.

References

Akiyama BM, Laurence HM, Massey AR, Costantino DA, Xie X, Yang Y, Shi PY, Nix JC, Beckham JD, Kieft JS. (2016) Zika virus produces noncoding RNAs using a multi–pseudoknot structure that confounds a cellular exonuclease. Science. 354(6316), 1148–1152.

Alam KK, Chang JL, Burke DH. (2015) FASTAptamer: A Bioinformatic Toolkit for High-throughput Sequence Analysis of Combinatorial Selections. Mol Ther Nucleic Acids. 4, e230. doi: 10.1038/mtna.2015.4.

Allen JE, Whelan S. (2014) Assessing the state of substitution models describing noncoding RNA evolution. Genome Biol Evol. 6(1), 65–75. doi: 10.1093/gbe/evt206.

Alvarez DE, De Lella Ezcurra AL, Fucito S, Gamarnik AV. (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 DE, Filomatori CV, Gamarnik AV. (2008) Functional analysis of dengue virus cyclization sequences located at the 5' and 3'UTRs. Virology. 375, 223–235. doi: 10.1016/j.virol.2008.01.014.

Anderson JR, Rico–Hesse R. (2006) Aedes aegypti vectorial capacity is determined by the infecting genotype of dengue virus. Am J Trop Med Hyg. 75(5), 886–892.

Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O, et al. (2013) The global distribution and burden of dengue. Nature 496, 504–507. doi: 10.1038/nature12060.

Bidet K, Dadlani D, Garcia–Blanco MA. (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. doi: 10.1371/journal.ppat.1004242.

Bustos-Arriaga J, Gromowski GD, Tsetsarkin KA, Firestone CY, Castro-Jiménez T, Pletnev AG, Cedillo-Barrón L, Whitehead SS. (2018). Decreased accumulation of subgenomic RNA in human cells infected with vaccine candidate DEN4Δ30 increases viral susceptibility to type I interferon. Vaccine. pii: S0264-410X(18)30607-8. doi: 10.1016/j.vaccine.2018.04.087. Carneiro AR, Cruz AC, Vallinoto M, Melo Dde V, Ramos RT, Medeiros DB, Silva EV, Vasconcelos PF. (2012) Molecular characterisation of dengue virus type 1 reveals lineage replacement during circulation in Brazilian territory. Mem Inst Oswaldo Cruz. 107(6), 805–812.

Carrillo-Valenzo E, Danis-Lozano R, Velasco-Hernández JX, Sánchez-Burgos G, Alpuche C, López I, Rosales C, Baronti C, de Lamballerie X, Holmes EC, Ramos-Castañeda J. (2010) Evolution of dengue virus in Mexico is characterized by frequent lineage replacement. Arch Virol. 155(9), 1401–1412. doi: 10.1007/s00705-

010-0721-1.

Chapman EG, Costantino DA, Rabe JL, Moon SL, Wilusz J, Nix JC, Kieft JS. (2014B) The structural basis of pathogenic subgenomic flavivirus RNA (sfRNA) production. Science. 344(6181), 307–310. doi: 10.1126/science.1250897.

Chapman EG, Moon SL, Wilusz J, Kieft JS. (2014A) RNA structures that resist degradation by Xrn1 produce a pathogenic Dengue virus RNA. Elife. 3, e01892. doi: 10.7554/eLife.01892.

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

Chen R, Vasilakis N. (2011) Dengue--quo tu et quo vadis? Viruses. 3(9), 1562–1608. doi: 10.3390/v3091562.

Darty K, Denise A, Ponty Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics. 25(15), 1974–1975. doi: 10.1093/bioinformatics/btp250.

Dash PK, Sharma S, Soni M, Agarwal A, Sahni AK, Parida M. (2015) Complete genome sequencing and evolutionary phylogeography analysis of Indian isolates of Dengue virus type 1. Virus Res. 195, 124–134. doi: 10.1016/j.virusres.2014.08.018.

de Castro MG, de Nogueira FB, Nogueira RM, Lourenço-de-Oliveira R, dos Santos FB. (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. doi: 10.1186/1743-422X-10-3.

Dupont-Rouzeyrol M, Aubry M, O'Connor O, Roche C, Gourinat AC, Guigon A, Pyke A, Grangeon JP, Nilles E, Chanteau S, Aaskov J, Cao-Lormeau VM. (2014) Epidemiological and molecular features of dengue virus type-1 in New Caledonia, South Pacific, 2001-2013. Virol J.11, 61. doi: 10.1186/1743-422X-11-61.

Durbin AP, Karron RA, Sun W, Vaughn DW, Reynolds MJ, Perreault JR, Thumar B, Men R, Lai CJ, Elkins WR, 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 AP, Kirkpatrick BD, Pierce KK, Elwood D, Larsson CJ, Lindow JC, Tibery C, Sabundayo BP, Shaffer D, Talaat KR, 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. doi: 10.1093/infdis/jis936.

Filomatori CV, Carballeda JM, Villordo SM, Aguirre S, Pallarés HM, Maestre AM, Sánchez-Vargas I, Blair CD, Fabri C, Morales MA, Fernandez-Sesma A, Gamarnik AV. (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(3), e1006265. doi: 10.1371/journal.ppat.1006265.

Gritsun DJ, Jones IM, Gould EA, Gritsun TS. (2014) Molecular archaeology of Flaviviridae untranslated regions: duplicated RNA structures in the replication enhancer of flaviviruses and pestiviruses emerged via convergent evolution. PLoS One. 9, e92056. doi: 10.1371/journal.pone.0092056.

Hadinegoro SR, Arredondo–García JL, Capeding MR, Deseda C, Chotpitayasunondh T, Dietze R, Muhammad Ismail HI, Reynales H, Limkittikul K, Rivera–Medina DM, et al. (2015) Efficacy and Long-Term Safety of a Dengue Vaccine in Regions of Endemic Disease. N Engl J Med. 373(13), 1195–1206. doi: 10.1056/NEJMoa1506223.

Halstead SB. (2014) Dengue Antibody-Dependent Enhancement: Knowns and Unknowns. Microbiol Spectr. 2(6). doi: 10.1128/microbiolspec.AID-0022-2014.

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

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

Kirkpatrick BD, Whitehead SS, Pierce KK, Tibery CM, Grier PL, Hynes NA, Larsson CJ, Sabundayo BP, Talaat KR, 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. doi: 10.1126/scitranslmed.aaf1517

Kumar SR, Patil JA, Cecilia D, Cherian SS, Barde PV, Walimbe AM, Yadav PD, Yergolkar PN, Shah PS, Padbidri VS, Mishra AC, Mourya DT. (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(3), 707–720. doi: 10.1099/vir.0.017954-0.

Lee KS, Lo S, Tan SS, Chua R, Tan LK, Xu H, Ng LC. (2012) Dengue virus surveillance in Singapore reveals high viral diversity through multiple introductions and in situ evolution. Infect Genet Evol. 12(1), 77–85. doi: 10.1016/j.meegid.2011.10.012.

Lequime S, Fontaine A, Ar Gouilh M, Moltini-Conclois I, Lambrechts L. Genetic Drift, Purifying Selection and Vector Genotype Shape Dengue Virus Intra-host Genetic Diversity in Mosquitoes. PLoS Genet. 12(6), e1006111. doi: 10.1371/journal.pgen.1006111.

Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. (2011) ViennaRNA Package 2.0. Algorithms Mol Biol. 6, 26. doi: 10.1186/1748-7188-6-26.

MacFadden A, O'Donoghue Z, Silva PAGC, Chapman EG, Olsthoorn RC, Sterken MG, Pijlman GP, Bredenbeek PJ, Kieft JS. (2018) Mechanism and structural

diversity of exoribonuclease-resistant RNA structures in flaviviral RNAs. Nat Commun. 9(1), 119. doi: 10.1038/s41467-017-02604-y.

Manakkadan A, Joseph I, Prasanna RR, Kunju RI, Kailas L, Sreekumar E. Lineage shift in Indian strains of Dengue virus serotype-3 (Genotype III), evidenced by detection of lineage IV strains in clinical cases from Kerala. (2013) Virol J. 10, 37. doi: 10.1186/1743-422X-10-37.

McElroy KL, Santiago GA, Lennon NJ, Birren BW, Henn MR, Muñoz-Jordán JL. Endurance, refuge, and reemergence of dengue virus type 2, Puerto Rico, 1986-2007. (2011) Emerg Infect Dis. 17(1), 64–71. doi: 10.3201/eid1701.100961.

Men R, Bray M, Clark D, Chanock RM, Lai CJ. (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 JP, Brady OJ, Pigott DM, Golding N, Kraemer MU, Scott TW, Wint GR, Smith DL, Hay SI (2015). The many projected futures of dengue. Nat Rev Microbiol. 13, 230–239. doi: 10.1038/nrmicro3430.

Mir D, Romero H, Fagundes de Carvalho LM, Bello G. (2014) Spatiotemporal dynamics of DENV-2 Asian-American genotype lineages in the Americas. PLoS One. 9(6), e98519. doi: 10.1371/journal.pone.0098519.

Myat Thu H, Lowry K, Jiang L, Hlaing T, Holmes EC, 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(2),163–172. doi: 10.1016/j.virol.2005.03.018.

OhAinle M, Balmaseda A, Macalalad AR, Tellez Y, Zody MC, Saborío S, Nuñez A, Lennon NJ, Birren BW, 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. doi: 10.1126/scitranslmed.3003084.

Pankhong P, Weiner DB, Ramanathan MP, Nisalak A, Kalayanarooj S, Nimmannitya S, 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(10), 481–491. doi: 10.1089/dna.2008.0835.

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

Pijlman GP, Funk A, Kondratieva N, Leung J, Torres S, van der Aa L, Liu WJ, Palmenberg AC, Shi PY, Hall RA, Khromykh AA. (2008) A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. Cell Host Microbe. 4, 579–591. doi: 10.1016/j.chom.2008.10.007

Pompon J, Manuel M, Ng GK, Wong B, Shan C, Manokaran G, Soto-Acosta R, Bradrick SS, Ooi EE, Missé D, Shi PY, Garcia-Blanco MA. (2017) Dengue subgenomic flaviviral RNA disrupts immunity in mosquito salivary glands to increase virus transmission. PLoS Pathog. 13(7), e1006535. doi: 10.1371/journal.ppat.1006535.

Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, Blazewicz J, Adamiak RW. (2012) Automated 3D structure composition for large RNAs. Nucleic Acids Res. 40(14), e112. doi: 10.1093/nar/gks339.

Proutski V, Gritsun TS, Gould EA, Holmes EC. (1999) Biological consequences of deletions within the 3′-untranslated region of flaviviruses may be due to rearrangements of RNA secondary structure. Virus Res. 64(2), 107–123.

Ren J, Rastegari B, Condon A, Hoos HH (2005). HotKnots: heuristic prediction of RNA secondary structures including pseudoknots. RNA. 11(10):1494 –1504.

Rivas E, Clements J, Eddy SR. (2017) A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. Nat Methods. 14(1), 45–48. doi: 10.1038/nmeth.4066.

Selisko B, Potisopon S, Agred R, Priet S, Varlet I, Thillier Y, Sallamand C, Debart F, Vasseur JJ, Canard B (2012). Molecular basis for nucleotide conservation at the ends of the dengue virus genome. PLoS Pathog. 8, e1002912. doi: 10.1371/journal.ppat.1002912.

Sessions OM, Wilm A, Kamaraj US, Choy MM, Chow A, Chong Y, Ong XM, Nagarajan N, Cook AR, Ooi EE. (2015) Analysis of Dengue Virus Genetic Diversity during Human and Mosquito Infection Reveals Genetic Constraints. PLoS Negl Trop Dis. 9(9), e0004044. doi: 10.1371/journal.pntd.0004044.

Silva RL, de Silva AM, Harris E, MacDonald GH. (2008) Genetic analysis of Dengue 3 virus subtype III 5' and 3' non-coding regions. Virus Res. 135(2):320-5. doi: 10.1016/j.virusres.2008.03.007.

Shurtleff AC, Beasley DW, Chen JJ, Ni H, Suderman MT, Wang H, Xu R, Wang E, Weaver SC, Watts DM, et al (2001). Genetic variation in the 3' non–coding region of dengue viruses. Virology. 281, 75–87. doi: 10.1006/viro.2000.0748.

Sim S, Aw PP, Wilm A, Teoh G, Hue KD, Nguyen NM, Nagarajan N, Simmons CP, Hibberd ML. (2015) Tracking Dengue Virus Intra-host Genetic Diversity during Human-to-Mosquito Transmission. PLoS Negl Trop Dis. 9(9), e0004052. doi: 10.1371/journal.pntd.0004052.

Steckelberg AL, Akiyama BM, Costantino DA, Sit TL, Nix JC, Kieft JS. (2018) A folded viral noncoding RNA blocks host cell exoribonucleases through a conformationally dynamic RNA structure. Proc Natl Acad Sci U S A. 115(25), 6404–6409. doi: 10.1073/pnas.1802429115.

Steel A, Gubler DJ, Bennett SN. (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(2), 505–512. doi: 10.1016/j.virol.2010.05.033.

Steel A, Gubler DJ, Bennett SN. (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(2), 505–512. doi: 10.1016/j.virol.2010.05.033.

Teoh BT, Sam SS, Tan KK, Johari J, Shu MH, Danlami MB, Abd-Jamil J, MatRahim N, Mahadi NM, AbuBakar S. (2013) Dengue virus type 1 clade replacement in recurring homotypic outbreaks. BMC Evol Biol. 13, 213. doi: 10.1186/1471-2148-13-213.

Vazeille M, Gaborit P, Mousson L, Girod R, Failloux AB. (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. doi: 10.1186/s12879-016-1666-0.

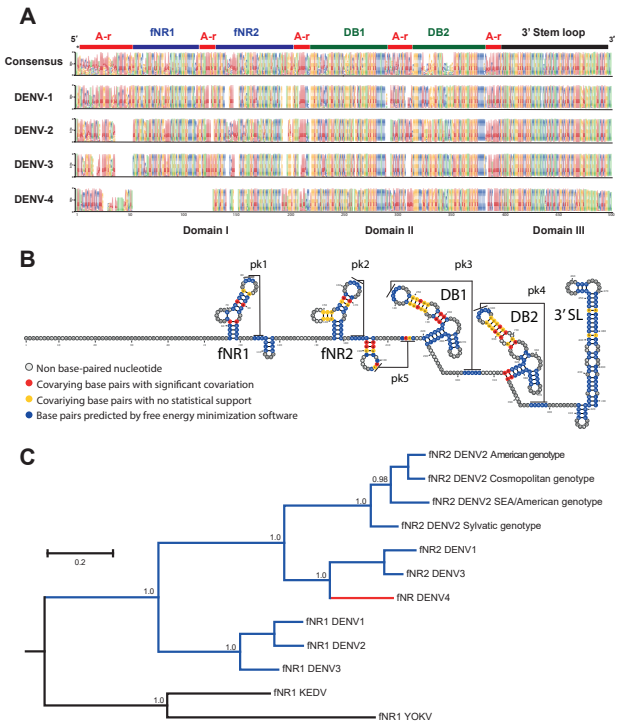
Villordo SM, Carballeda JM, Filomatori CV, Gamarnik AV. (2016) RNA Structure Duplications and Flavivirus Host Adaptation. Trends Microbiol. 24(4), 270–283. doi: 10.1016/j.tim.2016.01.002.

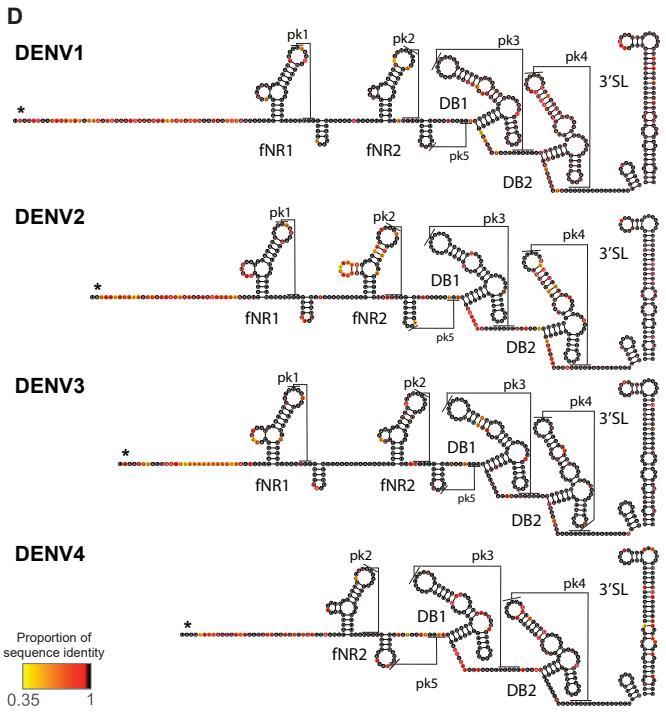
Villordo SM, Filomatori CV, Sánchez-Vargas I, Blair CD, Gamarnik AV. (2015) Dengue virus RNA structure specialization facilitates host adaptation. PLoS Pathog. 11, e1004604. doi: 10.1371/journal.ppat.1004604

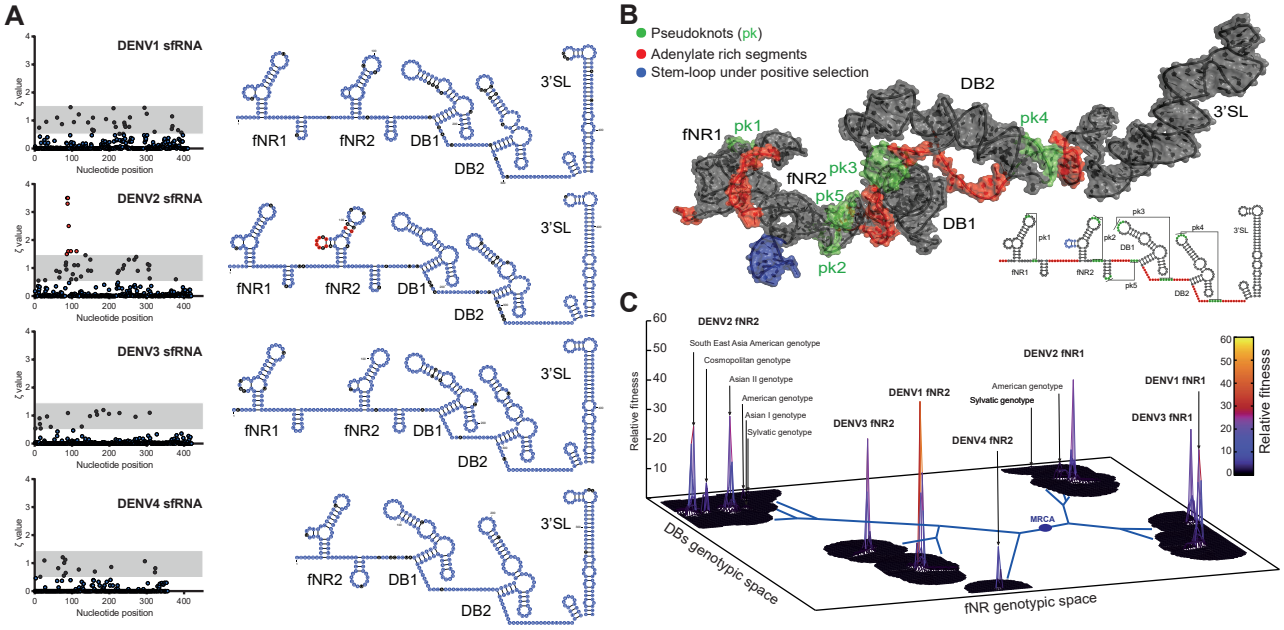
Weaver SC, Vasilakis N. (2009) Molecular evolution of dengue viruses: contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease. Infect Genet Evol. 9(4), 523–540. doi: 10.1016/j.meegid.2009.02.003.

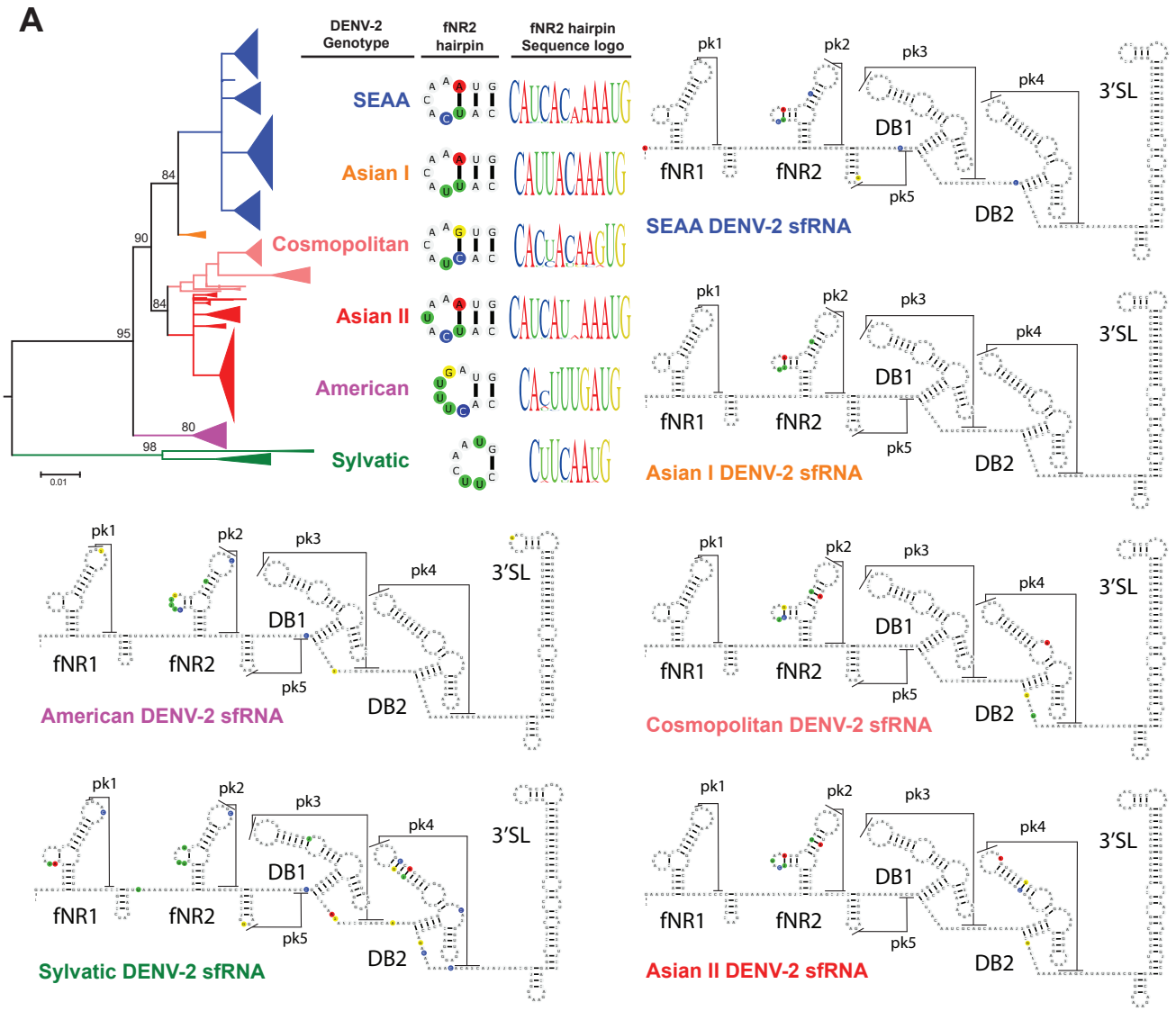
Wong, W. S. W., & Nielsen, R. (2004). Detecting selection in noncoding regions of nucleotide sequences. Genetics, 167(2), 949–958. doi:10.1534/genetics.102.010959

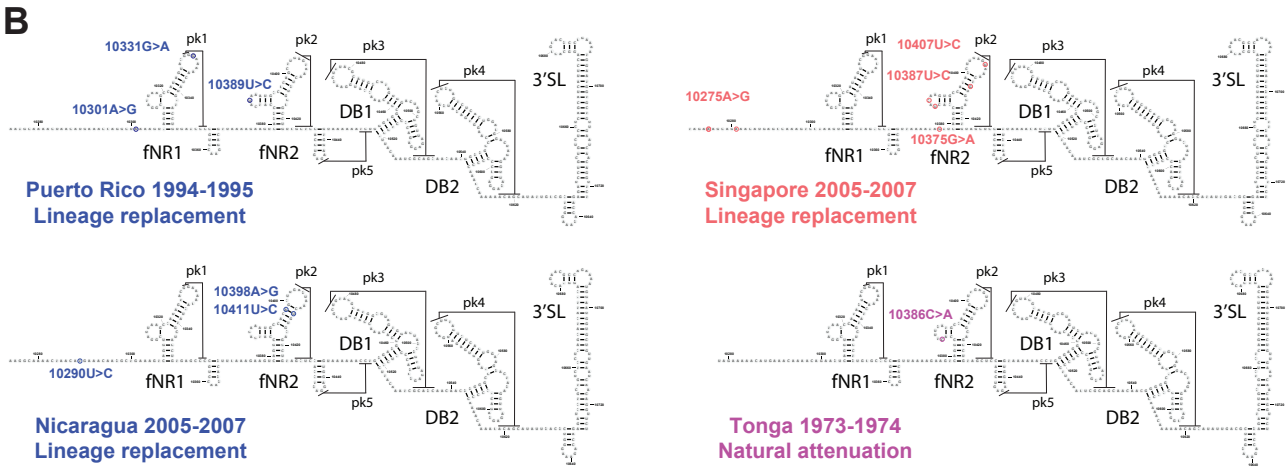
Zhou Y, Mammen MP Jr, Klungthong C, Chinnawirotpisan P, Vaughn DW, Nimmannitya S, Kalayanarooj S, Holmes EC, Zhang C. (2006) Comparative analysis reveals no consistent association between the secondary structure of the 3'-untranslated region of dengue viruses and disease syndrome. J Gen Virol. 87(9), 2595–2603.


A
5'
A-r
fNR1
A-r
fNR2
A-r
DB1
A-r
DB2
A-r
3' Stem loop
3'
Consensus
DENV-1
DENV-2
DENV-3
DENV-4
Domain I
Domain II
Domain III
B
pk1
pk2
pk3
pk4
pk5
DB1
DB2
3' SL
fNR1
fNR2
Non base-paired nucleotide
Covarying base pairs with significant covariation
Covariying base pairs with no statistical support
Base pairs predicted by free energy minimization software
C
fNR2 DENV2 American genotype
fNR2 DENV2 Cosmopolitan genotype
fNR2 DENV2 SEA/American genotype
fNR2 DENV2 Sylvatic genotype
fNR2 DENV1
fNR2 DENV3
fNR DENV4
fNR1 DENV1
fNR1 DENV2
fNR1 DENV3
fNR1 KEDV
fNR1 YOKV
0.2
0.98
1.0


D
DENV1
DENV2
DENV3
DENV4
pk1
pk2
pk3
pk4
pk5
DB1
DB2
3'SL
fNR1
fNR2
*
Proportion of sequence identity
0.35
1


A
DENV1 sfRNA
DENV2 sfRNA
DENV3 sfRNA
DENV4 sfRNA
ζ value
Nucleotide position
fNR1
fNR2
DB1
DB2
3'SL
B
Pseudoknots (pk)
Adenylate rich segments
Stem-loop under positive selection
pk1
pk2
pk3
pk4
pk5
C
DENV2 fNR2
South East Asia American genotype
Cosmopolitan genotype
Asian II genotype
American genotype
Asian I genotype
Sylvatic genotype
DENV3 fNR2
DENV1 fNR2
DENV4 fNR2
DENV2 fNR1
DENV1 fNR1
DENV3 fNR1
MRCA
Relative fitness
DBs genotypic space
fNR genotypic space


A
DENV-2 Genotype
fNR2 hairpin
fNR2 hairpin Sequence logo
SEAA
Asian I
Cosmopolitan
Asian II
American
Sylvatic
84
90
84
95
80
98
0.01
pk1
pk2
pk3
pk4
pk5
DB1
DB2
fNR1
fNR2
3'SL
SEAA DENV-2 sfRNA
Asian I DENV-2 sfRNA
American DENV-2 sfRNA
Cosmopolitan DENV-2 sfRNA
Sylvatic DENV-2 sfRNA
Asian II DENV-2 sfRNA


B
10331G>A
10301A>G
10389U>C
Puerto Rico 1994-1995
Lineage replacement
10275A>G
10407U>C
10387U>C
10375G>A
Singapore 2005-2007
Lineage replacement
10398A>G
10411U>C
10290U>C
Nicaragua 2005-2007
Lineage replacement
10386C>A
Tonga 1973-1974
Natural attenuation

Events	Geographical location	Year	Serotype/Genotype	Context	reference
Clade replacement	Puerto Rico*	1994	DENV-2 SEAA	Outbreak	McElroy et al., 2011
	Nicaragua*	2005	DENV-2 SEAA	Outbreak	OhAinle et al., 2011
	Singapore*	2005	DENV-2 Cosmopolitan	Outbreak	Lee et al., 2012
	Thailand	1990s	DENV-1	Outbreaks	Teoh et al., 2013
	Brazil	1990s	DENV-1 Genotype I	Outbreaks	Carneiro et al., 2013
	Mexico	2006 1999	DENV-1 Genotype III DENV-2 SEAA	Several outbreaks	Carrillo-Valenzo et al., 2010
	India*	2009-2011	DENV-1 genotype III	Outbreaks	Dash et al., 2015
	Malaysia	1987 1997 2004	DENV-1 Genotype I	Recurring outbreaks	Dupont-Rouzeyrol et al., 2014
	Tonga Island*	1974	DENV-2 American	Reduced severity	Steel et al., 2010
Genotype replacement	Sri Lanka*	1989	DENV-3 Genotype I> III	DHF emergence	Silva et al., 2008 Manakkadan et al., 2013
	Myanmar*	1974-2002	DENV-1 Genotype I> III	Several outbreaks	Myat Thu et al., 2005
	Indian Subcontinent*	1971	DENV-2 American>Cosmopolitan	Several outbreaks	Kumar et al., 2010
	Americas*	1983	DENV-2 American>SEAA	DHF emergence	Mir et al., 2014