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).
- Using RNA phylogeny and phylogenetics, we quantified natural selection on this ncRNA.
- Nuclease resistant (NR) structures in DENV 3'UTRs contributed to DENV speciation.
- A highly evolving NR structure appears to increase DENV-2 epidemiological fitness.

Abstract

Changes in dengue viruses (DENV) genome could give rise to fitter strains that spread epidemically. Mutations in the 3' untranslated region (3'UTR) of DENV-2 genome were recently shown to result in increased subgenomic flavivirus RNA (sfRNA) production. sfRNA inhibited host TRIM25 protein activity and reduced interferon (IFN) expression. It may thus enable DENV to reach higher viremia and higher infection rates of blood-feeding Aedes mosquitoes. Whether differences in 3'UTRs shaped DENV evolution remains incompletely understood. Herein, we applied a 'bigdata' approach - we retrieved 3544 dengue virus genomes from NBCI database - and combined RNA sequence covariation, RNA phylogenetics and site specific ncRNA natural selection estimation to gain insights into sfRNA evolution. We found that the second nuclease resistant (NR2) structure of DENV-2 sfRNA has undergone strong positive selection. Conversely, other sfRNA structures are under purifying selection and highly conserved despite sequence divergence. Epidemiological reports also suggest that nucleotide substitutions in NR2 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.

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 (DENV-1 to -4), all of which can cause the full spectrum of disease outcome. A tetravalent dengue vaccine has now been licensed in several countries for use to prevent dengue. However, its protective efficacy varied for each of 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 only licensed 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). While fluctuations in the relative prevalence of the DENV serotypes have caused epidemics in a background of low herd serotype-specific immunity, genetic differences in DENV also plays a distinct role in epidemic emergence (OhAinle et al., 2011). Indeed, we recently 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 resulted in increased sfRNA levels that bound TRIM25 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. Similarly, changes in the 3'UTR sequence that resulted in increased sfRNA production was also observed in the emergence of a new DENV-2 clade in 2005 that also 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 distal domains possess RNA structures necessary for 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 in DB1 generated attenuated DENV-1, -3 and -4 but not DENV-2 strains (Men et al., 1996; Durbin et al., 2001) that appear to be promising live vaccine candidates (Durbin et al., 2013); vaccination with this vaccine candidate 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, there is no evidence to rule out the involvement of other nucleases in the production of sfRNA. Therefore, we refer herein to these structures as nuclease resistant (NR) structures.

As the sequence and hence the RNA structures in the 3'UTR of the DENV genome appear to be an important determinant of epidemiologic 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 DENV 3’UTR.

Results

Sequence identity and RNA structures in the 3' untranslated region of dengue viruses

We aligned the 3'UTR sequences from each of the four DENV serotypes. Our analysis confirmed the existence of substantial differences in the sequence and length of the 3'UTR across and within serotypes as previously reported (Shurtleff et al., 2001; Proutski et al., 1999; Men et al., 1996). The 3'UTR of DENV-1 serotype is the longest (Mode 465; range 436–475) followed by DENV-2 (Mode 454; range 444–469), DENV-3 (Mode 443; range 429–455), and DENV-4 (Mode 387; range 387–407) **(Appendix Table 1A)**. As expected, 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 duplications, 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 ($p<0.05$) **(Appendix Table 1B)**. 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 appears to possess a semi variable region with a high level of nucleotide conservation (Average Identity ≥ 96% in all serotypes) **(Figure 1A and Appendix Table 1B),** suggesting the presence of conserved RNA structures across DENV serotypes. These conserved stretches appear 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, we adopted a 'divide, learn and conquer' approach. This strategy 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) validate, improve and build 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, we also applied G-test statistics to determine whether these covariations occur at a rate higher than phylogenetically expected **(Figure EV1)** (Rivas et al., 2017). The consensus DENV 3'UTR secondary structures derived from RNA phylogeny **(Figure 1B)** resembled the DENV-2 3'UTR structure previously obtained by chemical probing (Chapman et al., 2014A), suggesting the validity of our bioinformatics approach. The predicted NR structures in Domain I are compatible with the known crystal structures and substantially differs from the secondary structures suggested by Shurtleff et al., 2001, and to a lesser extent, from the one reported by Villordo et al., 2015. Our approach also predicted a novel RNA pseudoknot (pk5) in the variable region of all serotypes of DENV **(Figure 2B and 2D).**

NR structures and DENV evolution

Interestingly, DENV-1, -2 and -3 but not DENV-4 bear two NR structures **(Figure 1D)**. Among the duplicated NR structures, the first NR structure (NR1) exhibited greater conservation than NR2, suggesting that NR1 exerts the main nuclease resistant function for sfRNA production and is therefore necessarily highly conserved. The downstream NR (NR2) structure has thus relatively less constraint to evolve and adapt its structure possibly for additional function. The only NR structure in DENV4 appeared to be more related to the NR2 than NR1 structure of the other DENV serotypes **(Figure 1C)**.

To determine whether the variability of NR structures influence DENV evolution, we examined the selection pressure on each of the 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 a higher or lower substitution rate than the synonymous substitution rate in the

coding region of the genome, respectively. This approach thus provides an estimate on whether any given nucleotide position has undergone positive or negative selection relative to the coding region of the genome. Figure 2A shows the ζ values for every position in the four DENV sfRNA along with the 95% confidence intervals of the ζ parameter calculated for the coding region of the 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 DENV-1 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 2)**. This finding suggests that substitutions in NR2 may confer competitive advantage for DENV-2 strains.

To further assess the role of DENV-2 3'UTR evolution, a ML phylogenetic tree was constructed using a nucleotide substitution model of evolution for non-coding RNA, based on the alignment and consensus secondary RNA structures of DENV-2 sfRNA. The phylogenetic tree **(Figure 3A)** segregated the DENV-2 sfRNA sequences into six clades, consistent with the six genotypes that characterize DENV-2 evolution (Reviewed by Chen and Vasilakis, 2011). Remarkably, the positively selected hairpin in the NR2 structure differed in nucleotide composition and structure across DENV-2 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 DENV-2 genotypes in many part 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 NR2 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 thus provides a bioinformatic approach to glean insights to direct further 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.

Our findings indicate that the NR structures in the DENV 3'UTR are relatively well conserved. Any nucleotide substitution in the stem positions of the NR structures is often accompanied by compensatory mutation to maintain structural integrity. Against this backdrop, the finding of a positive selection in the NR2 structure of DENV 3'UTR is thus intriguing. NR2 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 DENV-2 in *Aedes albopictus* derived C6/36 cells resulted in multiple mutations in its NR2 and production of different sfRNA species. These mutations could have then been purified in subsequent human infection possibly based on their ability to bind host proteins for the suppression antiviral immune activation. Indeed, Bidet et al. (2014) showed that NR2 structure interacts with CAPRIN protein, mediating the sfRNA-induced repression of interferon-stimulated mRNAs in human-derived Huh7 cells. Moreover,

mutations in the NR2 structures produced higher replicative fitness in *Aedes albopictus* compared to the corresponding wild type DENV2 (Filomatori et al., 2017). Collectively, these findings indicate a strong evolutionary pressure on DENV-2 NR2. 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 NR 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 free to evolve and be selected based on advantageous RNA-protein interactions in human or mosquito cells for increased fitness. This evolutionary model would be consonant with the reduced sfRNA production and transmission fitness that NR1 mutations caused in DENV2 strains (Pompon et al. 2017), and the increased replicative, transmission and epidemiologic fitness that NR2 mutations conferred to some DENV-2 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 DENV-2 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 explains the lack of positive selection in DENV-4 as these viruses only possess one NR structure. It, however, raises questions on why no positive selection was seen on the 3'UTR of either DENV-1 or -3. We offer several interpretations. Firstly, if we assume that the observed distinct NR adaptation in DENV-2 sfRNA to interspecies transmission reported by 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 on the 3'UTR of DENV-1, -3 and -4 as compared to DENV2. A second and interesting scenario would suggest that the proposed model for the evolution of DENV2 sfRNA and its distinct NR adaptations do not occur in other DENV types, which indeed is already unlikely to occur in

DENV 4 – since its 3'UTR only possess one NR structure. This second postulate would help us understand why previous studies using RNA sequencing observed mutational hotspot in the 3'UTR of DENV-2 but not DENV-1 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 NR duplication has contributed to shape the overall divergence of dengue viruses. If the NR duplication occurred early during DENV evolution – as the RNA phylogenetic analysis suggested, it is likely that the later NR1 deletion in DENV4 imposed an evolutionary constraint in DENV-4 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 DENV-4 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 generated a complete attenuated phenotype in DENV-4 as well as in recombinant DENV-1 to -3 strains bearing a Δ30rDENV4 -3'UTR (Durbin et al., 2001 and 2013).

Collectively, these bioinformatic and previously published experimental data suggest that 3'UTR evolution and sfRNA development are important determinants of DENV adaptation, survival and epidemiological fitness. Our findings underscore the need for greater awareness in the field about the evolutionary, epidemiological and clinical consequences of mutations in the 3'UTR of DENV and other flaviviruses.

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 V. Dhanasekaran and M. Garcia-Blanco for their invaluable advice throughout this work.

Conflict of Interest

The authors declare that they have no conflict of interest.

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 solve 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 NR structures was optimized through comparative RNA modeling, using ZIKV NR 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)**.

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 colour-coded (Blue = Cytosine, Green = Uracil, Yellow = Guanine, Red = Adenine). Five conserved stretches were found across DENV 3'UTRs. They correspond to two nuclease resistant (NR) 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 NR structures in the 3'UTR of dengue viruses. As the sequence logos revealed (A), DENV4 3'UTR bears only NR structures, to determine whether this structure shares its most recent common ancestor with the NR1 or NR2 structures in other dengue viruses, Bayesian RNA phylogenetics was implemented under PHASE 3.0 software. It included all DENV NR structure sequences (Branches in blue) and a DENV4 NR (Branch in red), the NR 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 colour-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, the ribonucleotide positions in the DENV sfRNA secondary structures are colour-coded accordingly as well. 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 colours as in the secondary structures at the bottom right of the figure.

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 DENV-2 strain. A maximum-likelihood phylogenetic tree from DENV2 sfRNA was built using PHASE 3.0 software. This software applies a 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 DENV-2 genotypes.

(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 EV1. 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.

Appendix Table 1. 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 ($p<0.05$) (red). A 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.

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.

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.

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.

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.

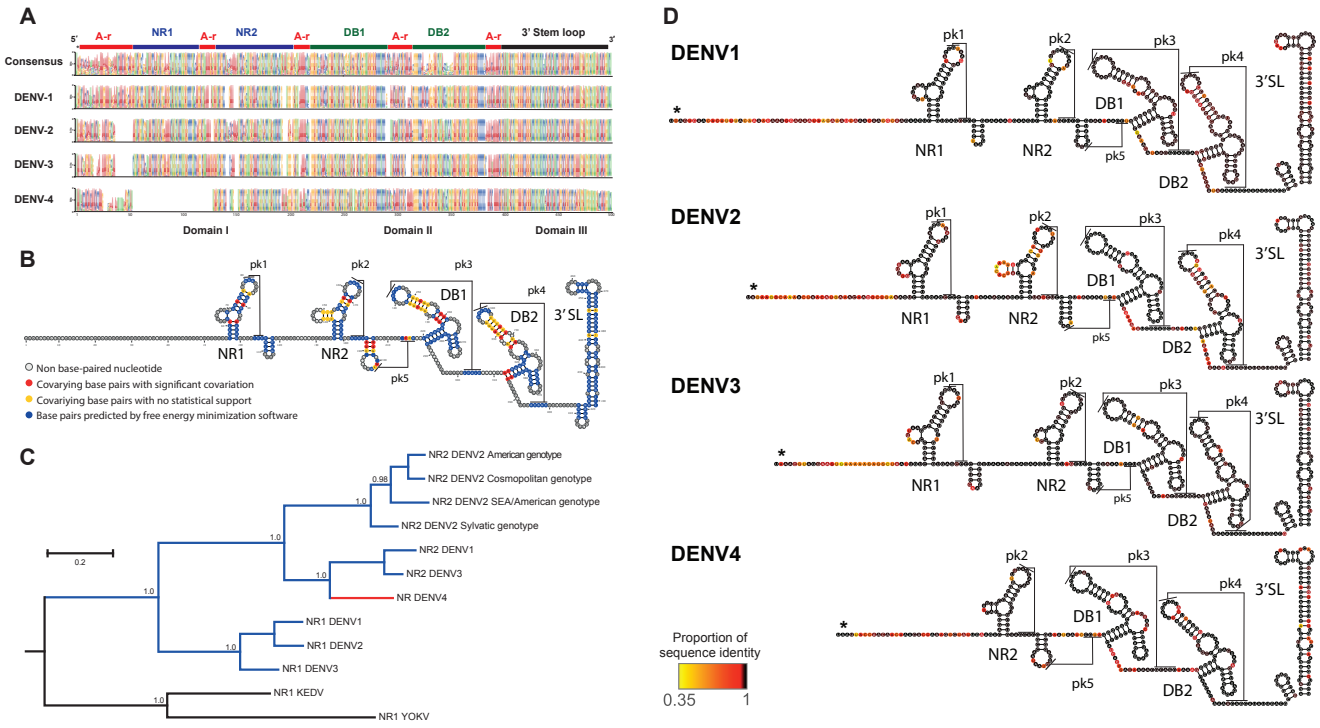
Villordo SM, Carballeda JM, Filomatori CV, Gamarnik AV. (2016) RNA Structure Duplications and Flavivirus Host Adaptation. Trends Microbiol. 24(4), 270–283. doi: 10.1016/j.tim.2016.01.002.

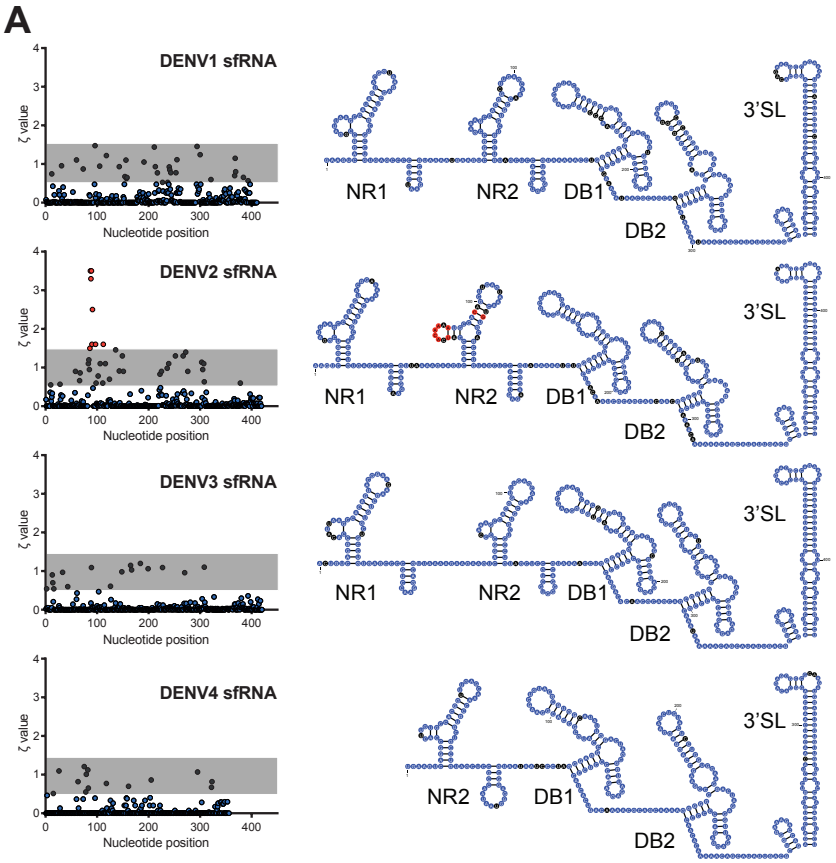
Villordo SM, Filomatori CV, Sánchez-Vargas I, Blair CD, Gamarnik AV. (2015) Dengue virus RNA structure specialization facilitates host adaptation. PLoS Pathog. 11, e1004604. doi: 10.1371/journal.ppat.1004604

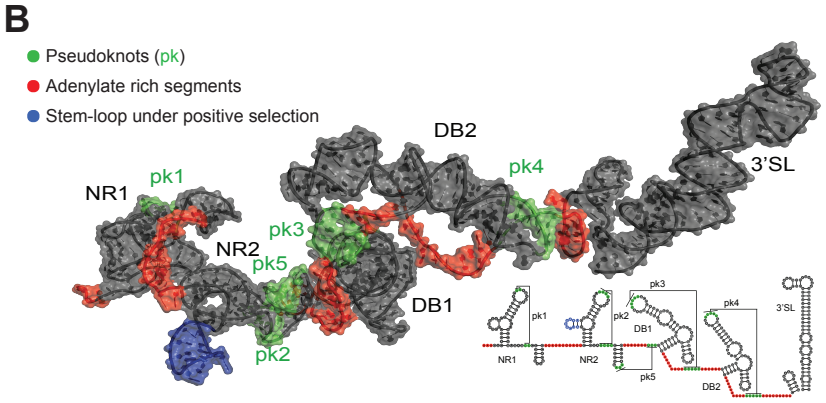
Weaver SC, Vasilakis N. (2009) Molecular evolution of dengue viruses: contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease. Infect Genet Evol. 9(4), 523–540. doi: 10.1016/j.meegid.2009.02.003.

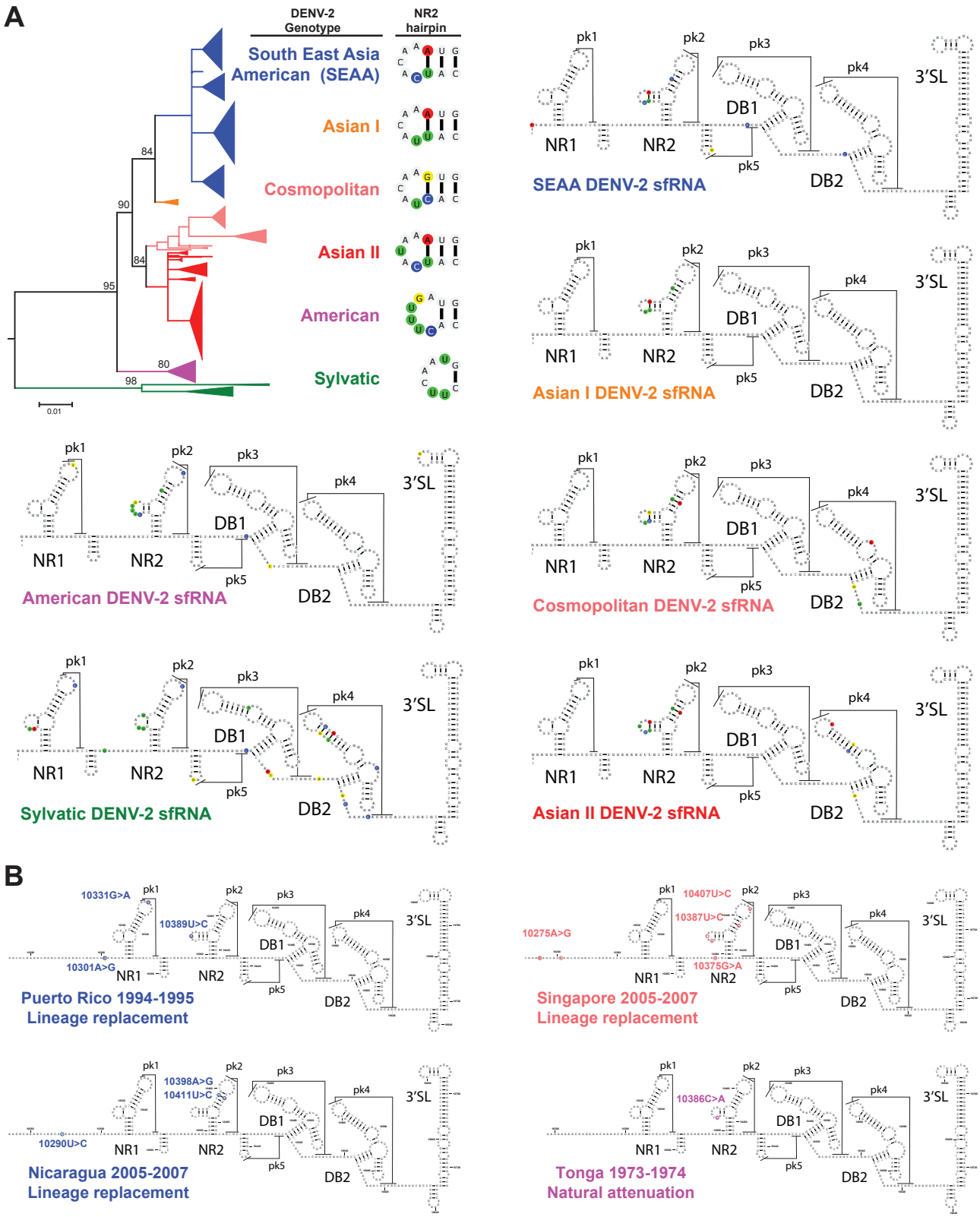
Wong, W. S. W., & Nielsen, R. (2004). Detecting selection in noncoding regions of nucleotide sequences. Genetics, 167(2), 949–958. doi:10.1534/genetics.102.010959

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
NR1
A-r
NR2
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
NR1
NR2
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
NR2 DENV2 American genotype
NR2 DENV2 Cosmopolitan genotype
NR2 DENV2 SEA/American genotype
NR2 DENV2 Sylvatic genotype
NR2 DENV1
NR2 DENV3
NR DENV4
NR1 DENV1
NR1 DENV2
NR1 DENV3
NR1 KEDV
NR1 YOKV
0.98
1.0
0.2
D
DENV1
DENV2
DENV3
DENV4
Proportion of sequence identity
0.35
1

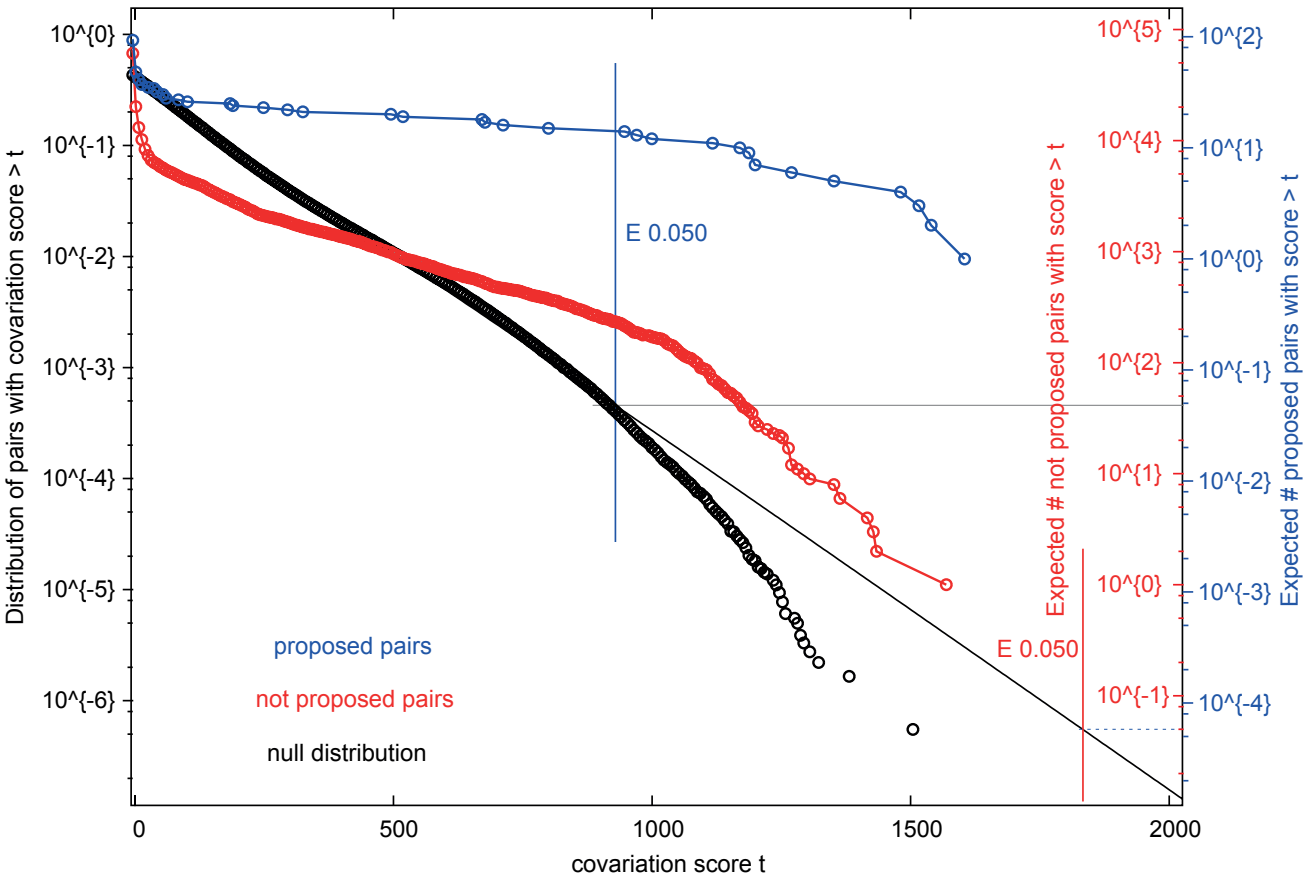

A
DENV1 sfRNA
DENV2 sfRNA
DENV3 sfRNA
DENV4 sfRNA
ζ value
Nucleotide position
NR1
NR2
DB1
DB2
3'SL


B
Pseudoknots (pk)
Adenylate rich segments
Stem-loop under positive selection
NR1
pk1
NR2
pk3
pk5
pk2
DB1
DB2
pk4
3'SL


A
DENV-2 Genotype
NR2 hairpin
South East Asia American (SEAA)
Asian I
Cosmopolitan
Asian II
American
Sylvatic
84
90
84
95
80
98
0.01
pk1
pk2
pk3
pk4
pk5
DB1
DB2
NR1
NR2
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
10389U>C
10301A>G
Puerto Rico 1994-1995
Lineage replacement
10407U>C
10387U>C
10275A>G
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

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