

RESEARCH ARTICLE

# Nonstructural Proteins Are Preferential Positive Selection Targets in Zika Virus and Related Flaviviruses

Manuela Sironi<sup>1</sup>, Diego Forni<sup>1</sup>, Mario Clerici<sup>2,3</sup>, Rachele Cagliani<sup>1\*</sup>

**1** Scientific Institute IRCCS E. MEDEA, Bioinformatics, Bosisio Parini, Italy, **2** Department of Physiopathology and Transplantation, University of Milan, Milan, Italy, **3** Don C. Gnocchi Foundation ONLUS, IRCCS, Milan, Italy

\* [rachele.cagliani@bp.lnf.it](mailto:rachele.cagliani@bp.lnf.it)



CrossMark  
click for updates

## OPEN ACCESS

**Citation:** Sironi M, Forni D, Clerici M, Cagliani R (2016) Nonstructural Proteins Are Preferential Positive Selection Targets in Zika Virus and Related Flaviviruses. PLoS Negl Trop Dis 10(9): e0004978. doi:10.1371/journal.pntd.0004978

**Editor:** Ann M Powers, Centers for Disease Control and Prevention, UNITED STATES

**Received:** April 6, 2016

**Accepted:** August 14, 2016

**Published:** September 2, 2016

**Copyright:** © 2016 Sironi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Viral coding sequences were retrieved from the NCBI database, GenBank IDs are reported in the Supporting Information.

**Funding:** The authors received no specific funding for this work.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

The *Flavivirus* genus comprises several human pathogens such as dengue virus (DENV), Japanese encephalitis virus (JEV), and Zika virus (ZIKV). Although ZIKV usually causes mild symptoms, growing evidence is linking it to congenital birth defects and to increased risk of Guillain-Barré syndrome. ZIKV encodes a polyprotein that is processed to produce three structural and seven nonstructural (NS) proteins. We investigated the evolution of the viral polyprotein in ZIKV and in related flaviviruses (DENV, Spondweni virus, and Kedougou virus). After accounting for saturation issues, alignment uncertainties, and recombination, we found evidence of episodic positive selection on the branch that separates DENV from the other flaviviruses. NS1 emerged as the major selection target, and selected sites were located in immune epitopes or in functionally important protein regions. Three of these sites are located in an NS1 region that interacts with structural proteins and is essential for virion biogenesis. Analysis of the more recent evolutionary history of ZIKV lineages indicated that positive selection acted on NS5 and NS4B, this latter representing the preferential target. All selected sites were located in the N-terminal portion of NS4B, which inhibits interferon response. One of the positively selected sites (26M/I/T/V) in ZIKV also represents a selection target in sylvatic DENV2 isolates, and a nearby residue evolves adaptively in JEV. Two additional positively selected sites are within a protein region that interacts with host (e.g. STING) and viral (i.e. NS1, NS4A) proteins. Notably, mutations in the NS4B region of other flaviviruses modulate neurovirulence and/or neuroinvasiveness. These results suggest that the positively selected sites we identified modulate viral replication and contribute to immune evasion. These sites should be prioritized in future experimental studies. However, analyses herein detected no selective events associated to the spread of the Asian/American ZIKV lineage.

## Author Summary

Zika virus is mainly transmitted by mosquitoes and is phylogenetically related to other human pathogens (e.g. dengue virus). After the outbreak in South America, the WHO declared that the spread of ZIKV should be regarded as a public health emergency. In fact, growing evidence suggests that ZIKV infection during pregnancy increases the risk of congenital birth defects. Moreover, ZIKV can trigger Guillain-Barré syndrome, a severe neurological disorder characterized by progressive muscle weakness. Evolutionary studies can help identify sites that allow viral adaptation—i.e. that increase viral fitness at least in some conditions. We analyzed the evolution of the polyproteins encoded by ZIKV and by related viruses and identified several sites in nonstructural proteins that were subject to natural selection. Most of these are located in protein regions that mediate interaction with the host immune system or that may regulate viral RNA synthesis. In ZIKV isolates, the NS4B protein was the preferential selection target with three selected residues. Changes at these sites are expected to modulate some aspect of viral fitness, either in mosquitoes or vertebrate hosts. Future studies to clarify the mechanisms of ZIKV pathogenicity should address the role of these sites in the modulation of viral phenotypes.

## Introduction

The *Flavivirus* genus (family *Flaviviridae*) comprises a large number of viral species, many of which are important human pathogens; these include dengue virus (DENV), yellow fever virus (YFV), Japanese encephalitis virus (JEV), West Nile virus (WNV), and the latest emerged pathogen, Zika virus (ZIKV).

ZIKV was first discovered in 1947 in Uganda, in a sentinel rhesus monkey, and subsequently in mosquitoes of the *Aedes* genus. Between 1947 and 2006, only sporadic human cases were reported in Africa and in Southeast Asia, until multiple outbreaks in the Pacific islands occurred. The first sizable outbreak was reported in the Federated States of Micronesia (Yap Island) in 2007, followed by an outbreak in French Polynesia in 2013. In 2014, the epidemic spread to Cook Islands, New Caledonia and Easter Island, and reached South America in late 2014–early 2015 [1–3]. As of May 18, 2016, sixty countries/territories have reported ZIKV cases (<http://www.who.int/emergencies/zika-virus/situation-report/en/>).

Although ZIKV infection is often asymptomatic or causes only mild symptoms, the WHO declared that the spread of ZIKV should be regarded as a public health emergency of international concern. In fact, growing evidence suggests that ZIKV infection during pregnancy increases the risk of microcephaly, brain damage, and congenital abnormalities [4–6]. Also, retrospective studies indicated that ZIKV can trigger Guillain-Barré syndrome (GBS), a severe neurological disorder characterized by progressive muscle weakness [7]. Moreover, even if *Aedes* mosquitoes species such as *Aedes aegypti* and *Aedes albopictus* represent the primary vectors for natural transmission, perinatal and congenital infections, as well as sexual transmission and infection through blood transfusion have been recently documented [1].

ZIKV is a member of the Spondweni (SPOV) serocomplex and, like other members of the *Flavivirus* genus, it is a single-stranded positive-sense RNA virus. Its genome consists of about 11,000 nucleotides with two flanking non-coding regions and a single long open reading frame. The encoded polyprotein is co- and post-translationally processed by viral and host proteases to produce three structural (capsid, C; pre-membrane, prM; envelope, E) and seven nonstructural (NS) proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5) [8].

Genetic and phylogenetic studies indicated that ZIKV has evolved into 2 major lineages: African and Asian/American, this latter responsible of the recent outbreaks and associated with reports of GBS and fetal malformations [9, 10].

Analysis of ZIKV genomes from microcephaly cases revealed no shared amino acid changes, suggesting that viral genetic features alone are not responsible for fetal abnormalities [2]. Likewise, inspection of amino acid differences between the Asian and African lineages provided no clear indication of viral genetic features that may result in altered virulence or increased pathogenicity, although no functional study of these variants has been performed to date [2]. It was thus proposed that, if the link with GBS and fetal abnormalities is confirmed, factors other than viral genetics, including infection with other viruses or the host genetic background [2], are responsible for these adverse effects. An alternative possibility is that all ZIKV lineages increase the risk of microcephaly and/or GBS, but the association has been previously missed due to the limited size of African outbreaks and to the lack of surveillance programs. Whereas addressing these questions will require extensive epidemiological and clinical surveys, analysis of all ZIKV strains and of their evolution within the wider perspective of closely related flaviviruses can identify positively selected amino acid changes. These latter are expected to entail a functional effect and can therefore be prioritized in further studies of viral pathogenesis. Indeed, evolutionary analyses in WNV have detected positively selected changes that modulate viral phenotypes such as virulence [11] and superinfection exclusion [12].

Herein we investigated the evolution of the viral polyprotein in ZIKV and in related flaviviruses. Results indicate that NS1 was a major selection target during flavivirus speciation and revealed ongoing selection in ZIKV strains in NS4B and NS5.

## Methods

### Sequences and alignments

Coding sequences were retrieved from the NCBI database (<http://www.ncbi.nlm.nih.gov/>), all flaviviruses analyzed were selected to have full coding sequence information. A list of accession number is reported in [S1 Table](#).

Alignment errors are common when divergent sequences are analyzed and can affect evolutionary inference. Thus, we used PRANK [13] to generate multiple sequence alignments and GUIDANCE2 [14] for filtering unreliably aligned codons (we masked codons with a score <0.90), as suggested [15].

### Substitution saturation and recombination

Substitution saturation was checked using the Xia's index implemented in DAMBE [16, 17]; this test compares an entropy-based index of saturation ( $I_{ss}$ ) with a critical value ( $I_{ss.c}$ ). If  $I_{ss}$  is significantly lower than  $I_{ss.c}$ , sequences have not experienced substitution saturation.

The presence of recombination was assessed using two methods, GARD [18] and Recco [19]. Whereas GARD uses phylogenetic incongruence among segments in the alignment to detect the best-fit number and location of recombination breakpoints, Recco is based on cost minimization and dynamic programming. In GARD, the statistical significance of putative breakpoints is evaluated through Kishino-Hasegawa (HK) tests; breakpoints were considered significant if their  $p$  value was  $< 0.05$ .

The Recco's output includes a  $p$  value for the whole dataset that, controlling for false positives, provides an indication as to whether a significant amount of recombination is detectable in the whole alignment. We concluded that no substantial recombination was present when the dataset  $p$  value was  $> 0.05$ . For alignments showing evidence of recombination in Recco (dataset  $p$  value  $< 0.05$ ), we considered sequences as recombinants if the number of savings was

>20, and the sequence  $p$  value was <0.001, as suggested [20]. Recombination breakpoints were defined accordingly.

## Evolutionary analyses

Phylogenetic trees were reconstructed using the program phyML with a maximum-likelihood approach, a General Time Reversible (GTR) model plus gamma-distributed rates and 4 substitution rate categories. Branch support was evaluated using a non parametric bootstrap analysis (100 replicates) [21].

The nonsynonymous/synonymous rate ratio ( $dN/dS$  or  $\omega$ ) is a widely used method to detect positive selection. Positive selection is inferred when the rate of nonsynonymous ( $dN$ ) substitutions is higher than that of synonymous ( $dS$ ) substitutions ( $dN/dS > 1$ ).

To test for the action of episodic positive selection in flaviviruses, we applied the branch-site test [22] from the codeml software [23]. The test estimates selective pressure changes among branches and sites in the phylogenetic tree. Two nested models (MA and MA1) are compared: MA allows positive selection on one or more lineages (called foreground lineages), and the MA1 does not allow such positive selection. Twice the difference of likelihood for the two models ( $\Delta \ln L$ ) is compared to a  $\chi^2$  distribution with one degree of freedom [22]. A false discovery rate correction was applied to take into account a multiple hypothesis issue generated by analyzing different branches on the same phylogeny [24].

A Bayes Empirical Bayes (BEB) analysis was used to evaluate the posterior probability that each codon belongs to the site class of positive selection on the foreground branch, only when  $2\Delta \ln L$  was statistically significant.

BUSTED (branch-site unrestricted statistical test for episodic diversification) [25] is an alternative approach implemented in the HyPhy package [26] designed to describe episodic positive selection acting on specific branches in the phylogenetic tree at a proportion of sites. A model that allows the action on positive selection on foreground branches is compared with a null model that doesn't allow  $\omega > 1$ . Twice the  $\Delta \ln L$  of the two models is then compared to a  $\chi^2$  distribution (with two degrees of freedom); if the null model is rejected, at least one site is under positive selection on the foreground branches. To detect selection at individual sites, twice the difference of the likelihood for the alternative and the null model at each site is compared to a  $\chi^2$  distribution (one degree of freedom).

To be conservative, we considered a site under episodic positive selection if it showed both a  $p$  value  $\leq 0.05$  in BUSTED and a BEB posterior probability  $\geq 0.90$ .

To better understand the evolution of ZIKV genomes, we also applied two random site (NSsite) models implemented in codeml: a null model (M7) that assumes that  $0 < \omega < 1$  and is beta distributed among sites in all branches of the phylogeny, and a positive selection model (M8); this model is the same as M7 but also includes an extra category of sites in the alignment with  $\omega > 1$ . A  $\chi^2$  distribution is used to assess statistical significance of  $2\Delta \ln L$  of the two models.

Positively selected sites were identified using the posterior probability ( $\geq 0.90$ ) from M8 BEB.

Individual sites under diversifying positive selection were also identified using Random effects likelihood (REL) [27] and Fast Unconstrained Bayesian AppRoximation (FUBAR) [28] methods from the HyPhy package.

REL estimates  $\omega$  at each site by inferring a gene distribution for both synonymous and nonsynonymous rate variations and assuming independent draw at each site from this distribution. We considered a site under positive selection if it showed a Bayes Factor  $> 50$ .

FUBAR is an approximate hierarchical Bayesian method that uses an unconstrained distribution of selection parameters by averaging over a large number of predefined site classes.

Given this distribution, FUBAR estimates the posterior probability of positive diversifying selection at each site in the alignment (with a cutoff  $\geq 0.90$ ).

In order to be conservative, we finally considered a site as under diversifying positive selection if it was detected by at least two different methods.

### Immune epitopes

Data on DENV experimentally verified immune epitopes were obtained from the NIAID Virus Pathogen Database and Analysis Resource (ViPR) online (<http://www.viprbrc.org>) [29].

Human epitopes were searched for by using the gene product name as a query. Linear epitopes with positive results in any assay type category (B cell, T cell and MHC binding) were included. We used ClustalOmega [30] to align epitopes onto the DENV protein sequence and from this onto the ZIKV sequence.

### Protein 3D structures, homology modeling, and membrane topology

The structure of NS1 of ZIKV was obtained by homology modeling using the West Nile virus NS1 (PDB ID: 4O6C) structure as a template; analysis was performed through the SWISS-MODEL server [31]. The accuracy of the model was assessed with VADAR (Volume, Area, Dihedral Angle Reporter), which uses several algorithms to calculate different parameters for individual residues and for the entire protein [32].

Images were created using PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.2 Schrödinger, LLC).

The membrane protein topology for the ZIKV NS4B protein was predicted by using TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>) [33].

## Results

### NS1 is the major target of positive selection during flavivirus speciation

As mentioned above, ZIKV belongs to the Spondweni group of mosquito-borne flaviviruses. In addition to Spondweni virus (SPOV), the viral species more closely related to ZIKV include the Kedougou (KEDV) and dengue (DENV) viruses [10]. To investigate selective events that took place during the speciation of ZIKV and closely related flaviviruses, we obtained complete coding sequence information for 21 ZIKV, 1 SPOV and 1 KEDV, as well as 11 DENV. ZIKV sequences were selected to represent viruses sampled in both African and in non-African countries, in distinct years, and from different hosts (see [S1 Table](#)). In the case of KEDV and SPOV, only one complete genome is available for each virus. For DENV, sequences were selected to belong to the four major serotypes (DENV1 to DENV4); for each serotype, sequences representative of the most common genotypes (based on complete E nucleotide sequences in [34]) were included. DENV sequences were also selected to cover different geographic locations and isolation dates.

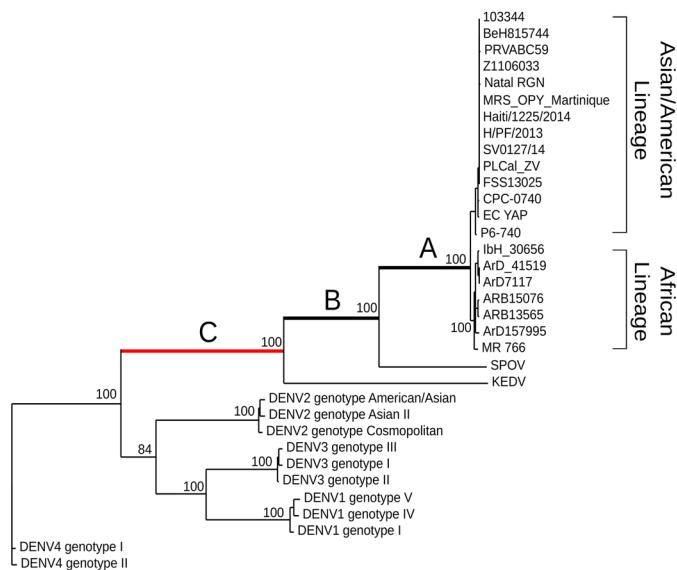
The structural and nonstructural coding regions were analyzed separately, and the alignments were pruned of unreliably aligned codons using the GUIDANCE utility (see [Methods](#)). This procedure resulted in the masking of 8.6% and 9.5% of codons in the structural and non-structural regions, respectively. A test for substitution saturation was performed using Xia's method [16] and indicated no substantial saturation in either alignment ([S2 Table](#)).

We next analyzed the alignments for the presence of recombination using two different methods, GARD (Genetic Algorithm Recombination Detection) [18] and Recco [19]. No evidence of recombination in the nonstructural region was detected using either program, whereas Recco (but not GARD) suggested the presence of a recombination breakpoint around

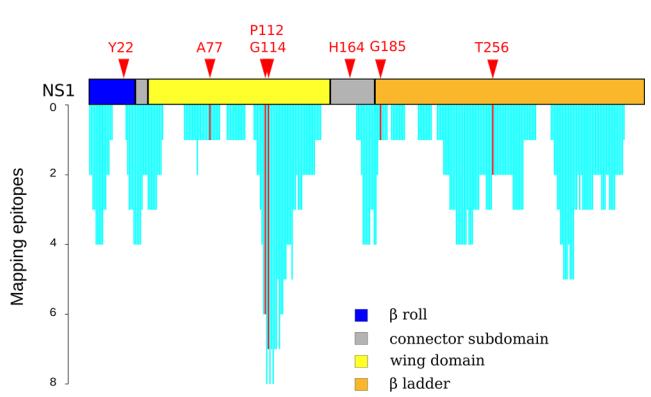
A



B



C



**Fig 1. Different selective pressures acting on flavivirus polyproteins.** (A) Schematic representation of the ZIKV polyprotein. Proteins are colored in hues of blue depending on the percentage of negatively selected sites in ZIKV strains. The location of recombination breakpoints in flaviviruses and ZIKV is shown by striped rectangles. Positively selected sites in the flavivirus phylogeny and in ZIKV strains are colored in red and green, respectively. (B) Maximum likelihood unrooted tree for the flavivirus phylogeny. Branches analyzed in the branch-site tests are indicated with capital letters, with red indicating statistically significant evidence of positive selection. Branch length is proportional to nucleotide substitutions per codon. Bootstrap values for internal branches >75% are shown. See [S1 Table](#) for accession number and full names of analyzed viruses. (C) Immune epitope mapping and schematic representation of NS1 domains. Cyan bars indicate the number of epitopes overlapping each NS1 residue. Positively selected sites are also shown.

doi:10.1371/journal.pntd.0004978.g001

position 1350–1360 (relative to AY632535 coding sequence) in the E region ([Fig 1A](#)). The structural region was thus split into two sub-regions for the following analyses, so as to avoid false positive inferences of positive selection as a result of unaccounted recombination [35].

Phylogenetic trees of the three regions were obtained with phyML. Trees were very similar and fully consistent with previously reported phylogenies for flaviviruses [36], with African and non-African ZIKV isolates forming distinct branches [3] ([Fig 1B](#)).

We next searched for evidence of episodic positive selection along the internal branches of flavivirus phylogenies using branch-site tests ([Fig 1B](#)). Two different methods were applied to ensure consistency: the branch-site unrestricted statistical test for episodic diversification (BUSTED) [25] and the maximum-likelihood models (MA/MA1) implemented in the PAML suite [23]. These two approaches rely on different assumptions of  $\omega$  (nonsynonymous/

synonymous rate ratio) variation among branches. Episodic positive selection at each tested branch was declared when statistically significant support was obtained with both methods. Using this criterion, we found no evidence of positive selection in the structural region. Conversely, both tests detected evidence of positive selection on one branch in the phylogeny of the nonstructural region ([Fig 1](#) and [Table 1](#)). Selected sites along this branch were identified using the Bayes empirical Bayes (BEB) procedure from model MA and with BUSTED; again, only sites detected by both methods were considered. A total of 16 positively selected sites were detected ([Fig 1A](#)); notably, seven of such sites are located in the NS1 protein. To test whether this number is higher than expected, we performed random sampling across codons in the nonstructural region (i.e. we assumed that all codons that were not masked by GUIDANCE in any sequence have the same probability of being called as positively selected). Results indicated that the likelihood of having 7 selected sites in NS1 amounts to 0.0039; thus, this protein represented a preferential selection target during flavivirus speciation.

We note that the percentage of codons masked by GUIDANCE ranged widely among protein regions, from 3.3% in NS3 to more than 30% in the 2k and NS2A region ([S3 Table](#)). Whereas this does not affect the significant enrichment we obtained for NS1 (as we accounted for masked codons), the power to detect selection in extensively masked regions is clearly reduced; in fact, the ultimate result of masking is fewer codons available for analysis or a shallow phylogeny at available codons.

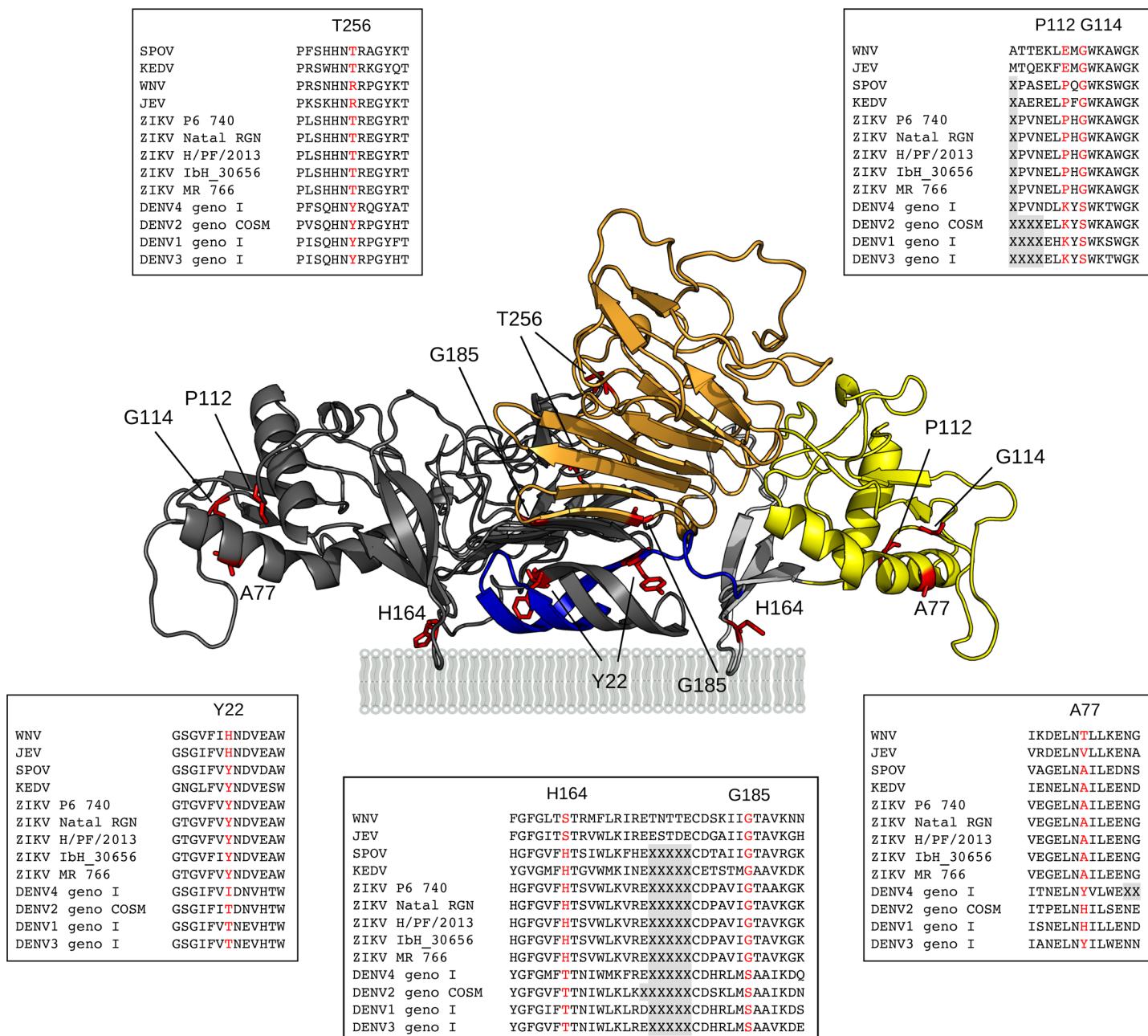
In NS1, the 7 positively selected sites are distributed along the entire protein region ([Fig 1C](#)). To gain insight into the role and the spatial localization of these sites, we performed homology modeling of ZIKV NS1 using the West Nile virus protein structure (PDB ID: 4O6C) as a template ([Fig 2](#)). We also retrieved the location of experimentally validated immune epitopes in NS1. Selected sites 112 and 114 map on a disordered loop of the wing domain; this loop is exposed, and several DENV immune epitopes were described in this region, most of them covering both positions ([Fig 1C](#)). In DENV, an alanine mutation at the 114 site affects virus particle production [37]. This residue was also suggested to have a role in the interaction between NS1 and the envelope glycoprotein [37]. Interestingly, three sites (residues 77, 112, 114) in the “wing” domain localize in close proximity on the protein structure ([Fig 2](#)), suggesting that they are involved in formation/stabilization of the same interactions. NS1 position 164 is located in a hydrophobic protruding loop, flanking a smaller loop that is essential for DENV viral replication [38]. Mutations in flanking positions (residues 160 and 162) affect both RNA synthesis and virus viability [38]. Indeed, residues 159–162 of the connector domain together with the β-roll (where the Y22 selected site maps) form a hydrophobic protrusion that faces the membrane ([Fig 2](#)). This hydrophobic structure is thought to be involved in the interaction between the NS1 homodimer and the replication complex through the NS4A and NS4B proteins [38]. The β-roll domain is also involved in NS1 dimerization [38]. Site 22 localizes in spatial proximity to the first β-strand of the β-ladder domain, where the G185 positively selected site also maps. Both sites are located at the dimerization interface [38, 39].

**Table 1. Branch-site analyses of the flavivirus polyprotein (nonstructural region).**

	MA vs MA1	p value (Bonferroni corrected p value)	BUSTED	p value (Bonferroni corrected p value)
Foreground Branch <sup>a</sup>	-2ΔlnL <sup>b</sup>		-2ΔlnL <sup>b</sup>	
Branch A	22.00	2.73x10 <sup>-6</sup> (8.19x10 <sup>-6</sup> )	0.32	1 (1)
Branch B	35.22	2.94x10 <sup>-9</sup> (8.83x10 <sup>-9</sup> )	4.36	1 (1)
Branch C	34.24	4.87x10 <sup>-9</sup> (1.46x10 <sup>-8</sup> )	8.10	0.016 (0.048)

<sup>a</sup> Branches are named as in [Fig 1](#)

<sup>b</sup> Twice the difference of likelihood for the null and the alternative models compared.



**Fig 2. 3D mapping of NS1 positively selected sites.** 3D structure of ZIKV NS1 dimer obtained by homology modeling of WNV NS1 (PDB ID: 4O6C).

Domain colors are as in Fig 1C. Positively selected sites are shown in red. Protein alignments of the regions surrounding selected sites are also shown for a few representative flavivirus species (JEV, NC\_001437; WNV, NC\_001563). Amino acids corresponding to codons filtered by GUIDANCE are represented by shaded X letters.

doi:10.1371/journal.pntd.0004978.g002

As for selected sites in proteins other than NS1, site V15 in NS2A maps to a hydrophobic protein region within the lumen of the endoplasmic reticulum (ER); mutations at nearby residues impair DENV virion assembly [40]. Positively selected sites were also detected in NS3 (M41, P82, T570, P577). Interestingly, site 570 flanks a conserved asparagine that is essential for NS3-NS5 binding in DENV [41]. Finally, the positively selected site in NS4B (I162) is located in a cytoplasmic loop involved in the interaction with NS3 and with host proteins [42].

## Ongoing positive selection at nonstructural proteins in ZIKV

We next investigated whether positive selection also occurred during the recent evolution of ZIKV. To this purpose, we retrieved coding sequence information for all complete ZIKV genomes ( $n = 39$ , as of March 26<sup>th</sup>, 2016) ([S1 Table](#)). Again, the structural and nonstructural regions were analyzed separately. In the structural region, GARD detected no recombination, whereas Recco inferred possible breakpoints at nucleotides 802–838 (relative to AY632535 coding sequence) in the M region. Both Recco and GARD detected a recombination breakpoint in the nonstructural region (position 8994, GARD; position 9040–9054, Recco) within the NS5 region. The portions encompassing the breakpoint positions were thus removed and the alignments were split into two sub-regions. Inspection of the Recco output indicated that in all cases recombination involved sequences from the African isolates. In fact, the phylogenetic trees for all sub-regions showed a clear separation of the African and non-African sequences ([S1 Fig](#)).

To obtain an estimate of the degree of constraint along ZIKV genomes, we used FUBAR to identify sites showing significant evidence of negative selection. This analysis indicated an uneven distribution of negatively selected sites, with the weakest selective pressure acting on the structural portion; conversely, more than 80% of sites are negatively selected in the NS1-NS4B region ([Fig 1A](#)).

We next tested for positive selection using both the codeml site models (M7 vs M8) and the branch-site models (MA1 vs MA). These latter models were used to test for selection on the branch of the phylogeny that separates the African and non-African sequences. No evidence of positive selection was obtained for the two sub-regions from the structural portion. Conversely, for the nonstructural region covering nucleotides 2371–8994, a model of evolution that allows a class of codons to evolve with  $\omega > 1$  (NSsite model M8) better fitted the data than the neutral model (NSsite models M7), supporting the action of positive selection ( $-2\Delta\ln L = 18.89$ , degrees of freedom = 2, Bonferroni- corrected  $p$  value for two tests =  $1.58 \times 10^{-4}$ ). Positively selected sites were identified using the BEB procedure from M8 and with two additional methods from the HyPhy suite, REL and FUBAR. Sites were defined as being positively selected if they were detected by at least two different methods. Using this conservative criterion, 5 positively selected sites were detected ([Table 2](#) and [Figs 1A](#) and [3](#)). Three of them are located in the relatively short NS4B region (M26, M87, and H88); using the same approach as above, we determined that this number is unlikely to occur by chance (random sampling,  $p$  value = 0.007), indicating that NS4B is the preferential positive selection target in ZIKV.

To gain insight into the location of positively selected sites in NS4B, we performed an *in silico* prediction of transmembrane helices. The resulting topology model was very similar to those previously proposed or determined for other flaviviruses [42] ([Fig 3B](#)). Residue M26 maps to the N-terminal region located in the ER lumen. Interestingly, the corresponding position was previously found to be positively selected in sylvatic DENV2 isolates; in JEV, a nearby

**Table 2. Positively selected sites in ZIKV polyprotein.**

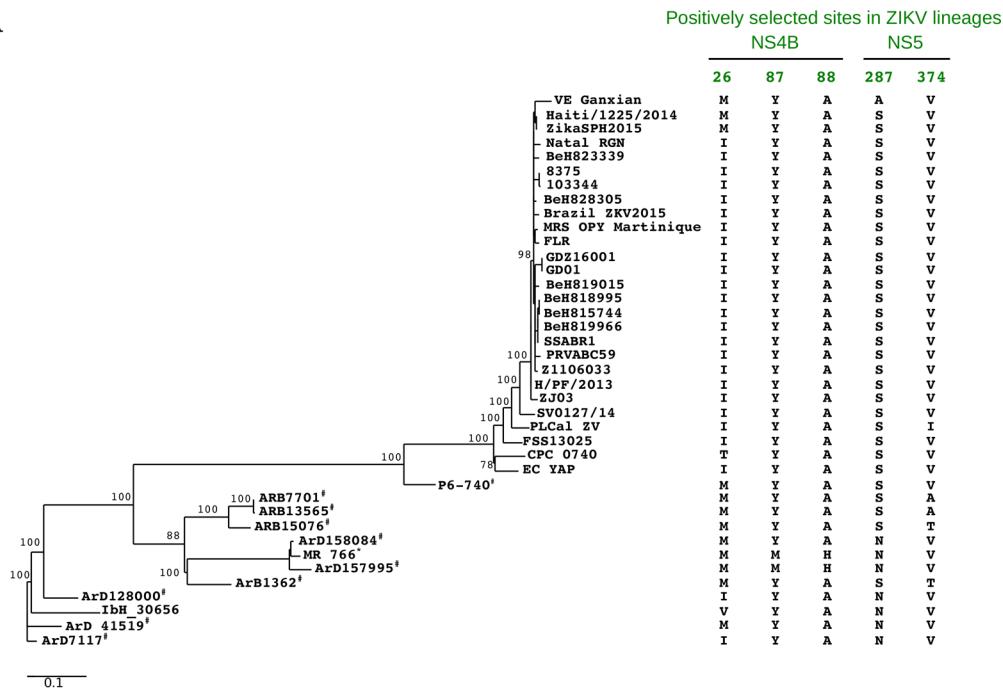
Position <sup>a</sup>	Protein	BEB pp <sup>b</sup>	REL Bayes Factor	FUBAR pp <sup>b</sup>
26	NS4B	0.92	1117.62	0.76
87	NS4B	0.99	2049.42	0.34
88	NS4B	0.94	187.395	0.20
287	NS5	0.89	526.902	0.91
374	NS5	0.96	129.954	0.23

<sup>a</sup> Relative position in protein

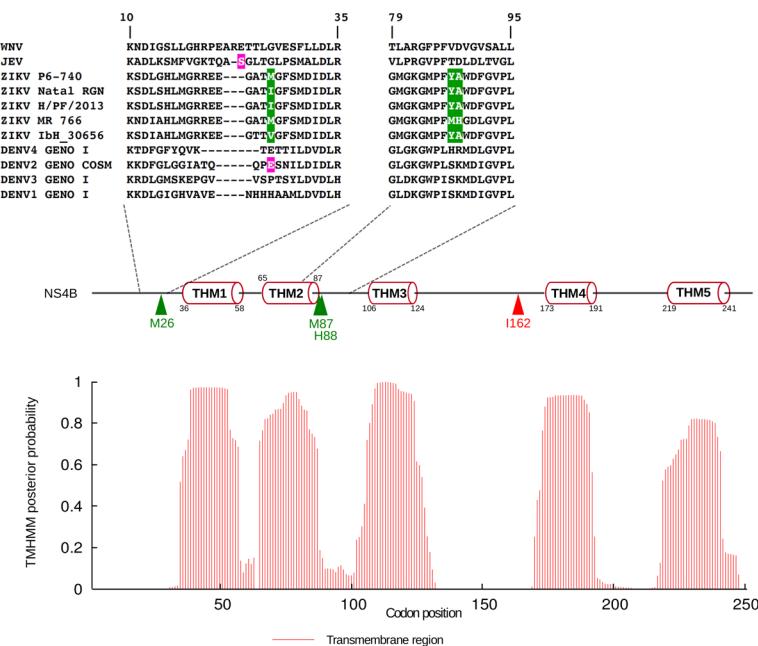
<sup>b</sup>Posterior Probability

doi:10.1371/journal.pntd.0004978.t002

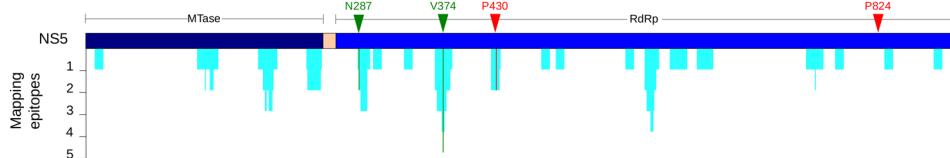
A



B



C



**Fig 3. Ongoing positive selection in ZIKV isolates.** (A) Maximum likelihood phylogeny for the nonstructural region (nucleotides 2371 to 8994). Amino acids at the five selected sites are shown for all ZIKV sequences.

Viruses isolated from mosquitos and macaques are denoted with hash and asterisk symbols, respectively. Branch length is proportional to nucleotide substitutions per codon. Bootstrap values for internal branches >75% are shown. The phylogenetic tree is unrooted. (B) TMHMM prediction of transmembrane helices (TMH1-5) for the ZIKV NS4B protein and schematic representation of protein topology. Positively selected sites in the flavivirus phylogeny and in ZIKV strains are indicated by red and green triangles, respectively. Amino acid alignments of the regions surrounding selected sites are shown for 5 representative ZIKV, for DENV sequences belonging to the four serotypes, for JEV (NC\_001437), and for WNV (strain NY-99, NC\_001563). In the alignment, positively selected sites in ZIKV are shown in green; sites that are positively selected in other flaviviruses are marked in magenta. (C) Immune epitope mapping on NS5. Cyan bars indicate the number of epitopes overlapping each NS5 residue. Positively selected sites are colored as above. Mtase: methyltransferase domain; RdRp: RNA-dependent RNA polymerase domain.

doi:10.1371/journal.pntd.0004978.g003

residue is positively selected, as well [43, 44] (Fig 3B). Residues 87 and 88 are also located in the ER lumen and reside in the second loop (Fig 3B), a region involved in NS4B-NS1 interaction in WNV [45].

Two other positively selected sites are located in NS5 (N287 and V374). Position 374 is part of the nuclear localization signal (NLS) region of NS5. Dengue virus serotypes have different nuclear localization, and these differences are due to changes in their NLS [46]. Analysis of DENV immune epitopes indicate that some of them comprise positions 374 and 287 (Fig 3C).

## Discussion

Herein we provide an analysis of the selective forces acting on ZIKV and related flaviviruses. We show that positive selection contributed to the genetic diversity of these human pathogens and we report ongoing adaptive evolution in ZIKV strains.

The evolutionary analysis of viral genomes poses challenges related to the possible presence of recombination, as well as to the high sequence divergence, with consequent saturation issues and alignment uncertainties. We accounted for all these possible confounding effects, which would otherwise affect inference of positive selection. Indeed, we adopted recommended alignment and filtering criteria to minimize erroneous codon alignments [15], and we tested for substitution saturation. As for recombination, we applied two methods, based on different features of the data, to screen the alignments and to infer the most likely position of breakpoints. These latter were used to split alignments into sub-regions that were separately analyzed. In this respect, it is worth noting that we did not detect recombination breakpoints in the non-structural region for the extended flavivirus phylogeny, whereas we found evidence of recombination when all ZIKV strains alone were analyzed. The explanation for this apparently contradictory finding is that one single ZIKV African sequence contributed to the recombination events and it was not represented in the flavivirus phylogeny. Moreover, the flavivirus alignment was partially masked to remove unreliably aligned codons. This procedure clearly determines the removal of the most divergent regions, which may derive from recombination events. This most likely accounts for the discrepancy between our results and those from a previous report that indicated recombination between Asian ZIKV strains and SPOV within NS2B [47]. Another previous study analyzed African and non-African ZIKV isolates and reported the presence of four recombination breakpoints in ZIKV genomes [48]. In our analysis we only detected two breakpoints. We believe that the main reason for the discrepancy with this previous analysis derives from the fact that the ArD142623 strain, which contributed most recombination events in Faye's dataset [48], was not included in our study because its genome sequence is not complete and because its polyprotein sequence is annotated as "nonfunctional due to mutation" in GenBank. In this respect, it is worth mentioning that despite our findings and those previously reported by others for ZIKV [47, 48] and DENV [34], experimental data have indicated that flaviviruses have very low propensity for recombination [49, 50]. Moreover,

under laboratory conditions specifically devised to detect recombination, extremely rare events were observed that generated aberrant JEV genomes with reduced growth properties [50]. These observations raise the possibility that recombination events identified through analysis of existing sequences in public databases are artifacts of laboratory contamination or incorrectly assembled sequence files. This was previously suggested to be the case for some “recombinant” DENV sequences [34, 51]. All the recombination events we detected involved one or few sequences from African ZIKV isolates. Whereas we cannot control for the accuracy of the deposited sequences, we have to take the possible recombination events into account; failure to do so would affect positive selection inference, irrespective of whether recombination actually occurred. Clarification of these potential caveats, though, is extremely relevant for epidemiological and preventive purposes. Because ZIKV, DENV, and other arboviruses can co-circulate during outbreaks [10], it will be extremely important to assess if and with what efficiency these viruses can recombine.

The branch-site tests we applied to analyze the flavivirus phylogeny were aimed at detecting episodic positive selection—i.e. selective events on one branch of the phylogeny and thus likely to have occurred during or after speciation. Using this approach we were able to show that positive selection acted on the branch that separates DENV from the other analyzed flaviviruses and mainly targeted NS1. It should be noted that the branch-site tests have low false positive rates and are largely insensitive to violations of the assumption of neutral evolution on the background branches [22, 24], but lack power [52]. Thus, selection may act on additional branches than the one we detected and more selected sites may exist. When analysis was performed on ZIKV genomes, which are characterized by much lower divergence compared to the flavivirus sequences in the inter-species analysis, tests to detect episodic and pervasive positive selection were applied. The branch-site test showed no evidence of episodic selection and we consequently identified no selective events leading to the spread of the Asian ZIKV lineage. We mention, however, that the branch-site test may have failed to detect weak selection or selection at a very limited number of codons. We also stress that the lack of selection signatures does not imply that amino acid differences between African and non-African ZIKV lineages are irrelevant or nonfunctional.

Conversely, we identified pervasive selection—i.e. selective events that involve all ZIKV lineages— and, again, selected sites were found to occur in nonstructural portions of the ZIKV genome (NS4B and NS5). These portions also display the strongest level of selective constraint.

Structural proteins (the E protein in particular) might be a priori considered to be preferential selection targets during flavivirus evolution for at least two reasons: these proteins [1] mediate the initial and essential steps of host infection via host cell binding and entry and [2] represent major targets for immune responses influencing antigenic selection [53]. Nevertheless, we found no evidence of positive selection in structural regions, either in flaviviruses or in ZIKV isolates. To our knowledge, no study has investigated the occurrence of positive selection in ZIKV or during flavivirus speciation, but efforts at detecting positive selection in DENV strains or isolates were performed. Depending on the serotype analyzed, on the geographic and temporal origin of the viruses, as well as on their transmission cycle (sylvatic or endemic), different genomic regions were found to represent targets of positive selection in DENV [54–59]. These regions were not limited to the structural portion, but also included nonstructural proteins [54–58]. Likewise, analysis of JEV sequences revealed selection in both structural and non-structural regions [43, 44]. Also, ample evidence indicated that although most neutralizing antibodies are directed against flavivirus E proteins [53], non-neutralizing anti-NS1 antibodies are protective against severe disease during DENV, YFV, and JEV infection [60–64].

Finally, cell-mediated immunity was shown to target both structural and nonstructural DENV proteins, with the vast majority of T-cell epitopes located in nonstructural proteins

[65]. In this respect, it is worth noting that several of the positively selected sites we detected in NS1 and in NS5 are located within immune epitopes. This observation suggests that the underlying selective pressure responsible for selection at these sites is exerted by the host adaptive immune system. These data are likely to be relevant for the current efforts to develop a ZIKV vaccine and to assess the possible cross-protection afforded by natural or vaccine-induced immunity against other related viruses.

Nonstructural proteins play different roles in flavivirus life cycles and several of them interact with innate immunity molecules. NS1, the major selection target in the flaviviruses we analyzed herein, is essential for viral RNA replication and is involved in immune system evasion. In particular, secreted hexameric NS1 represents a major antigenic marker of viral infection for all DENV serotypes. Soluble NS1 in the serum of patients has been found to correlate with severe clinical disease [66], suggesting that the NS1 protein also plays an important role in the pathogenesis of dengue. Importantly, the NS1 protein from WNV and DENV2 interacts with multiple components of the complement system (C1S, C4, C4-binding protein, CFH), as well as with toll-like receptors (TLR3, TLR2, and TLR6) [67, 68]. The molecular details of these interactions are presently unknown, but the presence of several positively selected sites in NS1 suggests a possible arms race with the host innate immune system. It will be extremely important to assess whether amino acid differences in flavivirus NS1 proteins affect the interaction with innate immunity components and consequently modulate the host response to ZIKV or DENV.

NS1 is also required for efficient viral genome replication. Recently, it has been proposed that dimeric NS1 plays an organizational role in the formation of the replication complex on the cytoplasmic side of the ER membrane [69], and that this function is mediated by interactions with NS4A and NS4B [45, 70]. Interestingly, the positively selected site in the connector domain (residue 164) is located in the hydrophobic protrusion that may contact NS4A and NS4B [38]. Although mutation of residue T164 to alanine has no effect on RNA replication or on the assembly of DENV particles [37], we cannot exclude its involvement in (de)stabilizing the interaction with the ER membrane; in fact, the introduction of a histidine (Fig 2) at this site might affect the protein function more importantly than the conservative alanine substitution. Scaturro and colleagues [37] also reported that NS1 plays a critical role in the biogenesis of DENV virions, a function that is mediated by interaction with structural proteins. In this context, a key role is played by two residues (114 and 115) in the flexible loop of the NS1 wing domain. Indeed, alanine mutation of residue S114 abrogates DENV2 NS1 binding to E, prM, and C [37]. Notably, we identified residue 114 as positively selected in the flavivirus phylogeny and two additional selected sites were located in close spatial proximity.

Overall, these observations suggest that positive selection at NS1 is acting to optimize viral fitness by modulating viral replication efficiency and/or favoring evasion from the host immune system. Similar considerations may apply to NS4B, which displays 3 positively selected sites in ZIKV isolates. This membrane protein has a role in the formation of the replication complex and in virus pathogenesis [71]. Several mutations in the NS4B region of JEV, YFV, and WNV were shown to modulate neurovirulence and/or neuroinvasiveness [42]. Notably, one of the positively selected sites we identified in NS4B (26M/I/T/V) was previously reported to represent a selection target in sylvatic DENV2 isolates but not in the endemic strains [57]. A nearby residue was also found to evolve adaptively in JEV, both in genotype I and genotype III isolates [43, 44]. However, variation at this site seems not to be associated with host preferences in JEV [44]. Although the functional significance of changes at position 26 in NS4B remains to be clarified, the fact that this residue or a flanking one is targeted by positive selection in three closely related flaviviruses suggests an important role in viral adaptation. Indeed, the N-terminal region of flavivirus NS4B (amino acids 1–125) inhibits interferon

(IFN) response by blocking IFN- $\alpha/\beta$  signaling [72]. This region includes two additional positively selected sites (M87 and H88) and is involved in host protein (e.g. STING) binding. It is interesting to note that YFV NS4B, but not DENV NS4B, can bind STING [73], suggesting that positive selection in this region results from adaptation to the host innate immune system to modulate binding of viral sensors. The region surrounding positions 87 and 88 is also responsible for NS1-NS4B binding and the same study demonstrated the importance of the F86C mutation in WNV NS4B to rescue viral replication in presence of NS1 nonfunctional mutations [45]. Finally, the DENV NS4B region spanning residues 84 to 146 is required for interaction with NS4A, another molecule involved in flavivirus replication [74].

Thus, based on data from other flaviviruses, the three positively selected sites we identified in NS4B of ZIKV are located in a protein region important for interaction with other viral proteins and with host molecules.

The relatively sparse sampling of ZIKV genomes and the paucity of ZIKV sequences isolated from humans in Africa and from mosquitoes in Asia/America, prevents drawing any definite conclusion about the role of selected sites on host preference, pathogenicity, or infectivity. Moreover, as anticipated above, a potential issue associated with viral sequence analysis concerns laboratory contaminations, especially during serial passages in culture. Contaminations were previously suggested to account for discrepancies in DENV phylogenies [34], and a few of the African ZIKV strains we included in the study were passaged several times in suckling mice or cell culture [9]. This process may also introduce variants that are not present in nature, potentially affecting evolutionary inference. These issues are unlikely to affect the analyses we performed on the extended flavivirus phylogeny, as variation on terminal branches has minor effects. However, these caveats should be kept in mind in the analysis of ZIKV strains, especially for positively selected sites showing variability in a minority of sequences. Further evolutionary analysis of ZIKV will greatly benefit from the sequencing and inclusion of additional isolates, not only from the ongoing American epidemic, but also from African countries.

Despite these limitations, we suggest that the positively selected sites we identified should be prioritized in future experimental studies. These amino acids changes are expected to modulate aspects of viral fitness, either in mosquitoes or vertebrate hosts. In this respect, reverse genetic approaches will be instrumental to assess the role of specific changes on different viral phenotypes including transmission by distinct Aedes mosquito vectors or alternative (e.g. human-to-human) transmission modes, increased viremia in humans, and altered tissue tropism.

Finally, we note that NS1 and NS4B are regarded as attractive candidates as direct or indirect targets for antiviral drugs in flavivirus infections [42, 75]. Nonetheless, these proteins are fast evolving in ZIKV and related flaviviruses, and the numerous selected sites are expected to entail functional differences among closely related viruses or even among viruses belonging to the same species. Thus, our data suggest that compounds developed against DENV NS4B [42] or drugs that result in DENV NS1 misfolding [75, 76] may not be active against ZIKV.

## Supporting Information

**S1 Fig. Effect of recombination on tree topology.** Maximum likelihood phylogeny for two subregions of ZIKV non-structural region based on the recombination breakpoint. Branch length is proportional to nucleotide substitutions per codon. The phylogenetic trees are unrooted. Bootstrap values for internal branches >75% are shown.

(PDF)

**S1 Table. List of viral sequences.**

(PDF)

**S2 Table. Substitution saturation analysis of flavivirus polyprotein.**  
(PDF)

**S3 Table. Percentage of GUIDANCE-masked codons in each protein region.**  
(PDF)

## Author Contributions

**Conceived and designed the experiments:** RC MS.

**Performed the experiments:** RC MS DF.

**Analyzed the data:** RC MS DF MC.

**Wrote the paper:** RC MS DF.

## References

1. Petersen LR, Jamieson DJ, Powers AM, Honein MA. Zika Virus. *N Engl J Med.* 2016 Mar 30.
2. Faria NR, Azevedo RD, Kraemer MU, Souza R, Cunha MS, Hill SC, et al. Zika virus in the Americas: Early epidemiological and genetic findings. *Science.* 2016 Mar 24.
3. Weaver SC, Costa F, Garcia-Blanco MA, Ko AI, Ribeiro GS, Saade G, et al. Zika Virus: History, Emergence, Biology, and Prospects for Control. *Antiviral Res.* 2016 Mar 17.
4. Mlakar J, Korva M, Tul N, Popovic M, Poljsak-Prijatelj M, Mraz J, et al. Zika Virus Associated with Microcephaly. *N Engl J Med.* 2016 Mar 10; 374(10):951–8. doi: [10.1056/NEJMoa1600651](https://doi.org/10.1056/NEJMoa1600651) PMID: [26862926](https://pubmed.ncbi.nlm.nih.gov/26862926/)
5. Brasil P, Pereira JP Jr, Raja Gabaglia C, Damasceno L, Wakimoto M, Ribeiro Nogueira RM, et al. Zika Virus Infection in Pregnant Women in Rio de Janeiro—Preliminary Report. *N Engl J Med.* 2016 Mar 4.
6. Calvet G, Aguiar RS, Melo AS, Sampaio SA, de Filippis I, Fabri A, et al. Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. *Lancet Infect Dis.* 2016 Feb 17.
7. Cao-Lormeau VM, Blake A, Mons S, Lastere S, Roche C, Vanhomwegen J, et al. Guillain-Barre Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study. *Lancet.* 2016 Feb 29.
8. Lindenbach BD, Rice CM. Molecular biology of flaviviruses. *Adv Virus Res.* 2003; 59:23–61. PMID: [14696326](https://pubmed.ncbi.nlm.nih.gov/14696326/)
9. Haddow AD, Schuh AJ, Yasuda CY, Kasper MR, Heang V, Huy R, et al. Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage. *PLoS Negl Trop Dis.* 2012; 6(2):e1477. doi: [10.1371/journal.pntd.0001477](https://doi.org/10.1371/journal.pntd.0001477) PMID: [22389730](https://pubmed.ncbi.nlm.nih.gov/22389730/)
10. Weaver SC, Costa F, Garcia-Blanco MA, Ko AI, Ribeiro GS, Saade G, et al. Zika Virus: History, Emergence, Biology, and Prospects for Control. *Antiviral Res.* 2016 Mar 17.
11. Brault AC, Huang CY, Langevin SA, Kinney RM, Bowen RA, Ramey WN, et al. A single positively selected West Nile viral mutation confers increased virogenesis in American crows. *Nat Genet.* 2007 Sep; 39(9):1162–6. PMID: [17694056](https://pubmed.ncbi.nlm.nih.gov/17694056/)
12. Campbell CL, Smith DR, Sanchez-Vargas I, Zhang B, Shi PY, Ebel GD. A positively selected mutation in the WNV 2K peptide confers resistance to superinfection exclusion in vivo. *Virology.* 2014 Sep; 464–465:228–32. doi: [10.1016/j.virol.2014.07.009](https://doi.org/10.1016/j.virol.2014.07.009) PMID: [25104615](https://pubmed.ncbi.nlm.nih.gov/25104615/)
13. Loytynoja A, Goldman N. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A.* 2005 Jul 26; 102(30):10557–62. PMID: [16000407](https://pubmed.ncbi.nlm.nih.gov/16000407/)
14. Sela I, Ashkenazy H, Katoh K, Pupko T. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res.* 2015 Jul 1; 43(W1): W7–14. doi: [10.1093/nar/gkv318](https://doi.org/10.1093/nar/gkv318) PMID: [25883146](https://pubmed.ncbi.nlm.nih.gov/25883146/)
15. Privman E, Penn O, Pupko T. Improving the performance of positive selection inference by filtering unreliable alignment regions. *Mol Biol Evol.* 2012 Jan; 29(1):1–5. doi: [10.1093/molbev/msr177](https://doi.org/10.1093/molbev/msr177) PMID: [21772063](https://pubmed.ncbi.nlm.nih.gov/21772063/)
16. Xia X, Xie Z, Salemi M, Chen L, Wang Y. An index of substitution saturation and its application. *Mol Phylogenetic Evol.* 2003 Jan; 26(1):1–7. PMID: [12470932](https://pubmed.ncbi.nlm.nih.gov/12470932/)
17. Xia X. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Mol Biol Evol.* 2013 Jul; 30(7):1720–8. doi: [10.1093/molbev/mst064](https://doi.org/10.1093/molbev/mst064) PMID: [23564938](https://pubmed.ncbi.nlm.nih.gov/23564938/)

18. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol.* 2006 Oct; 23(10):1891–901. PMID: [16818476](#)
19. Maydt J, Lengauer T. Recco: recombination analysis using cost optimization. *Bioinformatics.* 2006 May 1; 22(9):1064–71. PMID: [16488909](#)
20. Theze J, Lowes S, Parker J, Pybus OG. Evolutionary and Phylogenetic Analysis of the Hepacivirus and Pegasiviruses. *Genome Biol Evol.* 2015 Oct 21; 7(11):2996–3008. doi: [10.1093/gbe/evv202](#) PMID: [26494702](#)
21. Guindon S, Delsuc F, Dufayard JF, Gascuel O. Estimating maximum likelihood phylogenies with PhyML. *Methods Mol Biol.* 2009; 537:113–37. doi: [10.1007/978-1-59745-251-9\\_6](#) PMID: [19378142](#)
22. Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol.* 2005 Dec; 22(12):2472–9. PMID: [16107592](#)
23. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007 Aug; 24(8):1586–91. PMID: [17483113](#)
24. Anisimova M, Yang Z. Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. *Mol Biol Evol.* 2007 May; 24(5):1219–28. PMID: [17339634](#)
25. Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, et al. Gene-Wide Identification of Episodic Selection. *Mol Biol Evol.* 2015 Feb 19.
26. Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics.* 2005 Mar 1; 21(5):676–9. PMID: [15509596](#)
27. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* 2005 May; 22(5):1208–22. PMID: [15703242](#)
28. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, et al. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol.* 2013 May; 30(5):1196–205. doi: [10.1093/molbev/mst030](#) PMID: [23420840](#)
29. Pickett BE, Sadat EL, Zhang Y, Noronha JM, Squires RB, Hunt V, et al. ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res.* 2012 Jan; 40(Database issue):D593–8. doi: [10.1093/nar/gkr859](#) PMID: [22006842](#)
30. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 2011 Oct 11; 7:539. doi: [10.1038/msb.2011.75](#) PMID: [21988835](#)
31. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics.* 2006 Jan 15; 22(2):195–201. PMID: [16301204](#)
32. Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, et al. VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.* 2003 Jul 1; 31(13):3316–9. PMID: [12824316](#)
33. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001 Jan 19; 305(3):567–80. PMID: [11152613](#)
34. Chen R, Vasilakis N. Dengue—quo tu et quo vadis? *Viruses.* 2011 Sep; 3(9):1562–608. doi: [10.3390/v3091562](#) PMID: [21994796](#)
35. Anisimova M, Nielsen R, Yang Z. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics.* 2003 Jul; 164(3):1229–36. PMID: [12871927](#)
36. Alkan C, Zapata S, Bichaud L, Moureau G, Lemey P, Firth AE, et al. Ecuador Paraiso Escondido Virus, a New Flavivirus Isolated from New World Sand Flies in Ecuador, Is the First Representative of a Novel Clade in the Genus Flavivirus. *J Virol.* 2015 Dec; 89(23):11773–85. doi: [10.1128/JVI.01543-15](#) PMID: [26355096](#)
37. Scaturro P, Cortese M, Chatel-Chaix L, Fischl W, Bartenschlager R. Dengue Virus Non-structural Protein 1 Modulates Infectious Particle Production via Interaction with the Structural Proteins. *PLoS Pathog.* 2015 Nov 12; 11(11):e1005277. doi: [10.1371/journal.ppat.1005277](#) PMID: [26562291](#)
38. Akey DL, Brown WC, Dutta S, Konwerski J, Jose J, Jurkiw TJ, et al. Flavivirus NS1 structures reveal surfaces for associations with membranes and the immune system. *Science.* 2014 Feb 21; 343(6173):881–5. doi: [10.1126/science.1247749](#) PMID: [24505133](#)
39. Song H, Qi J, Haywood J, Shi Y, Gao GF. Zika virus NS1 structure reveals diversity of electrostatic surfaces among flaviviruses. *Nat Struct Mol Biol.* 2016 May; 23(5):456–8. doi: [10.1038/nsmb.3213](#) PMID: [27088990](#)

40. Xie X, Zou J, Puttikhunt C, Yuan Z, Shi PY. Two distinct sets of NS2A molecules are responsible for dengue virus RNA synthesis and virion assembly. *J Virol.* 2015 Jan 15; 89(2):1298–313. doi: [10.1128/JVI.02882-14](https://doi.org/10.1128/JVI.02882-14) PMID: [25392211](#)
41. Tay MY, Saw WG, Zhao Y, Chan KW, Singh D, Chong Y, et al. The C-terminal 50 amino acid residues of dengue NS3 protein are important for NS3-NS5 interaction and viral replication. *J Biol Chem.* 2015 Jan 23; 290(4):2379–94. doi: [10.1074/jbc.M114.607341](https://doi.org/10.1074/jbc.M114.607341) PMID: [25488659](#)
42. Zmurko J, Neyts J, Dallmeier K. Flaviviral NS4b, chameleon and jack-in-the-box roles in viral replication and pathogenesis, and a molecular target for antiviral intervention. *Rev Med Virol.* 2015 Jul; 25(4):205–23. doi: [10.1002/rmv.1835](https://doi.org/10.1002/rmv.1835) PMID: [25828437](#)
43. Carney J, Daly JM, Nisalak A, Solomon T. Recombination and positive selection identified in complete genome sequences of Japanese encephalitis virus. *Arch Virol.* 2012 Jan; 157(1):75–83. doi: [10.1007/s00705-011-1143-4](https://doi.org/10.1007/s00705-011-1143-4) PMID: [22033595](#)
44. Han N, Adams J, Chen P, Guo ZY, Zhong XF, Fang W, et al. Comparison of genotypes I and III in Japanese encephalitis virus reveals distinct differences in their genetic and host diversity. *J Virol.* 2014 Oct; 88(19):11469–79. doi: [10.1128/JVI.02050-14](https://doi.org/10.1128/JVI.02050-14) PMID: [25056890](#)
45. Youn S, Li T, McCune BT, Edeling MA, Fremont DH, Cristea IM, et al. Evidence for a genetic and physical interaction between nonstructural proteins NS1 and NS4B that modulates replication of West Nile virus. *J Virol.* 2012 Jul; 86(13):7360–71. doi: [10.1128/JVI.00157-12](https://doi.org/10.1128/JVI.00157-12) PMID: [22553322](#)
46. Hannemann H, Sung PY, Chiu HC, Yousuf A, Bird J, Lim SP, et al. Serotype-specific differences in dengue virus non-structural protein 5 nuclear localization. *J Biol Chem.* 2013 Aug 2; 288(31):22621–35. doi: [10.1074/jbc.M113.481382](https://doi.org/10.1074/jbc.M113.481382) PMID: [23770669](#)
47. Zhu Z, Chan JF, Tee KM, Choi GK, Lau SK, Woo PC, et al. Comparative genomic analysis of pre-epidemic and epidemic Zika virus strains for virological factors potentially associated with the rapidly expanding epidemic. *Emerg Microbes Infect.* 2016 Mar 16; 5:e22. doi: [10.1038/emi.2016.48](https://doi.org/10.1038/emi.2016.48) PMID: [26980239](#)
48. Faye O, Freire CC, Iamarino A, Faye O, de Oliveira JV, Diallo M, et al. Molecular evolution of Zika virus during its emergence in the 20(th) century. *PLoS Negl Trop Dis.* 2014 Jan 9; 8(1):e2636. doi: [10.1371/journal.pntd.0002636](https://doi.org/10.1371/journal.pntd.0002636) PMID: [24421913](#)
49. McGee CE, Tsetsarkin KA, Guy B, Lang J, Plante K, Vanlandingham DL, et al. Stability of yellow fever virus under recombinatory pressure as compared with chikungunya virus. *PLoS One.* 2011; 6(8):e23247. doi: [10.1371/journal.pone.0023247](https://doi.org/10.1371/journal.pone.0023247) PMID: [21826243](#)
50. Taucher C, Berger A, Mandl CW. A trans-complementing recombination trap demonstrates a low propensity of flaviviruses for intermolecular recombination. *J Virol.* 2010 Jan; 84(1):599–611. doi: [10.1128/JVI.01063-09](https://doi.org/10.1128/JVI.01063-09) PMID: [19864381](#)
51. Murphy BR, Blaney JE,Jr, Whitehead SS. Arguments for live flavivirus vaccines. *Lancet.* 2004 Aug 7–13; 364(9433):499–500. PMID: [15302187](#)
52. Gharib WH, Robinson-Rechavi M. The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. *Mol Biol Evol.* 2013 Jul; 30(7):1675–86. doi: [10.1093/molbev/mst062](https://doi.org/10.1093/molbev/mst062) PMID: [23558341](#)
53. Roehrig JT. Antigenic structure of flavivirus proteins. *Adv Virus Res.* 2003; 59:141–75. PMID: [14696329](#)
54. Twiddy SS, Woelk CH, Holmes EC. Phylogenetic evidence for adaptive evolution of dengue viruses in nature. *J Gen Virol.* 2002 Jul; 83(Pt 7):1679–89. PMID: [12075087](#)
55. Dash PK, Sharma S, Soni M, Agarwal A, Parida M, Rao PV. Complete genome sequencing and evolutionary analysis of Indian isolates of Dengue virus type 2. *Biochem Biophys Res Commun.* 2013 Jul 5; 436(3):478–85. doi: [10.1016/j.bbrc.2013.05.130](https://doi.org/10.1016/j.bbrc.2013.05.130) PMID: [23756811](#)
56. King CC, Chao DY, Chien LJ, Chang GJ, Lin TH, Wu YC, et al. Comparative analysis of full genomic sequences among different genotypes of dengue virus type 3. *Virol J.* 2008 May 21; 5:63,422X-5-63. doi: [10.1186/1743-422X-5-63](https://doi.org/10.1186/1743-422X-5-63) PMID: [18495043](#)
57. Vasilakis N, Holmes EC, Fokam EB, Faye O, Diallo M, Sall AA, et al. Evolutionary processes among sylvatic dengue type 2 viruses. *J Virol.* 2007 Sep; 81(17):9591–5. PMID: [17553878](#)
58. Bennett SN, Holmes EC, Chirivella M, Rodriguez DM, Beltran M, Vorndam V, et al. Selection-driven evolution of emergent dengue virus. *Mol Biol Evol.* 2003 Oct; 20(10):1650–8. PMID: [12832629](#)
59. Bennett SN, Holmes EC, Chirivella M, Rodriguez DM, Beltran M, Vorndam V, et al. Molecular evolution of dengue 2 virus in Puerto Rico: positive selection in the viral envelope accompanies clade reintroduction. *J Gen Virol.* 2006 Apr; 87(Pt 4):885–93. PMID: [16528038](#)
60. Chung KM, Nybakken GE, Thompson BS, Engle MJ, Marri A, Fremont DH, et al. Antibodies against West Nile Virus nonstructural protein NS1 prevent lethal infection through Fc gamma receptor-dependent and -independent mechanisms. *J Virol.* 2006 Feb; 80(3):1340–51. PMID: [16415011](#)

61. Lee TH, Song BH, Yun SI, Woo HR, Lee YM, Diamond MS, et al. A cross-protective mAb recognizes a novel epitope within the flavivirus NS1 protein. *J Gen Virol.* 2012 Jan; 93(Pt 1):20–6. doi: [10.1099/vir.0.036640-0](https://doi.org/10.1099/vir.0.036640-0) PMID: [21918007](#)
62. Schlesinger JJ, Brandriss MW, Walsh EE. Protection against 17D yellow fever encephalitis in mice by passive transfer of monoclonal antibodies to the nonstructural glycoprotein gp48 and by active immunization with gp48. *J Immunol.* 1985 Oct; 135(4):2805–9. PMID: [4031501](#)
63. Li Y, Counor D, Lu P, Duong V, Yu Y, Deubel V. Protective immunity to Japanese encephalitis virus associated with anti-NS1 antibodies in a mouse model. *Virol J.* 2012 Jul 24; 9:135,422X-9-135. doi: [10.1186/1743-422X-9-135](https://doi.org/10.1186/1743-422X-9-135) PMID: [22828206](#)
64. Schlesinger JJ, Foltzer M, Chapman S. The Fc portion of antibody to yellow fever virus NS1 is a determinant of protection against YF encephalitis in mice. *Virology.* 1993 Jan; 192(1):132–41. PMID: [8517015](#)
65. Mathew A, Townsley E, Ennis FA. Elucidating the role of T cells in protection against and pathogenesis of dengue virus infections. *Future Microbiol.* 2014; 9(3):411–25. doi: [10.2217/fmb.13.171](https://doi.org/10.2217/fmb.13.171) PMID: [24762312](#)
66. Libraty DH, Young PR, Pickering D, Endy TP, Kalayanarooj S, Green S, et al. High circulating levels of the dengue virus nonstructural protein NS1 early in dengue illness correlate with the development of dengue hemorrhagic fever. *J Infect Dis.* 2002 Oct 15; 186(8):1165–8. PMID: [12355369](#)
67. Morrison CR, Scholle F. Abrogation of TLR3 inhibition by discrete amino acid changes in the C-terminal half of the West Nile virus NS1 protein. *Virology.* 2014 May; 456–457:96–107. doi: [10.1016/j.virol.2014.03.017](https://doi.org/10.1016/j.virol.2014.03.017) PMID: [24889229](#)
68. Chen J, Ng MM, Chu JJ. Activation of TLR2 and TLR6 by Dengue NS1 Protein and Its Implications in the Immunopathogenesis of Dengue Virus Infection. *PLoS Pathog.* 2015 Jul 30; 11(7):e1005053. doi: [10.1371/journal.ppat.1005053](https://doi.org/10.1371/journal.ppat.1005053) PMID: [26226614](#)
69. Akey DL, Brown WC, Jose J, Kuhn RJ, Smith JL. Structure-guided insights on the role of NS1 in flavivirus infection. *Bioessays.* 2015 May; 37(5):489–94. doi: [10.1002/bies.201400182](https://doi.org/10.1002/bies.201400182) PMID: [25761098](#)
70. Lindenbach BD, Rice CM. Genetic interaction of flavivirus nonstructural proteins NS1 and NS4A as a determinant of replicase function. *J Virol.* 1999 Jun; 73(6):4611–21. PMID: [10233920](#)
71. Miller S, Sparacio S, Bartenschlager R. Subcellular localization and membrane topology of the Dengue virus type 2 Non-structural protein 4B. *J Biol Chem.* 2006 Mar 31; 281(13):8854–63. PMID: [16436383](#)
72. Munoz-Jordan JL, Laurent-Rolle M, Ashour J, Martinez-Sobrido L, Ashok M, Lipkin WI, et al. Inhibition of alpha/beta interferon signaling by the NS4B protein of flaviviruses. *J Virol.* 2005 Jul; 79(13):8004–13. PMID: [15956546](#)
73. Maringer K, Fernandez-Sesma A. Message in a bottle: lessons learned from antagonism of STING signalling during RNA virus infection. *Cytokine Growth Factor Rev.* 2014 Dec; 25(6):669–79. doi: [10.1016/j.cytogfr.2014.08.004](https://doi.org/10.1016/j.cytogfr.2014.08.004) PMID: [25212897](#)
74. Zou J, Xie X, Wang QY, Dong H, Lee MY, Kang C, et al. Characterization of dengue virus NS4A and NS4B protein interaction. *J Virol.* 2015 Apr; 89(7):3455–70. doi: [10.1128/JVI.03453-14](https://doi.org/10.1128/JVI.03453-14) PMID: [25568208](#)
75. Watterson D, Modhiran N, Young PR. The many faces of the flavivirus NS1 protein offer a multitude of options for inhibitor design. *Antiviral Res.* 2016 Jun; 130:7–18. doi: [10.1016/j.antiviral.2016.02.014](https://doi.org/10.1016/j.antiviral.2016.02.014) PMID: [26944216](#)
76. Rathore AP, Paradkar PN, Watanabe S, Tan KH, Sung C, Connolly JE, et al. Celgosivir treatment misfolds dengue virus NS1 protein, induces cellular pro-survival genes and protects against lethal challenge mouse model. *Antiviral Res.* 2011 Dec; 92(3):453–60. doi: [10.1016/j.antiviral.2011.10.002](https://doi.org/10.1016/j.antiviral.2011.10.002) PMID: [22020302](#)