Infection, Genetics and Evolution 51 (2017) 74–85

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Comparative analysis of protein evolution in the genome of pre-epidemic and epidemic Zika virus

Arunachalam Ramaiah [a,1], Lei Dai [b,c], Deisy Contreras [d], Sanjeev Sinha [e], Ren Sun [b,*], Vaithilingaraja Arumugaswami [d,f,g,**]

[a] Centre for Infectious Disease Research, Indian Institute of Science, Bangalore, KA 560012, India
[b] Department of Molecular and Medical Pharmacology, David Geffen School of Medicine, University of California at Los Angeles, CA 90095, United States
[c] Department of Ecology and Evolutionary Biology, University of California at Los Angeles, CA 90095, United States
[d] Board of Governors Regenerative Medicine Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048, United States
[e] All India Institute of Medical Sciences, New Delhi, India
[f] Department of Surgery, Cedars-Sinai Medical Center, Los Angeles, CA 90048, United States
[g] Department of Surgery, David Geffen School of Medicine, University of California at Los Angeles, CA 90095, United States

ARTICLE INFO

Article history:
Received 18 November 2016
Received in revised form 10 March 2017
Accepted 13 March 2017
Available online 14 March 2017

Keywords:
Zika virus
Host adaptation
Natural selection
Co-evolving sites
Immune epitopes

ABSTRACT

Zika virus (ZIKV) causes microcephaly in congenital infection, neurological disorders, and poor pregnancy outcome and no vaccine is available for use in humans or approved. Although ZIKV was first discovered in 1947, the exact mechanism of virus replication and pathogenesis remains unknown. Recent outbreaks of Zika virus in the Americas clearly suggest a human-mosquito cycle or urban cycle of transmission. Understanding the conserved and adaptive features in the evolution of ZIKV genome will provide a hint on the mechanism of ZIKV adaptation to a new cycle of transmission. Here, we show comprehensive analysis of protein evolution of ZIKV strains including the current 2015–16 outbreak. To identify the constraints on ZIKV evolution, selection pressure at individual codons, immune epitopes and co-evolving sites were analyzed. Phylogenetic trees show that the ZIKV strains of the Asian genotype form distinct cluster and share a common ancestor with African genotype. The TMRCA (Time to the Most Recent Common Ancestor) for the Asian lineage and the subsequently evolved Asian human strains was calculated at 88 and 34 years ago, respectively. The proteome of current 2015/16 epidemic ZIKV strains of Asian genotype was found to be genetically conserved due to genome-wide negative selection, with limited positive selection. We identified a total of 16 amino acid substitutions in the epidemic and pre-epidemic strains from human, mosquito, and monkey hosts. Negatively selected amino acid sites of Envelope protein (E-protein) (positions 69, 166, and 174) and NS5 (292, 345, and 587) were located in central dimerization domains and C-terminal RNA-directed RNA polymerase regions, respectively. The predicted 137 (92 CD4 TCEs; 45 CD8 TCEs) immunogenic peptide chains comprising negatively selected amino acid sites can be considered as suitable target for sub-unit vaccine development, as these sites are less likely to generate immune-escape variants due to strong functional constrains operating on them. The targeted changes at the amino acid level may contribute to better adaptation of ZIKV strains to human-mosquito cycle or urban cycle of transmission.

1. Introduction

Zika virus (ZIKV) is an emerging disease strongly linked to poor pregnancy outcomes and birth defects, specifically microcephaly (Oehler et al., 2014; Mlakar et al., 2016; Brasil et al., 2016; Victora et al., 2016). Zika virus is transmitted to humans primarily through the bite of an infected *Aedes* species mosquito, including both *Ae. aegypti* and *Ae. albopictus* and, sexual contact (Marchette et al., 1969; Foy et al., 2011; Musso et al., 2014). Zika viral particle contains a positive sense, single-stranded RNA genome of about 10.7 kb (Kuno and Chang, 2007). The genome is organized as 5′UTR-C-prM-E-NS1-NS2A-NS2B-NS3-NS4A-2K-NS4B-NS5-3′UTR, with untranslated regions (UTR) flanking a protein-coding region (Nandy et al., 2016). The latter encodes a single polyprotein (3423 amino acids) that is co- and post-translationally cleaved by cellular and viral proteases into multiple

* Corresponding author.
** Corresponding author at: Board of Governors Regenerative Medicine Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048, United States.
E-mail addresses: rsun@mednet.ucla.edu (R. Sun), arumugaswami@cshs.org (V. Arumugaswami).
[1] Present address: Rickettsial Zoonoses Branch, Centers for Disease Control and Prevention, Atlanta, GA 30329, United States.

http://dx.doi.org/10.1016/j.meegid.2017.03.012
1567-1348/

structural and non-structural proteins (Kuno and Chang, 2007; Wang et al., 2016). 5′ and 3′UTR stem loop RNA structures have been shown to be critical for the initiation of viral genome translation and replication in other Flaviviruses (Villordo et al., 2016).

ZIKV was originally isolated from a sentinel rhesus monkey in the Zika forest of Uganda in 1947 and the first human case based on serological evidence was reported in 1952 (Dick et al., 1952). Detailed history of ZIKV outbreaks and phylogeny are described

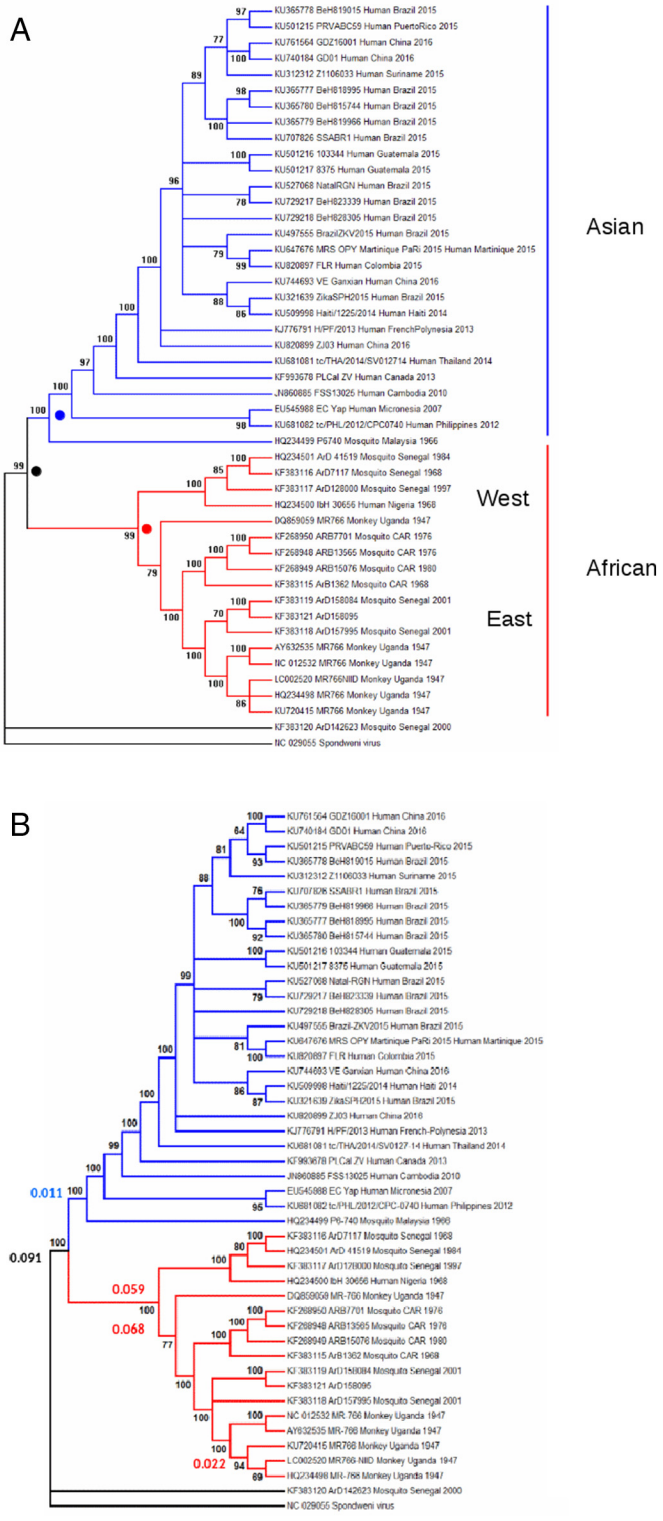


Fig. 1. Phylogenetic relationships of Zika viral strains of African and Asian lineages. (a) ML (b) NJ bootstrap consensus trees were reconstructed with 1000 bootstrap replications. Bootstrap values >60% are shown in the nodes. The African (18 strains) and Asian (28 strains) lineages are colored in red and blue, respectively. The Spondweni virus (NC_029055) was used as an out group. A total of 20 ZIKV strains responsible for current 2015/16 outbreaks were clustered together within the Asian lineage. Whereas, the West African strains were segregated from East African strains in the African clade. The calculated average of genetic distance within the strains from the three clusters and an overall mean are shown near the clades. ML tree was reconstructed under best GTR + G + I substitution model. Both trees show similar topology. (c) Bayesian consensus phylogenetic tree was reconstructed using the MCMC method under GTR + G + I model in MrBayes 3.2.1., which was viewed in FigTree. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

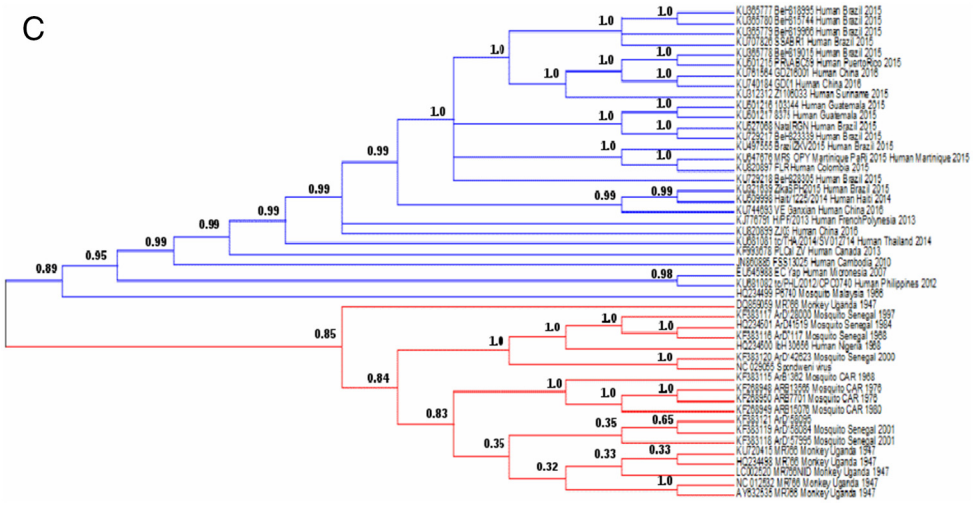


Fig. 1 (*continued*).

(Hayes, 2009; Faye et al., 2014; Lanciotti et al., 2016). Currently, the ZIKV strains are classified into two major lineages or genotypes, namely African and Asian. The African lineage comprises two clusters, namely West African (Nigerian cluster) and East African (MR766 prototype cluster) (Faye et al., 2014; Lanciotti et al., 2016). For the first time in 2007, a ZIKV outbreak was reported in the Pacific region in the Yap Island in the Federated States of Micronesia (Lanciotti et al., 2008). Subsequent epidemics have occurred in French Polynesia in 2013 and the Americas in 2015, including Brazil, Colombia, Guatemala and Puerto Rico (Musso et al., 2014; Baronti et al., 2014; Lazear et al., 2016). Genetic analysis revealed that the Asian genotype of ZIKV is responsible for the Pacific Island outbreaks and the Americas (Haddow et al., 2012; Faye et al., 2014; Baronti et al., 2014; Lanciotti et al., 2016; Enfissi et al., 2016; Zhu et al., 2016).

The high mutation rate in RNA viruses (Mahy, 2010) suggests that highly conserved features of the ZIKV genome may reflect elements critical for viral fitness. In contrast, genetic variations in the ZIKV genome may reveal how viruses have adapted to new environments, such as urban cycle. Sylvatic cycle comprising non-human primates and *Aedes* mosquitos with occasional involvement of human or urban cycle was the original mode of transmission (Dick et al., 1952; Baronti et al., 2014). Subsequent molecular adaptation of ZIKV to naïve population of humans with the circulating *Aedes* mosquito vectors (Weaver, 2017), is suspected to be responsible for increased incidence of microcephaly observed during the current outbreaks. Previous studies have shown patterns of selection pressure, recombination, and glycosylation events in pre-epidemic ZIKV strains sampled from 1947 to 2007 (Faye et al., 2014). Recent studies on the epidemic strains have mainly focused on the phylogeography of ZIKV isolates (Lanciotti et al., 2008; Haddow et al., 2012; Lanciotti et al., 2016). However, detailed analysis on the molecular evolution of epidemic ZIKV strains is still lacking.

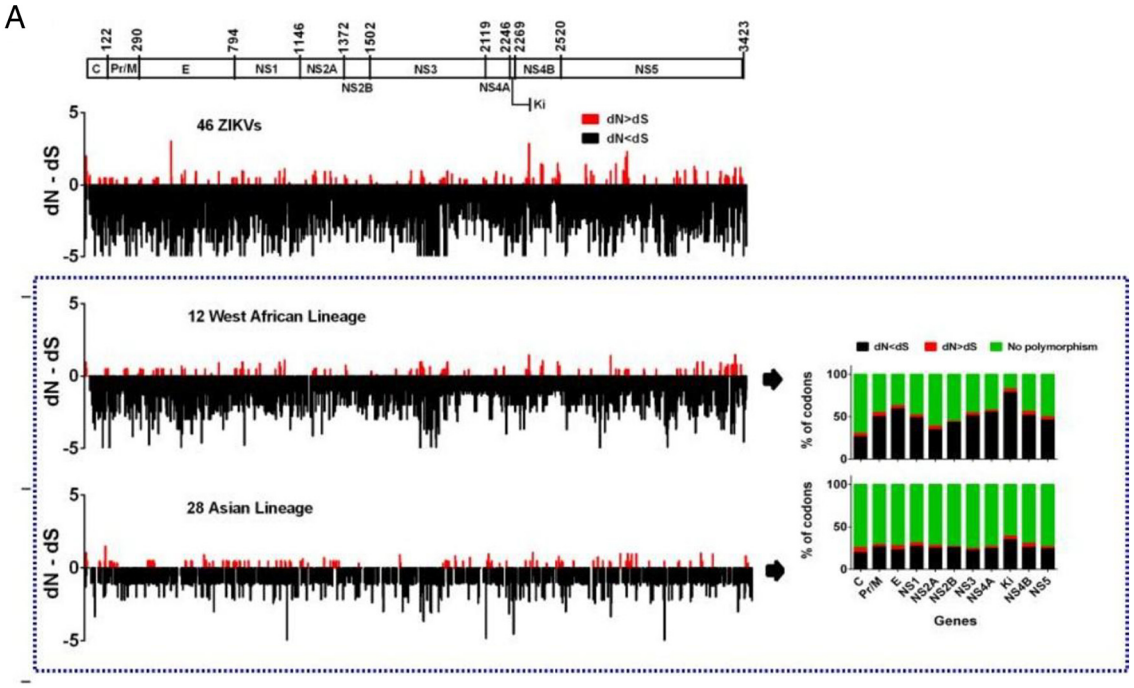
In this study, we performed a comprehensive analysis on the protein evolution of 46 ZIKV strains isolated from 1947 to 2016, including 20 strains responsible for the current 2015–16 outbreak. We defined the strains isolated during the current 2015/16 outbreak as 'epidemic strains' and the remaining as 'pre-epidemic strains'. We demonstrate how evolutionary forces and constraints have shaped the genome of current epidemic ZIKV strains, including genome-wide negative selection on amino acid substitutions, immune epitopes, as well as the co-evolving sites in the proteome in host adaptation and virus replication. We further discuss the implications of our results in the broader context of vaccine design to prevent emergence or re-emergence of pathogenic RNA viruses.

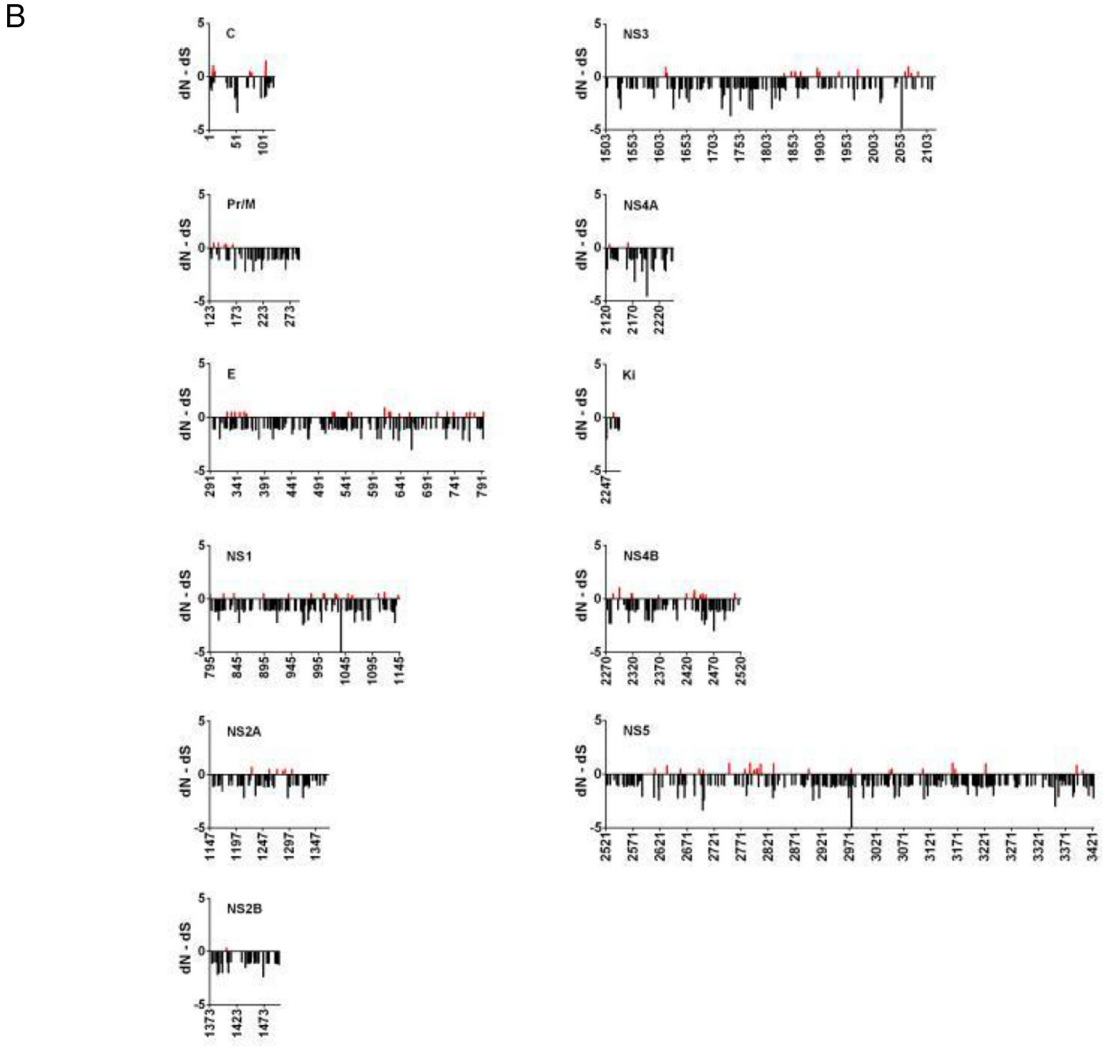
2. Materials and methods

2.1. Phylogenetic analysis of ZIKV strains

A total of 46 existing ZIKV genomic sequences including 20 strains (2015/16 isolates) that are responsible for the current epidemic outbreak were retrieved from GenBank (dated 10-Mar-2016). The evaluated ZIKV strains were isolated from human, monkey, and different species of mosquito from 1947 to 2016. Additionally, genome sequence of single Spondweni virus (NC_029055) from the Flavivirus family was included in our study as an out group species for phylogenetic tree reconstruction. The strain name, accession number, host, year of isolation, country, and strain lineage information of the 46 strains studied are listed in the Supplementary Table 1. Multiple sequence alignments for 47 viral genomes were performed using MUSCLE (Edgar, 2004), subsequently the phylogenetic trees were reconstructed using Maximum Likelihood (ML), Neighbor Joining (NJ), and Bayesian approaches. The NJ tree was reconstructed under Tajima-Nei model (Tamura and Nei, 1993) with 1000 bootstrap supports, and the GTR + G + I was identified as the best model by the Model Test program for reconstructing the ML tree in MEGA6 with 1000 bootstrap supports (Tamura et al., 2013). Bayesian consensus tree was constructed using MCMC method under GTR + G + I model in MrBayes 3.2.1. (Ronquist et al., 2012) and the tree was viewed in FigTree. A total of 40,000 MCMC generations were used to obtain a robust consensus Bayesian tree. The sequence alignment matrices were deposited in a TreeBASE (Sanderson et al., 1994) repository and the accession number was 19,900. The genetic distance (GD) between the strains in each clade and between the clades (East

Fig. 2. Schematic representation of dN-dS test statistics for 3423 codons of Zika viral strains. The dN-dS value for each of the 3423 codons are indicated as a bar, either above or below the X-axis. The positive values indicate an excess of non-synonymous substitutions (red bars) in the codons, whereas, negative values below the horizontal line indicate an excess of synonymous substitutions (black bars). Bars were arbitrarily trimmed at 5 to save space. (A) A total of 46 strains are shown in the graph, including 12 West African and 28 Asian viruses. A schematic of ZIKV genome is aligned on the top of the graph. dN-dS scores are presented in Y-axis. The stacked bar diagrams (right side) show the number of codons with dN < dS (black), dN > dS (red), and with no polymorphisms (green). Selected codons in each gene, for each Zika viral strain, belong to West African cluster and Asian lineage. (B) Autonomous deep view of dN-dS test statistic graphs of codons for 11 genomic segments (C, Pr/M, E, NS1, NS2A, NS2B, NS3, NS4A, Ki, NS4B, NS5) of 28 Asian ZIKV genomes. Codons with excess of dN or dS substitutions are shown, along with codons with no polymorphisms (zero as a score). The X-axis shows the respective number of codons and their range in the genes/segments. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)


A
122
290
794
1146
1372
1502
2119
2246
2269
2520
3423
C
Pr/M
E
NS1
NS2A
NS2B
NS3
NS4A
Ki
NS4B
NS5
46 ZIKVs
dN>dS
dN<dS
dN - dS
12 West African Lineage
28 Asian Lineage
dN<dS
dN>dS
No polymorphism
% of codons
Genes


B
C
Pr/M
E
NS1
NS2A
NS2B
NS3
NS4A
Ki
NS4B
NS5
dN - dS

African cluster, West African cluster, and Asian lineage) were calculated in MEGA6 using the NJ algorithm under Tajima-Nei model (Tamura and Nei, 1993; Tamura et al., 2013) (Supplementary Text 1). The time at which the most recent common ancestor (TMRCA) of ZIKV lived was estimated in DAMBE software package using Tip-Dating method (Xia, 2013).

2.2. Estimating dN-dS score and dN/dS ratio

The ML computations of dN and dS were conducted using HyPhy software package (Pond et al., 2005) implemented in MEGA6 (Tamura et al., 2013). The dN-dS score and dN/dS ratio were estimated for detecting ZIKV codons that have undergone positive or negative selection. Here dS is the number of synonymous substitutions per site (s/S) and dN is the number of non-synonymous substitutions per site (n/N).

2.3. Detection of site-specific selection pressure on epidemic human Zika viral genomes

The selection pressure operating on each codon/amino acid site of epidemic 2015/16 ZIKV strains was detected by computing the dN/dS (ω) ratio. The following multiple statistical methods from Datamonkey (Delport et al., 2010) were used for the selection analyses: SLAC, FEL (Pond and Frost, 2005), IFEL (Pond et al., 2006), FUBAR (Murrell et al., 2013), and MEME (Pond et al., 2011). These methods are described in detail elsewhere (Pond and Frost, 2005; Pond et al., 2006; Pond et al., 2011; Murrell et al., 2013; Arunachalam, 2013; Arunachalam, 2014; Ramaiah and Arumugaswami, 2016).

2.4. Analysis of co-evolving sites

Amino acid sequences of coding region (from site 1 to 3294) are aligned by MUSCLE. The sequences are separated into the African lineage (17 sequences) and the Asian lineage (28 sequences) to preserve phylogenetic homogeneity. The gaps are filled by amino acid of the consensus sequence of each lineage. The alignment is then transformed into a matrix of binary variables. For each site, if the amino acid is the same as the consensus sequence, it is represented by 0; otherwise, it is represented by 1. The analysis of correlation matrix is restricted to polymorphic sites, which we define as positions where at least 2 sequences have non-consensus amino acids.

The effect of phylogeny to the correlation matrix is cleaned up by removing the contributions of the largest two eigenvalues (African lineage) or the largest eigenvalue (Asian lineage) (Dahirel et al., 2011; Quadeer et al., 2014). The cleaned correlation matrix is reconstructed by the remaining eigenvalues. To compare the uncleaned and cleaned correlation matrix (Supplementary Figs. 3,4), the sites are assigned into sectors by the loadings of eigenvectors. The correlation between two sites is considered significant if the following criterions are met: 1) in the cleaned correlation matrix, the magnitude of the correlation is among the top 5%; 2) in the un-cleaned correlation matrix, the magnitude of correlation is larger than 0.5 to ensure that the Z score is larger than 2. The Z score for the correlation between two sites is estimated by 1000 random permutations at each column of the sequence alignment. The analyses are performed by custom MATLAB scripts. The network of sites with significant correlation is visualized by Cytoscape (Shannon et al., 2003).

2.5. Analysis of 3-dimensional (3-D) protein structures of epidemic ZIKV

For protein structure modeling, we utilized an online SWISS-MODEL program. We focused on envelope protein (E-protein) sequences of epidemic ZIKV from human host (KU527068), and pre-epidemic ZIKV from both mosquito (KF268948), and monkey hosts (AY632535) for generating 3D-models. These three protein queries were matched by the Hidden Markov Model system (Biasini et al., 2014) to an envelope protein structure of ZIKV strain H/PF/2013 (KJ776791) (PDB code 5IRE) in the Swiss-Model template library. Subsequently, the models for three target sequences were built based on the target and template (PDB code 5IRE) sequence alignments using Promod-II (Guex and Peitsch, 1997) implemented in a SWISS-MODEL workspace server (Arnold et al., 2006; Biasini et al., 2014). The overall and per-residue model qualities were verified using the QMEAN scoring function (Benkert et al., 2011). The region of four amino acids missing in the monkey E-protein was remodeled using a fragment library. PyMOL was used to create all the figures.

2.6. Predictions of CD4 and CD8 T-cell epitopes

We have considered 6 proteins: C, E (domain), NS2A, NS3-serine protease, NS4A, and NS5 from an epidemic human BeH819966 ZIKV (KU365779), as representative proteins for all 20 genetically conserved human 2015/16 ZIKV strains to predict all possible CD8 and CD4 T-cell epitopes (TCEs) using MHC-I and MHC-II TCE prediction tools (Wang et al., 2008) under default conditions implemented in the Immune Epitope Database (IEDB) (Vita et al., 2014). The predicted peptide epitopes were tested for their immunogenicity using an immunogenicity prediction tool (Calis et al., 2013). Here, the peptide with higher immunogenicity score indicates a greater probability of mounting an immune response. Additionally, immunodominant epitopes carrying negatively selected amino acid sites were shown in the 3-dimensional structures of the E (PDB ID - 5IRE) and NS5 (PDB ID - 5TFR) proteins using PyMOL.

3. Results and discussion

3.1. Phylogenetic trees reveal that 2015/16 epidemic strains belong to the Asian genotype

Identifying the origin and distribution of 2015/16 epidemic ZIKV strains would be crucial for diagnostics, vaccine development, and disease management (Faye et al., 2014). The evolutionary relationship of 46 ZIKV strains isolated from 1947 to 2016 showed one clade or lineage was formed by 28 Asian strains, which included all twenty 2015/16 epidemic ZIKV strains, while the other clade was formed by 18 African strains (Fig. 1), further confirming previous findings (Nandy et al., 2016; Ye et al., 2016; Faria et al., 2016; Zhu et al., 2016). Analysis also indicates that the causative agent of current epidemic is part of the Asian lineage. The evolutionary pressures exerted on Asian and African lineages were different. Interestingly, within the African lineage, strains belonging to East African and West African clusters were clearly segregated. This observation confirms the previous results of Faye et al. (2014) and Lanciotti et al. (2016). All three phylogenetic programs ML, NJ, and Bayesian show similar topology, however the differences in branching order were identified within the clusters (Fig. 1a–b) and/or between African clusters (Fig. 1c). As the West and East African clusters share a common ancestor, which is represented as a node in the phylogenetic tree (Fig. 1a; red dot) and this African ancestor share a common ancestor (black dot) with the Asian ancestor (blue dot). Overall data indicate that the Asian lineage is not closely related to any of the two African clusters, but share a common ancestor with them. The Asian strains have evolved and spread to geographically distinct continents since approximately 1960 (Lanciotti et al., 2016). The polyprotein synthesized by the coding region of 2015/16 ZIKV strains had 99.6–99.9% similarities, which suggests genetic conservation among current viral strains. These circulating strains had close genetic associations with pre-epidemic human H/PF/2013 ZIKV (KJ776791) and differed in only 16 non-synonymous substitutions (M166T, V313I, V346I, T769A, G894A, Y916H, H1857Y, M2074L, I2295M, I2445M, A2611V, M2634V, K2694R, N2778D and R3045C), suggesting that strain H/PF/2013 is likely the ancestor of currently circulating ZIKV strains of Asian genotype. The TMRCA for all ZIKV lineages was computed to have derived ~189 years ago and this common ancestor evolved into two autonomous groups that include the African and Asian lineages. The TMRCA of African and Asian lineages were estimated at 140 and 88 years ago, respectively. The TMRCA of Asian human strains and the current epidemic strains were estimated at 34 and 5 years ago, respectively. Our findings also indicate that the Asian lineage has dispersed to several

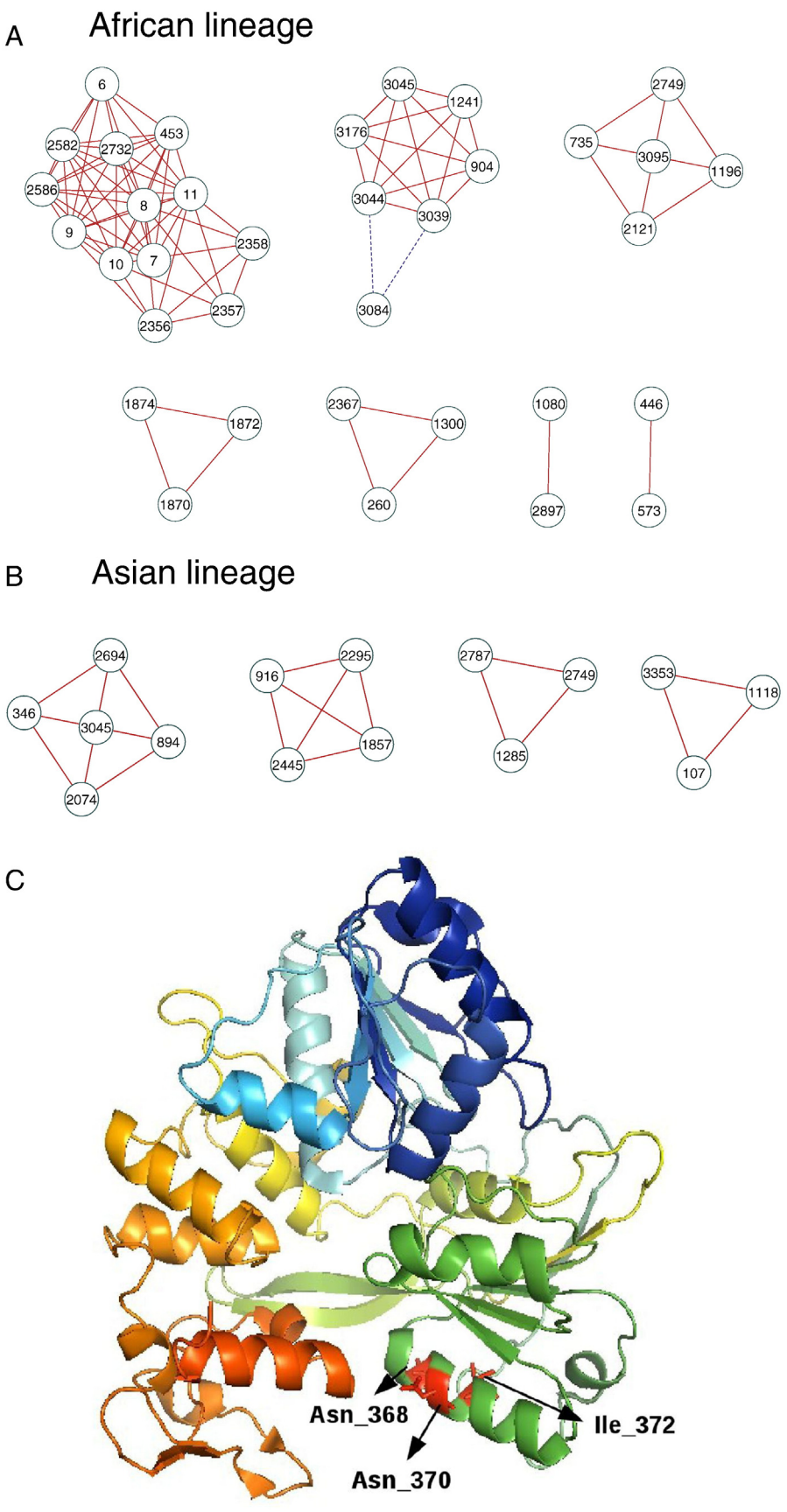


Fig. 3. Network of co-evolving sites of ZIKV polyprotein. (A) African lineage. (B) Asian lineage. The nodes are labeled by site number of Zika polyprotein (from 1 to 3423). The edges indicate significant positive (red solid lines) or negative correlations (blue dashed lines) between two sites. The table describes the position of correlating amino acids from different proteins. (C) Interaction of coevolving sites in ZIKV NS3 helicase (PDB: 5JMT). We have highlighted the three predicted co-evolving sites in the alpha helix domain II of NS3 helicase (368N-370N-372I or in the 1870–1872–1874 regions of the ZIKV polyprotein). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Protein	Range in polyprotein	Amino acid position in African lineage	Amino acid position in Asian lineage
C	1-122	6, 7, 8, 9, 10, 11	107
Pr/M	123-290	260	-
E	291-794	446, 453, 573, 735	346
NS1	795-1146	904	894, 916, 1118
NS2A	1147-1372	1080, 1196, 1241, 1300	1285
NS2B	1373-1502	-	-
NS3	1503-2119	1870, 1872, 1874	1857, 2074
NS4A	2120-2246	2121	-
Ki	2247-2269	-	-
NS4B	2270-2520	2356, 2357, 2358, 2367	2295, 2445
NS5	2521-3423	2582, 2586, 2732, 2749, 2897, 3039, 3044, 3045, 3084, 3095, 3176	2694, 2749, 2787, 3045, 3353

Fig. 3 (*continued*).

countries across the Asian and American continents, whereas the dissemination of the African lineage was restricted to countries within the African continent.

3.2. Proteome of the epidemic ZIKV strains genetically conserved due to genome-wide negative selection

We evaluated genome-wide amino acid substitutions to understand the nature of selection pressure acting on each amino acid site. Higher rate of synonymous (dS) substitutions than non-synonymous (dN) substitutions is usually considered as the signature of negative selection, also known as purifying selection. For the 46 ZIKV genomes isolated from 1947 to 2016, 70% of the 3423 codons were found to be polymorphic, indicating substantial evolution and viral-specific codon usage (van Hemert and Berkhout, 2016) in the viral proteome (Fig. 2a; Supplementary Table 2). Among the three clusters of two lineages, we found that 46–115 codons (1.3–3.4%) were identified with a higher rate of non-synonymous substitutions (dN > dS), while 560–1664 codons (16.4–48.6%) had a higher rate of synonymous substitutions (dN < dS). We identified 10 codons with dN > dS in both the Asian genotype and West African cluster, as well as 426 codons with dN < dS in these clusters (Supplementary Fig. 1; Supplementary Table 3). These results suggest that the adaptive evolutionary strategies of viruses from both Asian and African lineages were independent of each other as they exploited only a small percentage of common codons with excess of dN during the course of evolution. Notably, the codons with excess of dS exploited by two lineages are greater than the codons that are unique to the Asian lineage, further indicating that the fraction of patterns of synonymous changes that occurred in these two lineages are common. However, the Asian strains have evolved with unique dS changes as fitness effects, as they geographically spread in Asia and America continents.

We carried out statistical tests of dN-dS for codons from each of the 11 proteins of the Asian genotype and West African cluster. For example, in the Asian lineage, 31 (26%) of 122 codons in the capsid (C) were polymorphic. Among the polymorphic codons, a higher rate of non-synonymous (dN > dS) and synonymous (dN < dS) substitutions was observed in 7 (6%) and 24 (20%) codons, respectively (Fig. 2, Supplementary Fig. 1a–c). The selection profiles for all the other segments in the Asian lineage were found to be qualitatively similar. We also observed an overall similar pattern of dN and dS in the West African cluster.

The nature of selection pressure exerted on each amino acid of different ZIKV data sets can be measured by obtaining the ratio of dN and dS (Supplementary Table 4a). Our analysis of 46 ZIKV strains (total of 3 clusters from 2 lineages) provided a 'ω' ratio of 0.065, clearly indicating that these strains evolved under strong purifying selection pressures. Overall results of 2 lineages show that 'ω' score ranged from 0.065 to 0.076, suggesting no momentous differences in the overall evolution of ZIKV strains that belonged to different clusters. The mean of proportion of dN changes in all three clusters are comparatively higher than that of dS. The 'ω' score of both Asian genotype and West African cluster were much smaller than one and the highest values were observed in the capsid (0.239) of the Asian lineage and in the pre-membrane (0.118) of the West African cluster (Supplementary Table 4b). This result agrees with the findings of Zhu et al. (2016). In summary, our results suggest that non-synonymous mutations have a lower probability of being fixed than synonymous mutations, as changes in the protein sequence are on average more likely to decrease the replicative fitness of ZIKV. Thus, the proteomes of epidemic ZIKV strains are genetically conserved due to genome-wide negative selection on amino acid substitutions.

The results on positive and negative selection pressures acting on each amino acid site of polyprotein encoded by 2015/16 ZIKV genomes are shown in Supplementary Tables 5 and 6. A single positively selected site (894; IFEL, *p*-value 0.08) was identified by a single method with no statistical significance. Our data showed that a total of 11 of 24 negatively selected amino acid sites were inferred with statistical significance. The present result indicates that the 2015/16 epidemic ZIKV strains had evolved through purifying selection pressures, which are predicted to assist the viruses for better adaptation to the human cycle and to reduce the efficiency of active immunity. The results provided are consistent with the recent findings of other RNA viruses, such as Filovirus (Ramaiah and Arumugaswami, 2016) and influenza virus (Sant'Anna et al., 2014; Arunachalam, 2014), but different from the current prevailing hypothesis that genes encoding antigens can be highly variable to evade host immunity (Farci et al., 2000; Kawashima et al., 2009; Comas et al., 2010; Arunachalam, 2013). In our natural selection analysis, the significance of one method alone is not sufficient to infer that a given amino acid site underwent either positive or negative selection pressure (Pond and Frost, 2005; Pond et al., 2006; Delport et al., 2010). In the present study, no positively selected sites have been detected by more than one method, whereas, out of 11 statistically reliable negatively selected sites from polyprotein, only 2 sites (360 and 3358) were identified by more than one method (FEL: *p*-values 0.05, 0.02; FUBAR: posterior probabilities 0.90, 0.98). Therefore, these 2 amino acid sites are expected to be more reliable. Interestingly, we identified a single amino acid site (894; IFEL, *p*-value 0.08) in the polyprotein, which underwent positive Darwinian selection. Amino acid sites of ZIKV polyprotein under negative selection pressures were comparatively higher. A single negatively selected site (position 3358) identified by IFEL (*p*-value 0.07) was not statistically significant, but the same site identified by FEL (0.02) and FUBAR (posterior probability 0.98) was statistically significant. Moreover, 22 of 24 sites were identified by FEL without statistical significance. Therefore, only amino acid sites that have been detected by more than one method are finally considered as positively

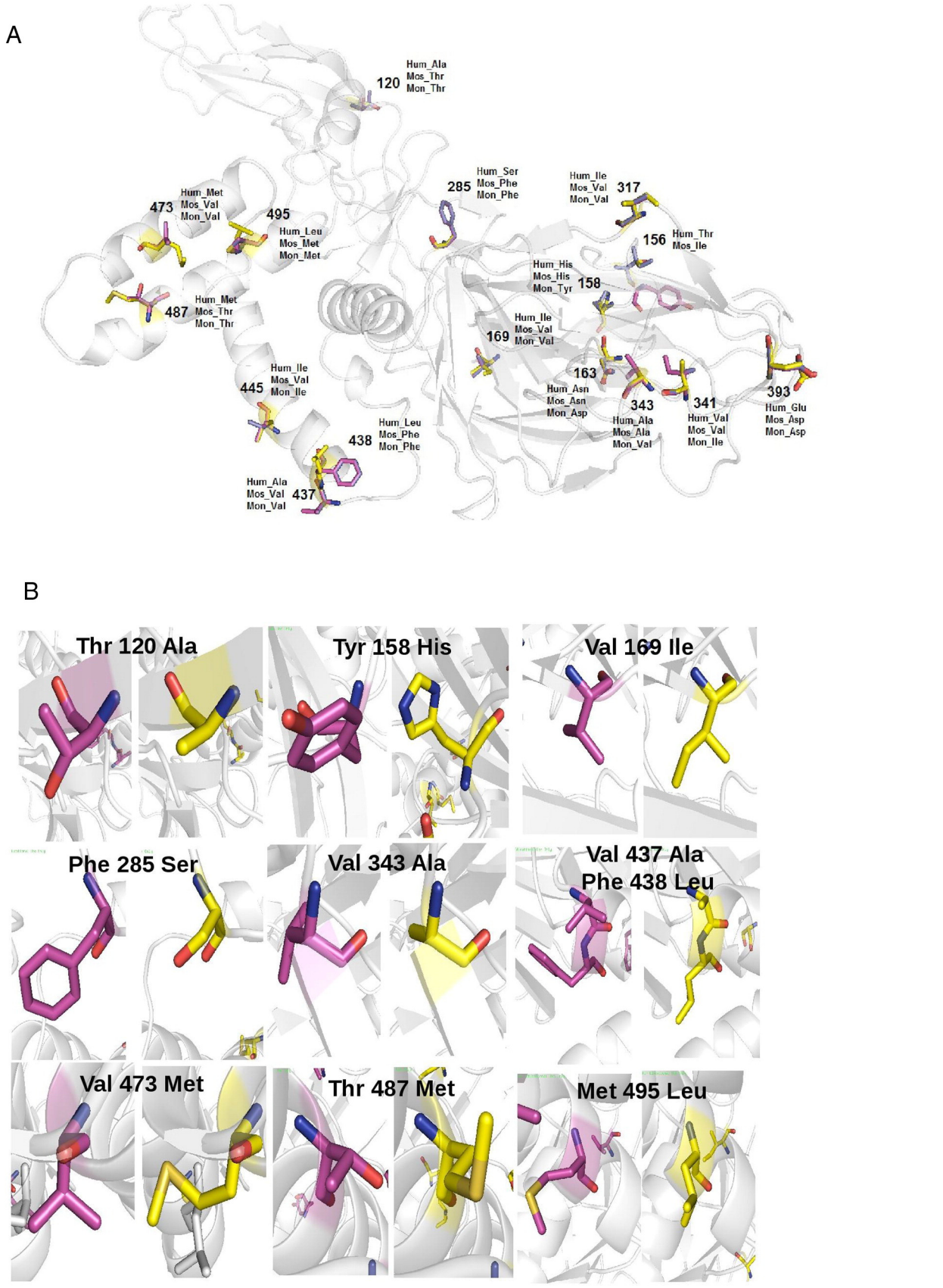


Fig. 4. Comparison of ZIKV E-protein structure from human, mosquito, and monkey hosts. (a) The modeled 3-D structure of 16 amino acid residues that were different between human (yellow), mosquito (blue), and monkey (magenta) E-proteins are shown as sticks in the cartoon. E-protein amino acid substitutions among isolates from three different hosts are shown in the table inset. (b) 3D-structures of the 10 crucial E-protein amino acid substitutions present between monkey (magenta) and human (yellow) isolates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

E-protein domains	Residue Number	Human	Mosquito	Monkey
1-302 (Central and dimerization domains)	120	Ala	Thr	Thr
	156	Thr	Ile	-
	158	His	His	Tyr
	163	Asn	Asn	Asp
	169	Ile	Val	Val
	285	Ser	Phe	Phe
311-403 (Immunoglobulin like domain III)	-317	Ile	Val	Val
	341	Val	Val	Ile
	343	Ala	Ala	Val
	393	Glu	Asp	Asp
408-504 (Stem/anchor domain)	437	Ala	Val	Val
	438	Leu	Phe	Phe
	445	Ile	Val	Ile
	473	Met	Val	Val
	487	Met	Thr	Thr
	495	Leu	Met	Met

Fig. 4 (*continued*).

or negatively selected sites. It should be noted that dN/dS tends to be biased towards one of the samples in the sequence samples that were analyzed. This is because the polymorphism observed is not actually due to fixed substitutions, which are assumed, and generally true in the case of divergent protein sequences. This bias is independent of sample size. Another concern is the small sample size, especially for the "per site" estimation of dN and dS, which is not averaged over the codons in the same protein. This uncertainty in dN and dS values makes it difficult to find sites under "significant" negative or positive selection. However, the actual sites under negative selection can be much more.

3.3. Coevolving amino acid sites in the African and Asian genotypes

To identify higher-order constraints on the evolution of ZIKV proteome, we inferred potentially coevolving amino acid sites in the African and Asian lineages by computing the amino acid variations at each pair of codons (Fig. 3a,b). The effect of phylogeny on site linkage was cleaned by removing the contribution of the largest eigenvalue(s) from the correlation matrix (Supplementary Figs. 3,4). Positive correlation between two sites implies that the corresponding double mutants are observed more frequently than if the mutations were to occur independently. In contrast, negative correlation between two sites implies that the double mutants are observed less frequently than if the individual mutations were to arise independently. In both cases, the dependence of one site on another implies epistasis. For example, the fitness effect of multiple mutations is considerably different from the additive effect of single mutations.

Among the 80 polymorphic sites in the African lineage, seven groups of sites showed significant pairwise correlation. The observed correlation between physical proximal sites, such as codon 1870–1872–1874 (Fig. 3a), may reflect epistatic interactions within a protein, and in this case NS3 (Fig. 3c) (Jain et al., 2016; Tian et al., 2016). Correlation between sites of different proteins may suggest potential protein-protein interactions, which can be further studied via mutagenesis and protein structures.

There are 32 polymorphic sites in the Asian lineage due to the limited genetic diversity (Fig. 3b). The similarity between the two networks is that most of the pairwise interactions are positive. One well-known example of positive interactions is compensatory mutations, where the fixation of a second mutation rescues the preceding deleterious mutation (Sanjuan et al., 2005). Interpreting coevolving site data requires additional caveats which are discussed in Supplementary Text 2. It would be surprising if there are significant differences between the two lineages in protein structure or protein-protein interactions. The differences in networks are probably due to limited divergence (i.e. weak statistical power to predict anything) or due to the predictions being false positives. The observed correlation can be further improved by utilizing larger sample size of fully sequenced Zika viral genomes.

3.4. 3-D structure-based comparison of E-protein from epidemic and pre-epidemic ZIKV strains

Sequence alignments confirmed insertions in the glycosylation motif of E-protein of the Asian lineage, which is not the case for the African lineage (Supplementary Fig. 2). The effect of amino acid changes in tertiary structures of the ZIKV E-protein from different hosts would provide insight on host specific adaptation. The published cryo-EM structure (PDB code 5IRE) of ZIKV strain H/PF/2013 (KJ776791) isolated during the 2013/14 French Polynesia epidemic (Sirohi et al., 2016) was identified as an appropriate template that shared 100%, 97.62% and 97.20% sequence identity with the target E-protein of the 2015/16 epidemic human ZIKV (KU527068), pre-epidemic mosquito ZIKV (KF268948), and monkey ZIKV (AY632535), respectively. A total of 16 amino acid substitutions in the E-protein of ZIKV strains isolated from mosquito, monkey, and human hosts led to the subtle structural changes (Fig. 4), however the global folds are expected to be similar. Out of 16 amino acid substitutions, a total of 6 substitutions were found in central and dimerization domains (1 − 302), 4 in the immunoglobulin-like

Table 1
Classification of predicted CD4 and CD8 TCEs of 2015/16 BeH819966 ZIKV (KU365779) based on their immunogenicity scores. In each of 6 proteins (Capsid, Env-domain, NS2A, NS3-serine protease, NS4A and NS5), the total number of predicted CD4 and CD8 TCEs were further classified into non-immunogenic, immunogenic, and highly immunogenic peptides based on immunogenicity scores. The number of TCEs carrying amino acid sites under purifying selection pressure was also provided (i.e. 12 of 46 immunogenic CD4 TCEs of Capsid protein were carrying negatively selected amino acid site(s)).

				CD4				CD8			
Protein	Protein length	Amino acid position under selection	Amino acid position under selection in polyprotein	Total # of predicted TCEs	Non-immunogenic TCEs (score < 0)	Immunogenic TCEs (score < 0.5–0)	Highly immunogenic TCEs (score ≥ 0.5)	Total # of predicted TCEs	Non-immunogenic TCEs (score < 0)	Immunogenic TCEs (score < 0.5–0)	Highly immunogenic TCEs (score ≥ 0.5)
Capsid	111	38 (A)	49	97	0/48	12/46	3/3	103	2/44	7/59	0/0
Env–domain	302	69 (A)	360	288	15/136	0/140	0/12	294	9/146	0/148	0/0
		166 (K)	456		2/136	13/140	0/12		2/146	7/148	0/0
		174 (P)	464		2/136	13/140	0/12		4/146	5/148	0/0
NS2A	215	13 (L)	1170	201	13/78	0/112	0/11	207	9/77	0/130	0/0
NS3-serine protease	151	40 (A)	1559	137	10/60	5/77	0/0	143	7/62	2/81	0/0
		121 (P)	1640		11/60	4/77	0/0		9/62	0/81	0/0
NS4A	145	66 (F)	2188	131	3/43	12/83	0/5	137	3/56	6/81	0/0
NS5	641	292 (I)	3063	627	0/232	8/320	7/75	633	0/262	9/366	0/5
		345 (S)	3116		15/232	0/320	0/75		6/262	3/366	0/5
		587 (I)	3358		0/232	15/320	0/75		3/262	6/366	0/5

domain III (311–403), and 6 in the stem/anchor domain (408–504). Collectively, our findings confirm the previous results that the overall E-protein structure of ZIKV (Sirohi et al., 2016), DENV, (Kuhn et al., 2002; Zhang et al., 2013) and WNV (Mukhopadhyay et al., 2003) are similar. Subsequently, we compared the E-protein amino acid changes among the ZIKV isolates of three different hosts. Interestingly, none of three strains presented with unique amino acids at the given 16 positions, meaning that one of the residues is common in any two strains. Out of 16 variable sites in 3 domains, the epidemic human strain shared a total of 4 and 1 common amino acids with the mosquito and monkey isolates, respectively (Implant table of Fig. 4). While in the African genotype monkey and mosquito ZIKV E-proteins, it is not surprising to note that 11 residues are conserved. It suggests that during the course of adaptation in mosquito cycle, the virus acquired only 5 amino acid substitutions in the E-protein. However, the virus may have undergone more changes to adapt in human cycle. These amino acid changes could be explained by the following possibilities, (i) ZIKV inter-species cycle of transmission, (ii) ZIKV isolates of monkey, mosquito and human hosts may belong to distinct lineages, and (iii) Host-specific immune selection pressure exerted on the epitopes present in the E-protein fusion loop. The differences in tertiary structures of ZIKV E-proteins reported in our study and other flaviviruses may influence adaptation to host, cellular response and disease outcome (Sirohi et al., 2016).

An insertion of 4–5 residues (positions 153–156/157) found in the E-protein of the Asian ZIKV relative to DENV, WNV, and JEV (Japanese encephalitis virus) reflects a rapidly evolving region. Previous studies showed that a loop region surrounding the glycosylation site (144–166) (Sirohi et al., 2016) may be linked to neurotropism (Beasley et al., 2005) in WNV. Three amino acid substitutions were found in the glycosylation site of the 2015/16 epidemic human strain as compared to the pre-epidemic strain isolated from the mosquito vector (mosquito cycle). These results suggest that the conformation of glycosylation surrounding sites varies among ZIKV strains and may enhance attachment and entry of the virus into human cells, further contributing to neurotropism, virion transmission, and pathogenesis (Sirohi et al., 2016). Further experiments are required to understand the effect of this structural change on ZIKV pathogenesis.

3.5. Identification of human immunogenic CD4 and CD8 TCEs of epidemic ZIKV strains

We then examined the potential selection pressure applied by the immune system on viral epitopes. We identified 106 CD4 TCEs and 5 CD8 TCEs that had higher probability to induce an immune response (Table 1). Among the 106 highly immunogenic CD4 TCEs, 10 peptide chains contained negatively selected amino acid sites. This observation indicates that immune escape variants can be strongly selected against because of fitness cost. The detailed description of immunodominant epitope screening for vaccine design is provided in Supplementary Text 3. Furthermore, we evaluated the 11 amino acids that were under purifying selection pressure in the context of TCEs (Table 1). These amino acid sites are less likely to generate immune-escape variants, due to strong functional constraints operating on them. Totally 11% (n = 163) of 1481 CD4 TCEs and 7% (n = 99) of 1517 CD8 TCEs comprise these 11 negatively selected amino acids (Table 1; Supplementary Tables 7a, 7b). For instance, a total of 32 (21 CD4; 11 CD8) and 48 (30 CD4; 18 CD8) immunodominant and/or highly immunodominant TCEs of E and NS5 proteins, respectively, contained the 5 out of 6 negatively selected sites (Fig. 5). Interestingly, negatively selected amino acid sites in E-protein (positions 69, 166, and 174) and NS5 (292, 345, and 587) were in central and dimerization domains, and in the C-terminal RNA-directed RNA polymerase region, respectively. Any substitution at these amino acid residues is likely to be lethal or intolerable for viral replication (Suzuki, 2004; Smith et al., 2004; Arunachalam, 2014). The highly immunogenic TCEs (n = 111) and an additional 137 (92 CD4 TCEs; 45 CD8 TCEs) immunogenic peptide chains (comprising negatively selected amino acid sites) can be used for sub-unit vaccine development.

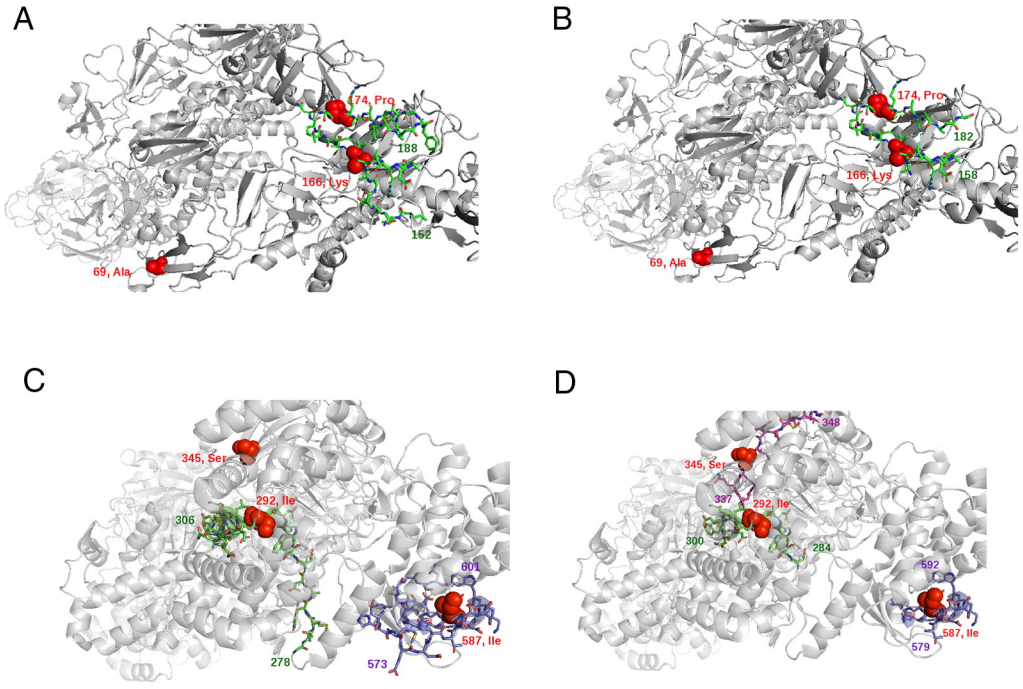


Fig. 5. Modeling amino acid sites under purifying selection pressure and predicted immunodominant (IMD) epitopes in E-protein and NS5 domains of epidemic ZIKV strains. The 3-D structures (PDB code: 5IRE) of E-protein show three negatively selected amino acid sites (Ala 69, Lys 166, and Pro 174; in red spheres). A total of (a) 21 CD4 IMD epitopes positioned from 152 to 188 (green sticks) and (b) 11 CD8 IMD epitopes positioned from 158 to 182 (green sticks) are presented in central and dimerization domains. Interestingly, none of the IMD epitopes contained the negatively selected site 69. NS5 structure (PDB code: 5TFR) also shows 3 negatively selected sites (Ile 292, Ser 345, and Ile 587, in red spheres), a total of (c) 30 CD4 IMD epitopes positioned from 278 to 306 (green) and from 573 to 601 (blue), carrying negatively selected amino acid sites (i.e. 292 and 587), and (d) 18 CD8 IMD epitopes positioned from 284 to 300 (green), 337 to 348 (magenta) and 579 to 592 (blue), carrying negatively selected sites. Notably, no CD4 IMD epitopes are predicted within site 345 (refer Table 1; Supplementary Tables 7a, 7b). In NS5, these epitopes are localized in the C-terminal RNA-directed RNA polymerase region.

Our detailed bioinformatic analyses of various Zika virus isolates determined that the Asian genotype is genetically conserved due to genome-wide negative selection, and identified specific amino acid residue changes, which may impact ZIKV virulence and host tropism. Reverse genetics approach by engineering point mutations and inter and intra-genotype chimeric viruses, would further address the biological significance of these observed changes. Deep sequencing analysis of the viral quasi-species present in various human organs, including fetal brain, placenta, blood, and bodily fluids can further improve our understanding of: 1) intra-host genetic diversity of virus; 2) the relation between intra-host genetic diversity and pathogenicity; 3) transmission bottleneck; 4) Zika tissue tropism in humans (adult vs. congenital infections); and 5) mutations in structural and non-structural proteins and their impact on viral quasi-species evolution.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.meegid.2017.03.012.

(For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Acknowledgments

This work was funded by the Cedars-Sinai Medical Center's Institutional Research Award to V.A., and NIH PO1 CA177322 to R.S. L.D. was supported by HHMI Postdoctoral Fellowship from Jane Coffin Childs Memorial Fund for Medical Research. The authors thank Dr. Laura Martinez, Cedars-Sinai Medical Center for her critical comments and help in improvising language of this manuscript.

References

Arnold, K., Bordoli, L., Kopp, J., Schwede, T., 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22: 195–201. http://dx.doi.org/10.1093/bioinformatics/bti770.

Arunachalam, R., 2013. Detection of site-specific positive Darwinian selection on pandemic influenza A/H1N1 virus genome: integrative approaches. Genetica 141:143–155. http://dx.doi.org/10.1007/s10709-013-9713-x.

Arunachalam, R., 2014. Adaptive evolution of a novel avian-origin influenza A/H7N9 virus. Genomics 104:545–553. http://dx.doi.org/10.1016/j.ygeno.2014.10.012.

Baronti, C., Piorkowski, G., Charrel, R.N., Boubis, L., Leparc-Goffart, I., de Lamballerie, X., 2014. Complete coding sequence of zika virus from a French polynesia outbreak in 2013. Genome Announc. 2 (3) pii:e00500-14. 10.1128/genomeA.00500-14.

Beasley, D.W., Whiteman, M.C., Zhang, S., Huang, C.Y., Scneider, B.S., Smith, D.R., Gromowski, G.D., Higgs, S., Kinney, R.M., Barrett, A.D., 2005. Envelope protein glycosylation status influences mouse neuroinvasion phenotype of genetic lineage 1 West Nile virus strains. J. Virol. 79:8339–8347. http://dx.doi.org/10.1128/JVI.79.13.8339-8347.2005.

Benkert, P., Biasini, M., Schwede, T., 2011. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27:343–350. http://dx.doi.org/10.1093/bioinformatics/btq662.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., Schwede, T., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 42 (W1):W252–W258. http://dx.doi.org/10.1093/nar/gku340.

Brasil, P., Pereira, J.P., Raja Gabaglia, C., Damasceno, L., Wakimoto, M., Ribeiro Nogueira, R.M., Carvalho de Sequeira, P., Machado Siqueira, A., Abreu de Carvalho, L.M., Cotrim da Cunha, D., Calvet, G.A., Neves, E.S., Moreira, M.E., Rodrigues Baião, A.E., Nassar de Carvalho, P.R., Janzen, C., Valderramos, S.G., Cherry, J.D., Bispo de Filippis, A.M., Nielsen-Saines, K., 2016. Zika virus infection in pregnant women in Rio de Janeiro - preliminary report. N. Engl. J. Med. 375:2321–2334. http://dx.doi.org/10.1056/NEJMoa1602412.

Calis, J.J.A., Maybeno, M., Greenbaum, J.A., Weiskopf, D., De Silva, A.D., Sette, A., Kesmir, C., Peters, B., 2013. Properties of MHC class I presented peptides that enhance immunogenicity. PLoS Comput. Biol. 9 (10), e1003266. http://dx.doi.org/10.1371/journal.pcbi.1003266.

Comas, I., Chakravarthi, J., Small, P.M., Galagan, J., Niemann, S., Kremer, K., Ernst, J.D., Gagneux, S., 2010. Human T cell epitopes of mycobacterium tuberculosis are evolutionarily hyperconserved. Nat. Genet. 42:498–503. http://dx.doi.org/10.1038/ng.590.

Dahirel, V., Shekhar, K., Pereyra, F., Miura, T., Artyomov, M., Talsania, S., Allen, T.M., Altfeld, M., Carrington, M., Irvine, D.J., Walker, B.D., Chakraborty, A.K., 2011. Coordinate linkage of HIV evolution reveals regions of immunological vulnerability. Proc. Natl. Acad. Sci. U. S. A. 108:11530–11535. http://dx.doi.org/10.1073/pnas.1105315108.

Delport, W., Poon, A.F., Frost, S.D.W., Pond, S.L.K., 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26:2455–2457. http://dx.doi.org/10.1093/bioinformatics/btq429.

Dick, G.W.A., Kitchen, S.F., Haddow, A.J., 1952. Zika virus. I. Isolations and serological specificity. Trans. R. Soc. Trop. Med. Hyg. 46:509–520 http://dx.doi.org/10.1016/0035-9203(52)90042-4.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinf. 5:113. http://dx.doi.org/10.1186/1471-2105-5-113.

Enfissi, A., Codrington, J., Roosblad, J., Kazanji, M., Rousset, D., 2016. Zika virus genome from the Americas. Lancet 387:227–228 http://dx.doi.org/10.1016/S0140-6736(16)00003-9.
Farci, P., Shimoda, A., Coiana, A., Diaz, G., Peddis, G., Melpolder, J.C., Strazzera, A., Chien, D.Y., Munoz, S.J., Balestrieri, A., Purcell, R.H., Alter, H.J., 2000. The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. Science 288: 339–344. http://dx.doi.org/10.1126/science.288.5464.339.
Faria, N.R., Azevedo Rdo, S., Kraemer, M.U., Souza, R., Cunha, M.S., Hill, S.C., Thézé, J., Bonsall, M.B., Bowden, T.A., Rissanen, I., Rocco, I.M., Nogueira, J.S., Maeda, A.Y., Vasami, F.G., Macedo, F.L., Suzuki, A., Rodrigues, S.G., Cruz, A.C., Nunes, B.T., Medeiros, D.B., Rodrigues, D.S., Nunes Queiroz, A.L., da Silva, E.V., Henriques, D.F., Travassos da Rosa, E.S., de Oliveira, C.S., Martins, L.C., Vasconcelos, H.B., Casseb, L.M., Simith Dde, B., Messina, J.P., Abade, L., Lourenço, J., Carlos Junior Alcantara, L., de Lima, M.M., Giovanetti, M., Hay, S.I., de Oliveira, R.S., Lemos Pda, S., de Oliveira, L.F., de Lima, C.P., da Silva, S.P., de Vasconcelos, J.M., Franco, L., Cardoso, J.F., Vianez-Júnior, J.L., Mir, D., Bello, G., Delatorre, E., Khan, K., Creatore, M., Coelho, G.E., de Oliveira, W.K., Tesh, R., Pybus, O.G., Nunes, M.R., Vasconcelos, P.F., 2016. Zika virus in the Americas: early epidemiological and genetic findings. Science 352:345–349. http://dx.doi.org/10.1126/science.aaf5036.
Faye, O., Freire, C.C.M., Iamarino, A., Faye, O., de Oliveira, J.V.C., Diallo, M., Zanotto, P.M.A., Sall, A.A., 2014. Molecular evolution of Zika virus during its emergence in the 20(th) century. PLoS Negl. Trop. Dis. 8, e2636. http://dx.doi.org/10.1371/journal.pntd.0002636.
Foy, B.D., Kobylinski, K.C., Chilson Foy, J.L., Blitvich, B.J., Travassos da Rosa, A., Haddow, A.D., Lanciotti, R.S., Tesh, R.B., 2011. Probable non-vector-borne transmission of Zika virus, Colorado, USA. Emerg. Infect. Dis. 17:880–882. http://dx.doi.org/10.3201/eid1705.101939.
Guex, N., Peitsch, M.C., 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723. http://dx.doi.org/10.1002/elps.1150181505.
Haddow, A.D., Schuh, A.J., Yasuda, C.Y., Kasper, M.R., Heang, V., Huy, R., Guzman, H., Tesh, R.B., Weaver, S.C., 2012. Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage. PLoS Negl. Trop. Dis. 6, e1477. http://dx.doi.org/10.1371/journal.pntd.0001477.
Hayes, E.B., 2009. Zika virus outside Africa. Emerg. Infect. Dis. 15:1347–1350. http://dx.doi.org/10.3201/eid1509.090442.
van Hemert, F., Berkhout, B., 2016. Nucleotide composition of the Zika virus RNA genome and its codon usage. Virol. J. 13:95. http://dx.doi.org/10.1186/s12985-016-0551-1.
Jain, R., Coloma, J., Garcia-Sastre, A., Aggarwal, A.K., 2016. Structure of the NS3 helicase from Zika virus. Nat. Struct. Mol. Biol. 23:752–754. http://dx.doi.org/10.1038/nsmb.3258.
Kawashima, Y., Pfafferott, K., Frater, J., Matthews, P., Payne, R., Addo, M., Gatanaga, H., Fujiwara, M., Hachiya, A., Koizumi, H., Kuse, N., Oka, S., Duda, A., Prendergast, A., Crawford, H., Leslie, A., Brumme, Z., Brumme, C., Allen, T., Brander, C., Kaslow, R., Tang, J., Hunter, E., Allen, S., Mulenga, J., Branch, S., Roach, T., John, M., Mallal, S., Ogwu, A., Shapiro, R., Prado, J.G., Fidler, S., Weber, J., Pybus, O.G., Klenerman, P., Ndung'u, T., Phillips, R., Heckerman, D., Harrigan, P.R., Walker, B.D., Takiguchi, M., Goulder, P., 2009. Adaptation of HIV-1 to human leukocyte antigen class I. Nature 458:641–645. http://dx.doi.org/10.1038/nature07746.
Kuhn, R.J., Zhang, W., Rossmann, M.G., Pletnev, S.V., Corver, J., Lenches, E., Jones, C.T., Mukhopadhyay, S., Chipman, P.R., Strauss, E.G., Baker, T.S., Strauss, J.H., 2002. Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. Cell 108:717–725. http://dx.doi.org/10.1016/S0092-8674(02)00660-8.
Kuno, G., Chang, G.J., 2007. Full-length sequencing and genomic characterization of Bagaza, Kedougou, and Zika viruses. Arch. Virol. 152:687–696. http://dx.doi.org/10.1007/s00705-006-0903-z.
Lanciotti, R.S., Kosoy, O.L., Laven, J.J., Velez, J.O., Lambert, A.J., Johnson, A.J., Stanfield, S.M., Duffy, M.R., 2008. Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. Emerg. Infect. Dis. 14:1232–1239. http://dx.doi.org/10.3201/eid1408.080287.
Lanciotti, R.S., Lambert, A.J., Holodniy, M., Saavedra, S., Signor Ldel, C., 2016. Phylogeny of Zika virus in Western Hemisphere, 2015. Emerg. Infect. Dis. 22:933–935. http://dx.doi.org/10.3201/eid2205.160065.
Lazear, H.M., Govero, J., Smith, A.M., Platt, D.J., Fernandez, E., Miner, J.J., Diamond, M.S., 2016. A mouse model of Zika virus pathogenesis. Cell Host Microbe 19:1–11 http://dx.doi.org/10.1016/j.chom.2016.03.010.
Mahy, B.W.J., 2010. The evolution and emergence of RNA viruses. Emerg. Infect. Dis. 16 (5):899. http://dx.doi.org/10.3201/eid1605.100164.
Marchette, N.J., Garcia, R., Rudnick, A., 1969. Isolation of Zika virus from *Aedes aegypti* mosquitoes in Malaysia. Am.J.Trop. Med. Hyg. 18 (3):411–415 https://doi.org/10.4269/ajtmh.1969.18.411.
Mlakar, J., Korva, M., Tul, N., Popović, M., Poljšak-Prijatelj, M., Mraz, J., Kolenc, M., Rus, K.R., Vipotnik, T.S., Vodušek, V.F., Vizjak, A., Pižem, J., Petrovec, M., Županc, T.A., 2016. Zika virus associated with microcephaly. N. Engl. J. Med. 374:951–958. http://dx.doi.org/10.1056/NEJMoa1600651.
Mukhopadhyay, S., Kim, B.S., Chipman, P.R., Rossmann, M.G., Kuhn, R.J., 2003. Structure of West Nile virus. Science 302:248. http://dx.doi.org/10.1126/science.1089316.
Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Pond, S.L.K., Scheffler, K., 2013. FUBAR: a fast, unconstrained Bayesian AppRoximation for inferring selection. Mol. Biol. Evol. 30:1196–1205. http://dx.doi.org/10.1093/molbev/mst030.
Musso, D., Nilles, E.J., Cao-Lormeau, V.M., 2014. Rapid spread of emerging Zika virus in the Pacific area. Clin. Microbiol. Infect. 20:O595–O596. http://dx.doi.org/10.1111/1469-0691.12707.
Nandy, A., Dey, S., Basak, S.C., Bielinska-Waz, D., Waz, P., 2016. Characterizing the Zika virus genome – a bioinformatics study. Curr. Comput. Aided Drug Des. 12:87–97. http://dx.doi.org/10.2174/1573409912666160401115812.
Oehler, E., Watrin, L., Larre, P., Leparc-Goffart, I., Lastere, S., Valour, F., Baudouin, L., Mallet, H., Musso, D., Ghawche, F., 2014. Zika virus infection complicated by Guillain-Barre syndrome - case report, French Polynesia, December 2013. Eurosurveillance 19 (9) pii20720. http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20720.
Pond, S.L.K., Frost, S.D.W., 2005. Not so different after all: a comparison of methods for detecting amino acid sited under selection. Mol. Biol. Evol. 22:1208–1222 https://doi.org/10.1093/molbev/msi105.
Pond, S.L.K., Frost, S.D.W., Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676–679 https://doi.org/10.1093/bioinformatics/bti079.
Pond, S.L.K., Frost, S.D., Grossman, Z., Gravenor, M.B., Richman, D.D., Brown, A.J., 2006. Adaptation to different human populations by HIV-1 revealed by codon-based analyses. PLoS Comput. Biol. 2:e62 http://dx.doi.org/10.1371/journal.pcbi.0020062.
Pond, S.L.K., Murrell, B., Fourment, M., Frost, S.D.W., Delport, W., Scheffler, K., 2011. A random effects branch-site model for detecting episodic diversifying selection. Mol. Biol. Evol. 28:3033–3043 https://doi.org/10.1093/molbev/msr125.
Quadeer, A.A., Louie, R.H., Shekhar, K., Chakraborty, A.K., Hsing, I.M., McKay, M.R., 2014. Statistical linkage analysis of substitutions in patient-derived sequences of genotype 1a hepatitis C virus nonstructural protein 3 exposes targets for immunogen design. J. Virol. 88:7628–7644. http://dx.doi.org/10.1128/JVI.03812-13.
Ramaiah, A., Arumugaswami, V., 2016. Ebola virus evolves in human to minimize the detection by immune cells by accumulating adaptive mutations. VirusDisease 27: 136–144. http://dx.doi.org/10.1007/s13337-016-0305-0.
Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542. http://dx.doi.org/10.1093/sysbio/sys029.
Sanderson, M.J., Donoghue, M.J., Piel, W., Eriksson, T., 1994. TreeBASE: a prototype database of phylogenetic analyses and an interactive tool for browsing the phylogeny of life. Am. J. Bot. 81:183. http://dx.doi.org/10.2307/2445447.
Sanjuan, R., Cuevas, J.M., Moya, A., Elena, S.F., 2005. Epistasis and the adaptability of an RNA virus. Genetics 170:1001–1008. http://dx.doi.org/10.1534/genetics.105.040741.
Sant'Anna, F.H., Borges, L.G., Fallavena, P.R., Gregianini, T.S., Matias, F., Halpin, R.A., Wentworth, D., d'Azevedo, P.A., Veiga, A.B., 2014. Genome analysis of pandemic and post-pandemic influenza A pH1N1 viruses isolated in Rio Grande do Sul, Brazil. Arch. Virol. 159:621–630. http://dx.doi.org/10.1007/s00705-013-1855-8.
Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13:2498–2504. http://dx.doi.org/10.1101/gr.1239303.
Sirohi, D., Chen, Z., Sun, L., Klose, T., Pierson, T.C., Rossmann, M.G., Kuhn, R.J., 2016. The 3.8 Å resolution cryo-EM structure of Zika virus. Science http://dx.doi.org/10.1126/science.aaf5316.
Smith, D.J., Lapedes, A.S., de Jong, J.V., Bestebroer, T.M., Rimmelzwaan, G.F., Osterhaus, A.D.M.E., Fouchier, R.A.M., 2004. Mapping the antigenic and genetic evolution of influenza virus. Science 305:371–376. http://dx.doi.org/10.1126/science.1097211.
Suzuki, Y., 2004. Negative selection on neutralization epitopes of poliovirus surface proteins: implications for prediction of candidate epitopes for immunization. Gene 328:127–133 http://dx.doi.org/10.1016/j.gene.2003.11.020.
Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substituions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 11: 715–724 https://doi.org/10.1093/oxfordjournals.molbev.a040023.
Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30:2725–2729 https://doi.org/10.1093/molbev/mst197.
Tian, H., Ji, X., Yang, X., Xie, W., Yang, K., Chen, C., Wu, C., Chi, H., Mu, Z., Wang, Z., Yang, H., 2016. The crystal structure of Zika virus helicase: basis for antiviral drug design. Protein Cell 7:450. http://dx.doi.org/10.1007/s13238-016-0275-4.
Victora, C.G., Schuler-Faccini, L., Matijasevich, A., Ribeiro, E., Pessoa, A., Barros, F.C., 2016. Microcephaly in Brazil: how to interpret reported numbers? Lancet 387:621–624. http://dx.doi.org/10.1016/S0140-6736(16)00273-7.
Villordo, S.M., Carballeda, J.M., Filomatori, C.V., Gamarnik, A.V., 2016. RNA Structure Duplications and Flavivirus Host Adaptation. Trends Microbiol. 24:270–283 http://dx.doi.org/10.1016/j.tim.2016.01.002.
Vita, R., Overton, J.A., Greenbaum, J.A., Ponomarenko, J., Clark, J.D., Cantrell, J.R., Wheeler, D.K., Gabbard, J.L., Hix, D., Sette, A., Peters, B., 2014. The immune epitope database (IEDB) 3.0. Nucleic Acids Res. 43:D405–D412. http://dx.doi.org/10.1093/nar/gku938.
Wang, P., Sidney, J., Dow, C., Mothé, B., Sette, A., Peters, B., 2008. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput. Biol. 4, e1000048. http://dx.doi.org/10.1371/journal.pcbi.1000048.
Wang, L., Valderramos, S.G., Wu, A., Ouyang, S., Li, C., Brasil, P., Bonaldo, M., Coates, T., Nielsen-Saines, K., Jiang, T., Aliyari, R., Cheng, G., 2016. From mosquitos to humans: genetic evolution of Zika virus. Cell Host Microbe 19:561–565. http://dx.doi.org/10.1016/j.chom.2016.04.006.
Weaver, S.C., 2017. Emergence of epidemic Zika virus transmission and congenital Zika syndrome: are recently evolved traits to blame? MBio 8 (1) pii:e02063-16. 10.1128/mBio.02063-16.
Xia, X., 2013. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. Mol. Biol. Evol. 30:1720–1728. http://dx.doi.org/10.1093/molbev/mst064.
Ye, Q., Liu, Z.-Y., Han, J.-F., Jiang, T., Li, X.-F., Qin, C.-F., 2016. Genomic characterization and phylogenetic analysis of Zika virus circulating in the Americas. Infect. Genet. Evol. 43: 43–49. http://dx.doi.org/10.1016/j.meegid.2016.05.004.
Zhang, X., Ge, P., Yu, X., Brannan, J.M., Bi, G., Zhang, Q., Schein, S., Zhou, Z.H., 2013. Cryo-EM structure of the mature dengue virus at 3.5-Å resolution. Nat. Struct. Mol. Biol. 20:105–110. http://dx.doi.org/10.1038/nsmb.2463.
Zhu, Z., Chan, J.F., Tee, K.M., Choi, G.K., Lau, S.K., Woo, P.C., Tse, H., Yuen, K.Y., 2016. Comparative genomic analysis of pre-epidemic and epidemic Zika virus strains for virological factors potentially associated with the rapidly expanding epidemic. Emerg. Microbes Infect. 5, e22. http://dx.doi.org/10.1038/emi.2016.48.