Infection, Genetics and Evolution 71 (2019) 98–107

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Research paper

Complete genome of DENV2 isolated from mosquitoes in Mexico

Selene Zárate[a,*], Fabiola Hernández-Perez[a], Blanca Taboada[b], Norma E. Martínez[c], Sofía L. Alcaráz-Estrada[d], Oscar del Moral[e], Martha Yocupicio-Monroy[a]

[a] Posgrado en Ciencias Genómicas, Universidad Autónoma de la Ciudad de México, México
[b] Instituto de Biotecnología, Universidad Nacional Autónoma de México, México
[c] Servicios Estatales de Salud de Guerrero, México
[d] División de Medicina Genómica, CMN 20 de Noviembre-ISSSTE, México
[e] Facultad de Ciencias Químico Biológicas, Universidad Autónoma de Guerrero, México

ARTICLE INFO

Keywords:
Mosquito-derived sequences
DENV2
Sequence polymorphism

ABSTRACT

Dengue virus is the most prevalent arbovirus in Mexico, and although the diversity of this virus has been studied, the vast majority of sequences have been derived from viruses isolated from the human host. In this work, we aimed to sequence and to analyze DENVs derived from wild mosquitoes captured in Acapulco Guerrero, Mexico. We succeeded in determining three full genome sequences of such viruses and were able to compare them with other reported sequences from human and mosquito-derived DENVs. We found 15 nonsynonymous and 88 synonymous substitutions that were present more frequently in mosquito viruses than what would be expected by chance, although the limited number of genomes reported so far puts a constraint on the conclusions that can be derived from these analyses. Also, given the high depth of coverage attained in one of the genomes a variant analysis was carried out, finding 68 polymorphic sites in this genome. Interestingly, six of them corresponded to SNV that were detected as potentially differential between mosquitoes and humans, indicating that a that at least some positions may be maintained as polymorphic, which may facilitate host transmission.

1. Introduction

Dengue virus (DENV) is an arbovirus transmitted by *Aedes aegypti* and *Aedes albopictus* mosquitoes that causes around 100 million reported infections per year, affecting the tropical and subtropical regions of the world. Although the estimates of disease burden vary, all agree that there has been an increase in the number of cases in the last decades (Wilder-Smith and Byass, 2016). DENV is classified into four serotypes, each representing a separate transmission event to the urban cycle and constituting a monophyletic group (Holmes and Twiddy, 2003). Each serotype is further classified into genotypes that also correspond to monophyletic groups, but the diversity threshold that it is used to separate genotypes varies with the serotype. The sequence of gene E that encodes for the surface glycoprotein is routinely used to assign a genotype to a newly sequenced strain. There are five genotypes within DENV2 serotype, which roughly correlate with the geographic distribution of the virus (Weaver and Vasilakis, 2009). However, this distribution is rapidly changing, due to the genotypes broadening their range. In Mexico, the four serotypes of DENV have circulated, albeit not at the same time, and in recent years DENV1 and DENV2 have been the most prevalent (Torres-Galicia et al., 2014). Few studies have described the molecular epidemiology of DENV in Mexico, but the results so far indicate that DENV diversity is characterized by the successive introduction of DENV lineages that replace pre-existing varieties (Carrillo-Valenzo et al., 2010; Díaz et al., 2006; Gardella-Garcia et al., 2008; Rivera-Osorio et al., 2011). This observation is consistent with what has been reported in other regions of the world (Weaver and Vasilakis, 2009).

Although DENV is an RNA virus with a mutation rate approximate to other fast-evolving viruses, natural populations do not accumulate mutations at the expected rate. Thus, it has been proposed that DENV evolution is constrained, at least in part, by its transmission cycle (Lambrechts and Lequime, 2016). For instance, it has been reported that DENVs grown in insect cells accumulate fewer mutations than those cultured in human cells, despite both viral populations showing similar fitness gains in their respective cell type (Vasilakis et al., 2009). Furthermore, in DENV transmitted from infected patients to mosquitoes, the variance acquired in most positions in the human host reverts to consensus when it is transmitted back to mosquitoes (Sessions et al., 2015). Additionally, it has been recently shown, that DENVs from

* Corresponding author at: Posgrado en Ciencias Genómicas, Universidad Autónoma de la Ciudad de México, San Lorenzo 290, Ciudad de México C. P. 03100, Mexico.
E-mail address: selene.zarate@uacm.edu.mx (S. Zárate).

https://doi.org/10.1016/j.meegid.2019.03.018
Received 5 October 2018; Received in revised form 13 January 2019; Accepted 20 March 2019
Available online 21 March 2019

mosquitoes that were fed patients' blood attained a similar level of diversity of DENV replicating in humans. However, the particular single nucleotide variants (SNV) changed with the host, with only around 10% of the SNV being maintained after transmission, in part due to population bottlenecks (Sim et al., 2015). Also, cold spots for mutations were identified in DENV replicating in mosquitoes, which revealed evolutionary constraints imposed by this host.

Moreover, it has been reported that in experimental transmission of DENV from human to mosquitoes, the diversity of the viral population decreased at first due to a genetic bottleneck upon transmission. Then, it was followed by an increase in diversity that was later restricted due to a strong purifying selection and that the levels of attained diversity varied with the host genotype (Lequime et al., 2016). Altogether, these findings indicate that DENV replication in the mosquito limits the evolution of the natural populations of this virus, although the mechanisms are unclear. For instance, in West Nile virus it has been reported that mosquito RNAi response can drive viral diversification in the targeted genome regions (Brackney et al., 2009), and a similar mechanism could also affect DENV diversity.

One of the challenges of understanding the molecular evolution of DENV natural populations is the bias in the available sequences in public databases. There are only a few reported sequences obtained from wild mosquitoes, which limits our understanding of how the transmission between vector and host shapes the diversity of this virus in the wild. In this study, we determined the full genome sequence of DENV from mosquitoes captured in the neighborhood of confirmed DENV cases, during the rainy season of 2012, in Acapulco, Guerrero, and analyzed their diversity.

2. Methods

2.1. Sample preparation and sequencing

Pools of DENV positive mosquitoes were collected by the Laboratorio Estatal de Salud Pública del Estado de Guerrero as previously described (Dzul-Manzanilla et al., 2016). Briefly, adult *Ae. aegypti* were collected indoors during the rainy season of 2012, in the neighborhood of confirmed and suspected DENV cases in Acapulco, Gro. After collection, the specimens were maintained at 4–8 °C. Adult female mosquitoes were grouped in pools of 15 to 30; each pool was homogenized, aliquoted, and stored at −70 °C. Total RNA was extracted from the samples using Zymo-Spin RNA extraction kit according to the manufacturer's instructions. DENV presence and serotype were verified using a real-time reverse-transcriptase polymerase chain reaction (RT-PCR) protocol to amplify a conserved region of NS5 (Chao et al., 2005; Lanciotti, 2003). Afterward, 200 ng of total RNA from the five DENV positive pools samples were depleted of rRNA using NEBNext rRNA Depletion Kit (Human/Mouse/Rat) according to manufacturer instructions. Then, rRNA free samples were sent to the USMB-IBT UNAM where RNAseq Illumina shotgun libraries were built, followed by paired-end sequencing of 75 base pair (bp) length using a MySeq system.

2.2. Sequence assembly

First, for each sample, the next pre-process analysis was performed: i) FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to verify reads quality, ii) adapters were removed with Cutadapt v1.11 (Martin, 2011) and iii) the 5' and 3' ends with no-call sites (N residues) and with low-quality scores (< 20, corresponding to an error rate of > 1%) were trimmed with AfterQC (Chen et al., 2017) ensuring reads with a minimum length of 50 bp long and with no more than 10 N nucleotides.

The remaining sequences were considered valid reads. Then, identification of candidate *Flaviviridae* reads was done by aligning valid reads, with at least 70% identity, with a flavivirus genomes database using Bbmap (Bushnell et al., 2014). This database was built from NCBI complete genomes annotated as belonging to the *Flaviviridae* family and consists of 7020 sequences from 94 different flaviviruses although database bias makes that 95% of the records correspond to only six viruses. Afterward, the reads that matched with this *Flaviviridae* database were used to do a *de novo* assembly with the IDBA software (Peng et al., 2012) to assemble large contigs, which were analyzed with blastn. Finally, depth coverage and map quality was assessed using Qualimap 2.2.1 (Okonechnikov et al., 2015).

2.3. Genotyping

DENV genotyping was carried out using phylogenetics, in which reference sequences from each known genotype were included. To perform this analysis, we downloaded 31 gene E sequences, from the ViPR database (Pickett et al., 2012), that represent the six reported genotypes of DENV2. Using these reference and problem sequences a NJ tree was built with 100 bootstrap replicates using SeaView4.6 (Gouy et al., 2010), and genotype was called out based on clustering.

2.4. Phylogeography

Two phylogeographic trees were reconstructed using the programs BEAST2.4.3, BEAUti, and TreeAnnotator (Drummond et al., 2012), introducing the country as a trait and the collection year. For the first tree, the 391 complete genomes of DENV2 AsianAmerican genotype from North and South American countries available from from the ViPR database were downloaded. These sequences were aligned to the genomes assembled in this work using MUSCLE for a total of 394 sequences (Edgar, 2004). Given the high variability in the lengths of the reported UTRs, only the CDS was used for the phylogenetic reconstruction. As there were only 17 full genome sequences reported from Mexico in ViPR, a second tree was built using the E coding region only. For this second phylogeny, we employed 97 sequences from the E coding region of DENV2 AsianAmerican genotype reported for Mexico found in ViPR and the E coding region of the viruses reported in this work. In both cases, the MCMC was run for 25 million generations, using the TN93 substitution model and a gamma distribution with four categories as the site heterogeneity model. The resulting file was analyzed with Tracer 1.6 (Rambaut et al., 2018) to check for convergence and to determine the Burnin proportion. Finally, TreeAnnotator was used to build the maximum clade credibility tree, which was visualized with FigTree 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree).

2.5. Interhost diversity

To identify single nucleotide variants (SNVs) in the mosquito-derived sequences the MSA built to reconstruct the full genome phylogeny was employed. To determine the presence of SNV observed in mosquito-derived sequences, but absent or rare in human-derived sequences, we carried out an analysis using the tool metaCATS (Pickett et al., 2013). This tool compares the distribution of polymorphisms between groups of sequences, in our case we grouped the sequences by host of origin. For positions where a significant difference between groups was found, the polymorphisms were mapped over a tree to determine if the substitutions were synapomorphic or homoplasic; this was carried out using stochastic mutation mapping with the program SIMMAP 1.5 (Bollback, 2006).

2.6. Intrahost diversity

To determine the presence of minority variants in the viral populations sequenced from mosquitoes, we carried out variant analysis using the program VirVarSeq (Verbist et al., 2015). In our case, we only considered as true variants the ones present both in forward and reverse sequences, and that were present in at least 1% of the reads, with a

mean quality score of at least 20.

2.7. Relative codon usage

The sequences from the full genome MSA were analyzed to determine their Relative Synonymous Codon Usage (RSCU) value using CAIcal (http://genomes.urv.cat/CAIcal/E-CAI). To compare RSCU values between viruses a hierarchical clustering analysis was carried out using the pvclust package (Suzuki and Shimodaira, 2006) in R (R Core Team, 2016).

3. Results

3.1. Sequence generation

The pools of mosquitoes that were reported as positive for DENV2 by Dzul-Manzanilla et al. (2016) were re-analyzed to determine if the aliquot used in this work was also positive to DENV2 virus; this is a necessary step due to the heterogeneity of the sample. Of the seven positive pools previously reported, we could only confirm five pools as positive by real-time RT-PCR. Characteristics of the pools, including sampling locations and dates, are presented in Table A.1. DENV positive samples were treated to remove rRNA and sequenced using the MySeq platform at the USMB-IBT UNAM as described in Methods. After preprocessing the sequences to eliminate adapters, low quality, and short reads, a *Flaviviridae* sequence database was used to separate the reads that had at least 70% identity with the reference index, using BBmap. The number of reads of each sample that matched to the index is shown in Table 1. The disparities in the number of matched reads in each sample could be due to differences in viral load or to the presence of more than one positive mosquito in the pool.

For each sample, a *de novo* assembly was built using the matched sequences as described in Methods. The number of assembled contigs and their length are shown in Table 1, along with the mean coverage attained. For sample S2, two large contigs were obtained, which have similar coverage of around 14.5x, but the sequence of 71 nucleotides starting at position 3805 could not be determined, we dubbed this genome as G2AE. For samples S1 and S5, full genome sequences were obtained. In sample S1, a single contig of 10617 bp was achieved, with a mean coverage of 29.47x and 95% of the positions being covered with a depth of at least 10x, the resulting genome was named G1AE. In the case of S5 a single contig of 10705 nucleotides was assembled with a mean coverage of 906.37x and 99% of the sites being covered at a depth of at least 170x, this genome was dubbed G3AE. For samples S3 and S4, it was impossible to assemble large contigs due, in part, to the limited number of reads that matched the *Flaviviridae* genomes database.

In addition to obtaining the complete coding sequence, partial sequences from the UTRs were also determined. For G1AE, the length of 5′UTR and the 3′UTR were 35 and 406, for G2AE 66 and 409, and for G3AE 87 and 442, respectively.

3.2. Genotyping

Using 31 reference sequences of the E coding segment that represent the six different DENV2 genotypes and the mosquito-derived sequences G1AE, G2AE and G3AE, a neighbor-joining tree with 100 bootstrap replicas was reconstructed. All three assembled genomes were identified as belonging to the AsianAmerican genotype (data not shown). As expected, the bootstrap support for the clades grouping each genotype was 100%.

3.3. MCC tree

The phylogenetic reconstruction using full genome sequences was carried out with BEAST as explained in the methods section. First, we constructed a multiple sequence alignment that included all full genome sequences classified as AsianAmerican genotype obtained from human or mosquito hosts in North and South American countries. Before this report, only two full genome mosquito-derived sequences that belonged to the AsianAmerican genotype had been reported; these sequences were from Yucatán, Mexico from the years 2009 and 2010. The alignment consisted of 389 human-derived sequences and five mosquito-derived sequences (for a total of 394 sequences). The tree in Fig. 1 shows clustering of human-derived sequences by country and more broadly by geographical area, (i.e. clustering of South American vs. Central/North America sequences). Also, within these clades, sequences cluster also by year of sampling. In this regard, South American sequences cluster into two groups, while USA/Puerto Rico sequences are in a separate cluster. In the case of Mexican sequences, they group into two clades, both of which have sequences from Nicaragua as sister clades. Moreover, sequences from Guatemala and Belize fall in the same global cluster. Regarding mosquito-derived sequences, not surprisingly our three sequences cluster together but are also closely related to the mosquito-derived sequences from Yucatan. Nevertheless, the lack of more sequences from mosquito limits the possible interpretation of this phylogenetic relationship.

One of the limitations of the tree shown in Fig. 1 is that most of the full genome sequences are human-derived, and there are only 17 full genome Mexican sequences in the ViPR database, and therefore only 20 (17 plus the three reported in this study) in the tree. Moreover, all of the full genome mosquito sequences are from Mexico, and it is not very surprising that the mosquito-derived sequences appear to cluster together. To gain a better insight into the distribution of mosquito and human-derived sequences sampled in Mexico, we built a second tree, using only the coding region of E. We downloaded all 97 sequences of the DENV2 E region of AsianAmerican genotype available from Mexico and included the three sequences described in this study. In the phylogeny shown in Fig. 2, it is clear that each of the mosquito-derived sequences from Guerrero clusters with a separate group of human-derived sequences that have been circulating in Mexico in the last eight years. We find this remarkable given that the mosquito-derived sequences determined in this study were sampled over a short period within the same year and in the same city. Since most of the sequences downloaded from the databases do not have the information of the states from where they were sampled, it is difficult to make a hypothesis about the patterns of circulation of DENV in Mexico. However, it is likely that the downloaded sequences were sampled from various states of the country and therefore the DENV circulating in Mexico lack

Table 1
Number of mapped reads and contigs assembled from 5 DENV2 positive mosquito pools.

Sample	Number of total paired-end reads	Number of matched reads	Number of contigs	Length of contigs (nt)	Mean coverage (s.d.)	Name of assembled genome (AC number)
S1	13 653 991	4 195	1	10 617	29.47 (13.69)	G1AE (MH781014)
S2	11 957 390	1 736	2	3804 6775	15.35 (5.65) 14.20 (6.07)	G2AE (MH781013)
S3	14 190 444	1 026	ND	–	–	–
S4	11 745 024	1 176	ND	–	–	–
S5	7 790 660	130 086	1	10705	906.37 (293.43)	G3AE (MH781015)

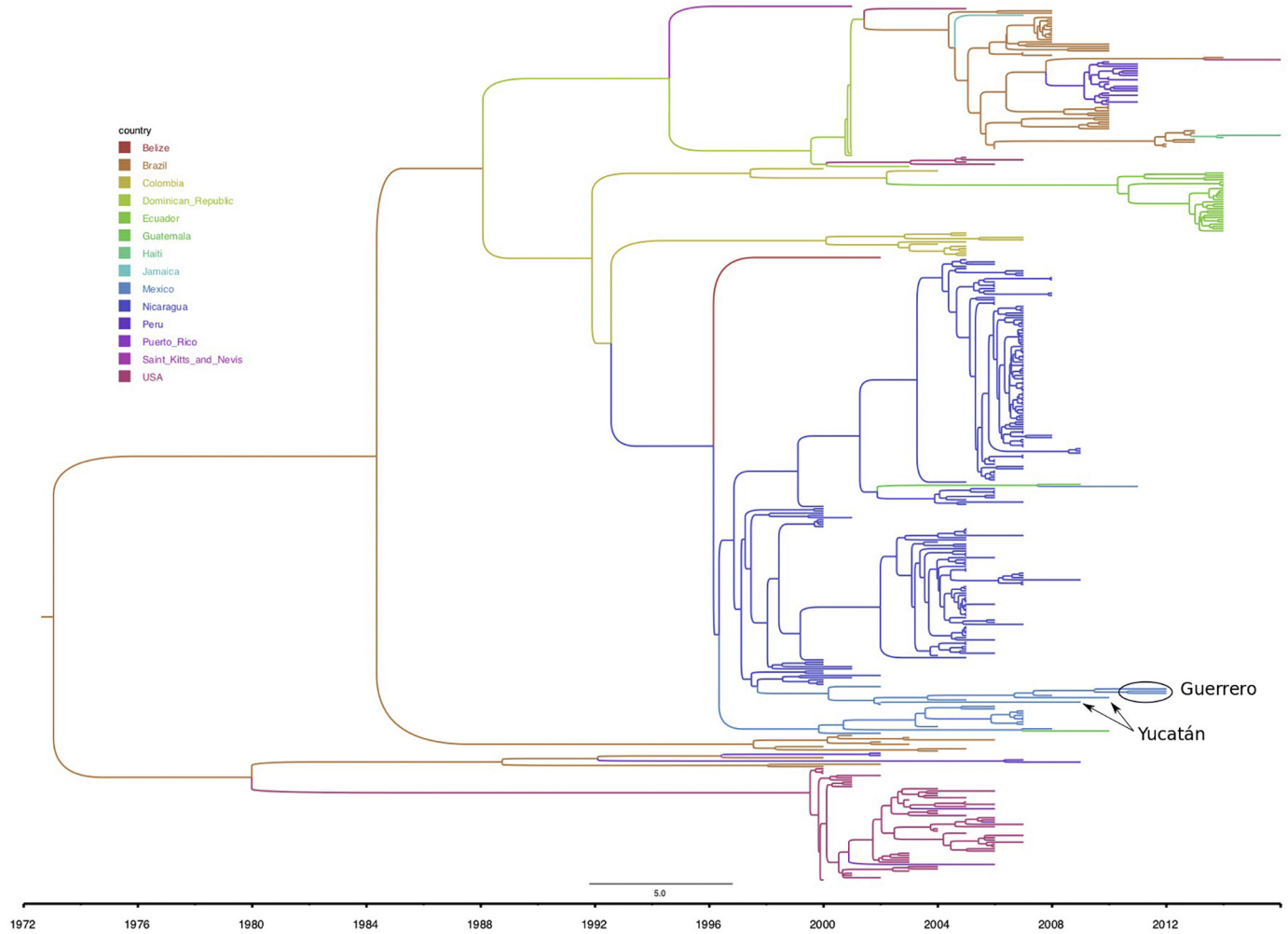

Fig. 1. Maximum clade credibility (MCC) tree of DENV2 Asian-American genotype full genome sequences. 394 full genome sequences from DENV2 viruses were used to reconstruct an MCC tree using BEAST. The MCMC was run for 25 million generations, and the tree was built using a 0.25 Burnin proportion. The reported country and year of collection were included as data in the reconstruction. In the figure, the branches are colored by country of origin according to the code shown in the figure, and clustering by country of origin can be observed. Mosquito-derived sequences from this work (Guerrero) and the databases (Yucatán) are indicated with arrows.

geographical and temporal structure, and behave like a single population.

3.4. Inter-host variants

To assess the presence of SNV in the genomes G1AE, G2AE, and G3AE, we used the multiple sequence alignment used to reconstruct the complete genome phylogeny. As mentioned earlier, the alignment consisted of 389 human-derived sequences and five mosquito-derived sequences and was analyzed using the tool metaCATS, that looks for differences in the nucleotide frequency of polymorphic sites between groups. This analysis showed 15 nonsynonymous (shown in Table 2) and 88 synonymous (shown in Table A.2) SNVs that were more frequent in mosquito viruses than what would be expected by chance. To account for multiple comparisons, the cutoff value to consider statistical significance was a Bonferroni corrected p-value of $5x10^{-6}$.

To determine if the sequences that share a particular SNV cluster in the tree, indicating they correspond to synapomorphic changes, or if their acquisition was independent, we mapped the SNV that resulted in nonsynonymous substitutions described in Table 2 over the phylogeny using SIMMAP. We found that mosquito-related SNVs found at positions 5173, 6156, and 7552 were also present in 6 to 9 closely related Mexican sequences, despite that one more of the mosquito-derived sequences did not have this particular SNV. This observation may indicate that these polymorphisms are common in DENV2 sequences circulating in Mexico from the year 2002 to 2012. In the case of the SNV in position 794 it was shared by all Mexican sequences included in the analysis but was a minority variant when compared with non-Mexican sequences. On the other hand, the SNV on position 9826 was shared not only with Mexican human-derived sequences from 2002–2009 but also with closely related Nicaraguan sequences from 2000 and 2001 but not with other Central or South America sequences. These SNVs are likely to be acquired by common descent and may have appeared only once in the phylogeny.

In contrast, the SNV in position 1915 was only shared with a Brazilian sequence from 2005, and the polymorphism 6511 was common to sequences from a closely related Mexican sequence from 2008 but also present in sequences from Peru from 2010 and 2011 and sequences from Ecuador 2014, but not with ore closely related sequence of Central America. Meanwhile, the SNV in position 6796 was only shared with a sequence form Belize 2002. These SNVs seem to be have appeared multiple times in the phylogeny and may be the result of parallel evolution. Finally, the SNVs found in positions 152, 307, 772, 2633, 8272, 8350, and 10160 were unique to one or more of the mosquito-derived sequences present in the alignment.

Also, the number of SNVs that are more frequent in mosquito-

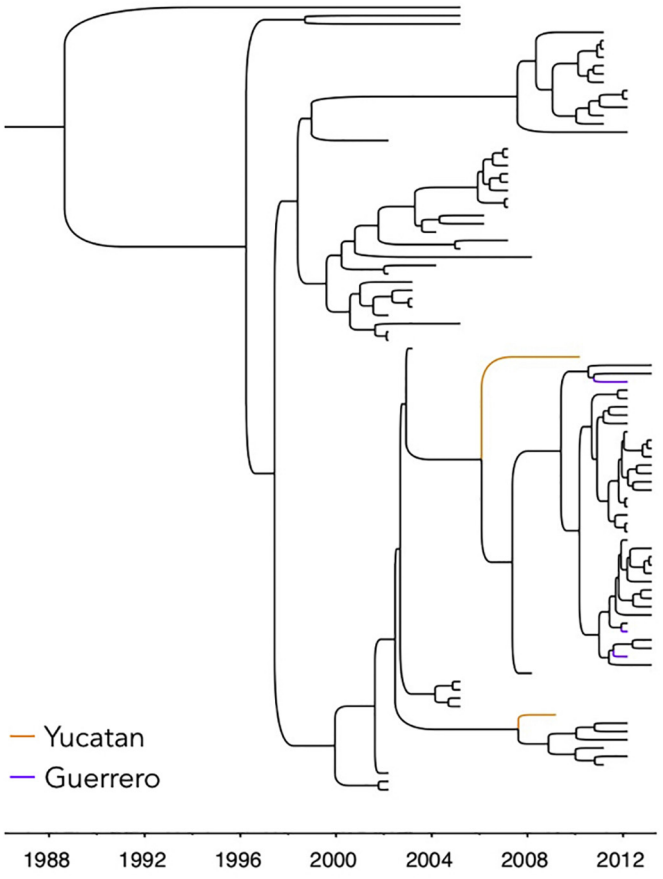

Fig. 2. Maximum clade credibility (MCC) tree of DENV2 AsianAmerican genotype Mexican sequences of the E encoding region. One hundred available sequences from the region encoding the protein E of DENV2, that were collected in Mexico, were used to reconstruct an MCC tree using BEAST. The MCMC was run for 25 million generations, and the tree was built using a 0.25 Burnin proportion. The reported year of collection was included in the reconstruction. In the figure, the branches of mosquito-derived sequences are colored, in purple the sequences from Guerrero, and in orange, the sequences form Yucatan are shown. It can be observed that these mosquito-derived sequences do not cluster together.

Table 2
Non-synonymous substitutions present in higher frequency in mosquito-derived sequences.

Position	Gene	G1	G2	G3	Freq in MDseq[a]	Freq in HDseq[b]	AA	p-Value
152	C		✓		0.20	0	V > A	8.23×10^{-17}
307	C			✓	0.02	0.0026	G > C	8.3×10^{-17}
772	M		✓		0.20	0	I > V	8.2×10^{-17}
794	M	✓	✓	✓	1	0.013	R > K	6.7×10^{-42}
1915	E		✓		0.20	0.0026	T > A	2.6×10^{-8}
2633	NS1		✓		0.20	0.013	S > F	8.2×10^{-17}
5173	NS3	✓	✓	✓	0.80	0.0026	E > Q	5.2×10^{-54}
6156	NS3	✓	✓		0.80	0.01	M > V	2.5×10^{-33}
6511	NS4A	✓	✓	✓	0.80	0.02	I > V	7.1×10^{-22}
6796	NS4B	✓	✓	✓	0.60	0.0026	S > P	9.1×10^{-38}
7552	NS5	✓	✓	✓	1	0.01	I > V	1.1×10^{-46}
8272	NS5	✓		✓	0.40	0	I > V	1.1×10^{-33}
8350	NS5			✓	0.20	0	H > Y	8.2×10^{-17}
9826	NS5	✓	✓	✓	1	0.062	S > P	9.1×10^{-14}
10160	NS5			✓	0.20	0	A > V	8.2×10^{-17}

[a] MDseq Mosquito-derived sequences.
[b] HDseq Human-derived sequences.

Table 3
Distribution by of SNVs with higher frequency in mosquito-derived sequences.

Gene	Gene length	Nonsynonymous	Synonymous	Frequency of SNVs
C	342	2	1	0.0088
prM	273	0	3	0.0110
M	225	2	2	0.0178
E	1485	1	14	0.0101
NS1	1056	1	4	0.0047
NS2A	654	0	4	0.0061
NS2B	390	0	4	0.0103
NS3	1854	2	22	0.0129
NS4A	381	1	2	0.0079
2K	69	0	2	0.0290
NS4B	744	1	6	0.0094
NS5	2787	5	24	0.0104

derived sequences roughly correlates with the length of the gene in which they are located, as shown in Table 3. Interestingly, the frequency of non-synonymous variants present in gene E is much lower than the frequency of the synonymous ones; this is of note since the gene E which encodes for the surface glycoprotein is considered as a highly variable one. Interestingly, most of the nonsynonymous variants are located in nonstructural genes and particularly in NS5. Moreover, some of the nonsynonymous substitutions are nonconservative such as the substitutions of Ser by Pro in the proteins NS4B and NS5, Glu by Gln in NS3 and His by Tyr in NS5. Also of interest, the regions encoding for M and NS3 proteins have a higher frequency of SNV than average, which is 0.015 SNV per position, while NS1 and NS2A have the lowest proportion of SNVs.

Given the high number of synonymous SNV observed in the alignment we carried out a codon usage analysis by calculating the RSCU for each of the sequences of the alignment and then using a hierarchical clustering algorithm to determine If there was a difference in codon usage between mosquito and human-derived sequences. However, the clustering obtained reflected a geographical grouping similar to the pattern observed in the phylogeny (data not shown).

3.5. Intra-host variants

Given the depth of coverage attained in the sequencing of genome G3AE, an intra-host diversity analysis was carried out. Using the program VirVarSeq, we found 52 polymorphic positions, in which the polymorphism was found in at least one percent of the lectures, both in the forward and reverse reads. Of these variants, 17 were highly frequent, as shown in Table 4, meaning that they were present in at least 20% of the forward and reverse reads. Additionally, 28 variants resulted in nonsynonymous substitutions, of which 6 were high-frequency ones. On the other hand, 22 substitutions were synonymous, with half being present at high frequency. Interestingly, there is an excess of nonsynonymous low-frequency substitutions, (ChiSquare, $p = 0.02$). The specific nonsynonymous polymorphic sites are shown in Table 5, while the synonymous polymorphic sites are shown in Table A.3. The distribution of the variants was evenly spread throughout the genome, as shown in Table 6, but interestingly, the frequency of polymorphic sites was ten times higher in the coding regions for C and NS4A.

Interestingly, six sites, 432, 1362, 3265, 6156, 8350 and 10160 that were identified as polymorphic in G3AE and were also found to have a different SNV frequency when comparing human and mosquito-derived

Table 4
Polymorphic site classification by frequency and type

	Low Frequency	High Frequency
Synonymous	11	12
Nonsynonymous	22	5
Missense	2	0

Table 5
High frequency polymorphisms in G3AE

Position in the CDS	Gene	SNV	Frequency in forward reads	Frequency in reverse reads	Type pf change (S[a] or NS[b])	AA change
136	C	T > C	0.33	0.34	S	L
489	prM	T > C	0.35	0.33	S	I
1326	E	T > C	0.38	0.42	S	I
1362	E	G > A	0.39	0.39	S	E
2354	NS1	A > G	0.4	0.39	NS	N > S
3018	NS1	G > C	0.41	0.44	S	L
5427	NS3	T > C	0.39	0.37	S	D
6156	NS3	G > A	0.32	0.32	NS	M > I
6369	NS4A	G > A	0.34	0.35	S	V
6493	NS4A	C > T	0.4	0.35	S	L
6555	NS4A	T > C	0.39	0.38	S	S
7758	NS5	A > G	0.45	0.44	S	K
8039	NS5	C > T	0.4	0.46	NS	S > L
8350	NS5	T > C	0.41	0.39	NS	Y > H
9006	NS5	G > C	0.37	0.34	S	G
9867	NS5	C > T	0.43	0.37	S	H
10160	NS5	T > C	0.24	0.22	NS	V > A

[a] S Synonymous substitution.
[b] NS Nonsynonymous substitution.

Table 6
Distribution of polymorphic sites present in G3AE by gene

Gene	Gene length	Non-synonymous		Synonymous	
		HF	LF	HF	LF
C	342			1	2
prM	273			1	
M	225		1		
E	1485		3	2	1
NS1	1056	1	2	1	
NS2A	654				1
NS2B	390		1		
NS3	1854	1	3	1	1
NS4A	381		3	3	
2K	69		2		
NS4B	744		3		
NS5	2787	3	5	2	5

sequences. The polymorphisms in positions 432 and 1362 result in synonymous substitutions and the minority variants are present in other human-derived DENV sequences reported in the database. On the other hand, the minority variant in position 3265 introduced a nonsynonymous substitution in genome G3AE that has not been previously reported; however, it is worth noting that also a synonymous SNV was identified in this position when comparing human- and mosquito-derived sequences. The SNV in position 6156 results in a nonsynonymous substitution of Met for Val, while the polymorphisms in G3AE results in a change of Met for Ile. Finally, in positions 8350 and 1060 the nonsynonymous SNVs are only present in G3AE, but the positions are polymorphic in this genome. The observation that some positions of DENV are maintained as polymorphic at the intrahost level (as observed in genome G3AE) and at the inter-host level (as evidenced by the detected SNVs) suggest that the polymorphic states may facilitate DENV transmission between mosquitoes and humans, however more evidence is needed to confirm the observations made in the G3AE genome.

4. Discussion

In this work, the full genome sequences of three mosquito-derived DENV2 were determined. These new sequences are a significant contribution to the databases since only two other DENV2 genomes derived from mosquitoes have been reported. Although it is generally recognized that the diversity and evolution of dengue virus are constrained by the necessity of carrying out their replication cycle in two distinct hosts, there are only a few studies that aim to determine the diversity of DENV in wild populations of mosquitoes. We found several positions that have different frequency of SNV in mosquito-derived viruses as compared to those isolated from humans, but the majority correspond to synonymous substitutions. This observation is not surprising considering that it has been proposed that genetic drift and negative selection seem to dominate DENV evolution in mosquitoes (Lequime et al., 2017, 2016), albeit this has been determined only in experimentally infected mosquitoes.

The SNVs with higher frequency in mosquito-derived sequences were distributed throughout the genome at an average frequency of 0.0115 SNV per position, however, in some genes, there were deviations to this frequency. It remains to be determined if these differences in SNVs numbers by gene reflect the overall diversity of the genes or are particular for mosquito-derived sequences. For instance, the presence of cold spots of mutations in DENV2 replicated in artificially infected mosquitoes has been reported (Sim et al., 2015) indicating a constraint in mutation acquisition during virus replication in mosquitoes in some genomic regions that were not constrained to accumulate mutations when replicating in the human host. It is worth noting that the E encoding region had a slightly lower diversity than the average gene, and also that 14 out of 15 SNVs identified in this genomic region resulted in synonymous changes; this was surprising since this region Is often thought of as the most diverse one and therefore is frequently sequenced to carry out phylogenetic analyses and to call out the genotype of DENVs. As full genome sequences accumulate, it will be useful to examine if other regions will be better markers for studying DENV diversity and evolution.

Of note, the distribution of SNVs among mosquito and human-derived sequences showed a different geographical distribution. Some of these SNVs were present only in Mexican sequences or in Mexican and closely related Nicaraguan sequences, possibly indicating that these SNVs constitute shared derived states (synapomorphies) that only arose once in the phylogeny. On the other hand, some SNVs that were found in one or a few Mexican sequences were also present in distantly related sequences (i.e., South American sequences), which may be evidence of parallel evolution.

Interestingly, the majority of the SNVs associated with mosquito-derived sequences identified in this work correspond to synonymous substitutions, but a codon usage analysis revealed no relation between codon usage and host of origin, despite a recent report indicating that in arboviruses selective forces influence codon usage (Velazquez-Salinas et al., 2016). Interestingly, three of synonymous SNVs were found to be polymorphic in G3AE, which may indicate the maintenance of a polymorphic state, at least in a few positions of the genome, upon transmission. Whether or not these polymorphisms contribute to the viral fitness remains to be established. Also, some of the intrahost polymorphic positions are unique to the virus analyzed, or correspond to interhost synonymous SNVs, indicating that these positions are not likely to be mutationally constrained. However, with the data presently available is not possible to determine if the polymorphic state of these positions is favored by a rapid transmission dynamic that prevents the fixation of a single state or is a result of balancing selection, which maintains polymorphic states to facilitate mosquito-human transmission. To determine how diversity is lost/preserved during DENV transmission it will be necessary to determine the presence of polymorphic positions in both naturally infected humans and mosquitoes and to look for evidence of parallel evolution that will indicate positive selection. As mentioned above, the occurrence of parallel evolution has been established in mosquitoes artificially infected with DENV1 (Sessions et al., 2015), suggesting that at least some genomic sites may be evolving under positive selection.

Additionally, most of the nonsynonymous substitutions identified in this work were located in the nonstructural genes, of particular interest

are NS4B and NS5 which seem to be the ones with the highest accumulation rate. To address the evolutionary mechanism that gives rise to the observed diversity more mosquito-derived sequences need to become available. For instance, it is worth noting that in the full genome phylogenetic analysis all of the mosquito-derived sequences cluster together, but when using only the gene E for which many more sequences are available the mosquito-derived sequences fall into different clusters. It would be interesting to determine if the observed SNVs are common to DENV2 viruses circulating in Mexico or if they have arisen multiple times. However, this type of analysis can only be carried out presently in the region that encodes for E since only a few sequences from other genomic regions are currently available.

As mentioned earlier, some evidence of the role that DENV replication in mosquitoes has on viral diversity has been generated using artificially infected mosquitoes. For instance, no observable differences in the consensus sequences were found between input and salivary glands viruses in mosquitoes artificially infected with DENV1 (Lequime et al., 2017). Also, no evidence of directional selection has been found when the diversity of DENV was determined in various mosquito anatomical compartments (Lequime et al., 2017, 2016). It remains to be established if the observations from artificially infected mosquitoes will be confirmed in wild-caught mosquitoes. For instance, evolutionary stasis of DENV is inferred from the available sequences which are mostly derived from human samples that are determined at a population level, i.e., they constitute the consensus sequence, and do not reflect the mutations that might be generated in the mosquito, nor the intra-host variability that is generated while the virus replicates in the human host.

One of the most substantial caveats of this work is that very few mosquito-derived DENV2 sequences have been reported, and only five sequences were available for this analysis. Additionally, all sequences are from Mexico, albeit from different states, two previously reported which are from Yucatan in the Caribean, from mosquitoes collected in 2009 and 2010 and the sequences determined in this study which are from Acapulco which is located on the Pacific Coast, 2012. Despite the geographical and time differences between these five sequences they cluster together in the phylogeny, but as mentioned earlier, there is the issue that no other full genome sequence from Mexico has been reported in 2012 and later years. On the other hand, when the phylogeny was carried out using only the E gene sequence, the mosquito-derived sequences do not form a cluster even though they were collected in the same city during the same year. This observation indicates that geographical clustering of DENV2 sequences inside Mexico does not occur, contrary to what is observed at a larger scale in with South American and Central American sequences form separate clusters.

5. Conclusions

In this work, we determined the full genome sequence of three DENV2 derived from mosquitoes caught in the wild; this is a significant contribution given the limited number of DENV mosquito-derived sequences available overall. Additionally, we have examined the sequences with the aim of identifying polymorphic sites by comparing them with sequences reported in the databases. Since most of the sequences in databases are human-derived, the conclusions cannot be definitive, but the results suggest that some SNVs are more frequent in mosquito-derived sequences and that there is some overlapp between these SNVs and the positions that were maintained as polymorphic in one of the mosquito-derived genomes. Other interesting observations were made, for instance, most of the nonsynonymous polymorphisms were found in non-structural proteins, indicating that these proteins, rather than the envelope one, may be involved in the ability of this virus to maintain fitness upon host switching.

Declaration of interest

The authors declare no conflicts of interest.

Acknowledgments

This work was funded by the National Council for Science and Technology (CONACyT) of Mexico (183448 and 89571-457481).

Appendix A

Table A.1
Characteristics of mosquito pools.

Sample	Collection date	Neighborhood of collection inside Acapulco	Type of dengue case	Number of female mosquitoes in the pool	Confirmation of DENV presence by RT-PCR
S1	07/18/2012	Morelos	Confirmed case	30	Confirmed
S2	07/26/2012	Coloso	Confirmed case	33	Confirmed
S3	10/12/2012	La Garita	Probable case	33	Confirmed
S4	10/23/2012	Jardín Mangos	Probable case	22	Confirmed
S5	10/23/2012	Jardín Mangos	Probable case	22	Confirmed
S6	11/07/2012	Sta. Rosita	Probable case	23	Not confirmed
S7	11/07/2012	Centro	Probable case	13	Not confirmed

Table A.2
Synonymous substitutions present in higher frequency in mosquito-derived sequences.

Position	Gene	G1	G2	G3	Freq in MDseq[a]	Freq in HDseq[b]	S[c] or NS[d]	AA	p-Value
288	C	✓	✓	✓	0.60	0.01	S	N	3.4×10^{-21}
432	prM			✓	0.20	0.0026	S	G	2.6×10^{-8}
489	prM			✓	0.20	0	S	I	8.2×10^{-17}
552	prM	✓	✓	✓	1	0.044	S	S	1.9×10^{-18}
702	M	✓	✓	✓	0.60	0.015	S	V	5.5×10^{-38}
726	M	✓	✓	✓	0.60	0.0026	S	L	1.1×10^{-49}
915	E	✓	✓	✓	0.60	0	S	L	1.1×10^{-50}
961	E	✓	✓	✓	1	0.013	S	L	6.7×10^{-42}
1084	E		✓		0.20	0.0026	S	L	3.6×10^{-8}
1296	E	✓	✓	✓	0.60	0.046	S	G	1.3×10^{-6}
1362	E			✓	0.20	0.0026	S	G	2.6×10^{-8}
1419	E	✓	✓	✓	0.60	0.008	S	F	7.2×10^{-25}

(*continued on next page*)

Table A.2 (*continued*)

1506	E	✓	✓	✓	0.80	0.05	S	P	3.7×10^{-15}
1593	E	✓	✓	✓	1	0.01	S	V	1.1×10^{-46}
1653	E	✓			0.20	0	S	Q	8.2×10^{-17}
1710	E		✓		0.40	0.01	S	D	7.7×10^{-11}
2067	E		✓		0.20	0	S	A	6.3×10^{-17}
2133	E	✓	✓	✓	0.60	0.04	S	F	2.3×10^{-7}
2271	E	✓			0.20	0.0026	S	V	2.6×10^{-8}
2299	E	✓			0.20	0.0026	S	L	2.6×10^{-8}
2355	NS1	✓	✓	✓	1	0.036	S	N	1.6×10^{-21}
2562	NS1			✓	0.20	0.0026	S	L	2.6×10^{-8}
2883	NS1	✓	✓	✓	1	0.067	S	A	8.1×10^{-13}
2898	NS1	✓	✓	✓	0.80	0.005	S	N	8.2×10^{-45}
3399	NS2A	✓	✓	✓	1	0.13	S	D	1.4×10^{-6}
3489	NS2A	✓	✓		0.60	0.028	S	V	3.2×10^{-10}
3711	NS2A	✓	✓	✓	0.60	0	S	L	1.1×10^{-50}
3759	NS2A	✓	✓	✓	1	0.16	S	A	2.35×10^{-19}
4029	NS2B	✓	✓	✓	0.80	0.013	S	K	2.5×10^{-33}
4170	NS2B	✓			0.20	0	S	T	8.2×10^{-17}
4257	NS2B	✓	✓	✓	1	0.01	S	L	1.1×10^{-46}
4368	NS2B		✓		0.20	0	S	P	8.2×10^{-17}
4533	NS3	✓	✓	✓	1	0.059	S	I	2.1×10^{-20}
4671	NS3	✓	✓	✓	0.80	0.0026	S	G	5.2×10^{-54}
4716	NS3		✓		0.20	0.0026	S	V	2.6×10^{-8}
4723	NS3	✓	✓	✓	0.60	0.0026	S	L	9.1×10^{-38}
4734	NS3	✓	✓	✓	0.80	0.013	S	G	1.6×10^{-29}
4749	NS3	✓	✓	✓	0.60	0.19	S	A	1×10^{-50}
4929	NS3	✓	✓	✓	0.60	0.023	S	T	4.9×10^{-12}
5181	NS3		✓		0.40	0.008	S	T	4.2×10^{-17}
5250	NS3	✓	✓	✓	0.80	0.01	S	V	2.5×10^{-33}
5370	NS3		✓		0.20	0	S	T	8.2×10^{-17}
5442	NS3	✓	✓	✓	1	0.02	S	I	5.9×10^{-32}
5481	NS3	✓			0.40	0.008	S	T	4.2×10^{-12}
5571	NS3		✓		0.20	0.0026	S	V	2.6×10^{-8}
5610	NS3	✓			0.40	0.005	S	V	3×10^{-30}
5676	NS3	✓	✓	✓	0.80	0.005	S	F	8.2×10^{-45}
5763	NS3	✓	✓	✓	0.80	0.0026	S	P	5.3×10^{-54}
5823	NS3	✓	✓	✓	0.80	0.005	S	K	8.2×10^{-45}
5847	NS3	✓	✓	✓	0.60	0	S	Y	1.1×10^{-50}
5997	NS3	✓	✓	✓	0.60	0.005	S	R	5.1×10^{-30}
6021	NS3	✓	✓	✓	0.80	0.005	S	T	8.2×10^{-45}
6126	NS3	✓	✓	✓	0.80	0.038	S	I	1.8×10^{-13}
6186	NS3	✓	✓	✓	0.80	0.0026	S	E	5.2×10^{-54}
6283	NS4A	✓	✓	✓	1	0.01	S	L	1.1×10^{-46}
6438	NS4A	✓	✓	✓	0.80	0.008	S	E	3×10^{-38}
6699	2K	✓	✓	✓	0.60	0.005	S	A	5×10^{-30}
6702	2K	✓	✓	✓	1	0.015	S	I	5.5×10^{-38}
6744	NS4B	✓	✓	✓	0.60	0.0026	S	F	9.1×10^{-38}
6747	NS4B	✓	✓	✓	0.60	0.0026	S	L	9.1×10^{-38}
6783	NS4B	✓	✓	✓	0.60	0.026	S	I	4.7×10^{-11}
7170	NS4B	✓	✓	✓	0.80	0.0026	S	V	$5{,}2 \times 10^{-54}$
7254	NS4B	✓	✓	✓	0.80	0.0026	S	I	5.2×10^{-54}
7329	NS4B	✓	✓	✓	0.80	0.005	S	P	8.2×10^{-45}
7531	NS5	✓		✓	0.40	0.01	S	L	7.7×10^{-11}
7653	NS5	✓	✓	✓	1	0.008	A	A	1.9×10^{-55}
7849	NS5	✓	✓	✓	1	0.01	S	L	1.1×10^{-46}
7926	NS5	✓		✓	0.40	0	S	S	1.1×10^{-33}
7992	NS5	✓	✓	✓	0.60	0	S	N	1.1×10^{-50}
8364	NS5	✓	✓	✓	1	0.015	S	D	5.5×10^{-38}
8391	NS5	✓	✓	✓	1	0.01	S	H	1.1×10^{-46}
8505	NS5	✓	✓	✓	0.60	0.005	S	T	5×10^{-30}
8523	NS5			✓	0.20	0	S	G	8.2×10^{-17}
8565	NS5	✓	✓	✓	0.80	0.0026	S	Q	5.2×10^{-54}
8808	NS5	✓	✓	✓	0.80	0.01	S	G	2.5×10^{-33}
9015	NS5	✓	✓	✓	0.60	0	S	R	1.1×10^{-50}
9024	NS5	✓	✓	✓	1	0.01	S	Y	1.1×10^{-46}
9225	NS5	✓	✓	✓	1	0.01	S	T	$1{,}1 \times 10^{-46}$
9255	NS5			✓	0.20	0	S	S	8.3×10^{-17}
9441	NS5	✓	✓	✓	0.60	0	S	S	1.1×10^{-50}
9459	NS5	✓	✓	✓	1	0.01	S	G	1.1×10^{-46}
9477	NS5	✓	✓	✓	0.60	0	S	K	1.1×10^{-50}
9489	NS5	✓	✓	✓	1	0.013	S	D	6.7×10^{-42}
9826	NS5	✓	✓	✓	1	0.062	NS	S > P	9.1×10^{-14}
10041	NS5		✓		0.20	0	S	R	8.2×10^{-17}
10160	NS5			✓	0.20	0	NS	A > V	8.2×10^{-17}
10175	NS5			✓	0.20	0	Stop	Stop	8.2×10^{-17}
10176	NS5	✓			0.040	0.0026	Stop	Stop	9×10^{-33}

[a] MDseq Mosquito-derived sequences

[b] HDseq Human-derived sequences
[c] S Synonymous substitutions
[d] NS Nonsynonymous substitutions

Table A.3
Low frequency polymorphisms in G3AE.

Position in the CDS	Gene	SNV	Frequency in forward reads	Frequency in reverse reads	Type pf change (S[a] or NS[b])	AA change
15	C	A > G	0.08	0.07	S	R
297	C	C > T	0.01	0.02	S	R
647	M	G > A	0.01	0.01	NS	G > E
1000	E	G > C	0.01	0.02	NS	A > P
1183	E	A > C	0.01	0.02	NS	T > P
1524	E	A > G	0.01	0.01	S	G
1622	E	A > C	0.01	0.03	NS	H > P
3067	NS1	G > C	0.01	0.02	NS	A > P
3226	NS1	A > C	0.01	0.02	NS	T > P
3714	NS2A	T > G	0.03	0.04	S	L
4180	NS2B	G > C	0.01	0.02	NS	A > P
5160	NS3	A > C	0.01	0.02	S	P
5212	NS3	G > C	0.01	0.03	NS	A > P
5668	NS3	G > C	0.01	0.02	NS	A > P
5773	NS3	A > C	0.01	0.03	NS	T > P
5827	NS3	G > T	0.01	0.01	missense	E > stop
6286	NS4A	A > C	0.01	0.02	NS	T > P
6476	NS4A	G > C	0.01	0.01	NS	G > A
6645	NS4A	A > C	0.02	0.01	NS	E > D
6682	2K	A > C	0.03	0.03	NS	T > P
6721	2K	A > C	0.03	0.03	NS	T > P
6754	NS4B	A > C	0.02	0.04	NS	T > P
6827	NS4B	G > C	0.01	0.03	NS	R > P
6853	NS4B	G > C	0.02	0.01	NS	A > P
7692	NS5	G > C	0.02	0.01	S	P
7721	NS5	G > C	0.03	0.01	NS	G > A
7749	NS5	G > C	0.01	0.04	S	G
7935	NS5	C > A	0.02	0.01	S	P
8474	NS5	T > G	0.01	0.04	NS	V > G
8695	NS5	G > C	0.01	0.02	NS	A > P
9150	NS5	A > G	0.01	0.01	S	G
9654	NS5	A > C	0.02	0.02	NS	P
9939	NS5	G > C	0.02	0.03	S	P
9946	NS5	G > A	0.01	0.01	Missense	E > stop
9979	NS5	G > C	0.02	0.01	NS	V > L

[a] S Synonymous substitution
[b] NS Nonsynonymous substitution

References

Bollback, J.P., 2006. SIMMAP: Stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics 7, 1–7. https://doi.org/10.1186/1471-2105-7-88.

Brackney, D.E., Beane, J.E., Ebel, G.D., 2009. RNAi targeting of West Nile virus in mosquito midguts promotes virus diversification. PLoS Pathog. 5, 1–9. https://doi.org/10.1371/journal.ppat.1000502.

Bushnell, B., Egan, R., Copeland, A., Foster, B., Clum, A., Sun, H., Labutti, K., Singan, V., Tritt, A., Spunde, A., Martin, J., Han, J., Scholz, M., 2014. BBMap: A Fast, Accurate, Splice-Aware Aligner. eScholarship. https://doi.org/10.1186/1471-2105-13-238.

Carrillo-Valenzo, E., Danis-Lozano, R., Velasco-Hernández, J.X., Sánchez-Burgos, G., Alpuche, C., López, I., Rosales, C., Baronti, C., de Lamballerie, X., Holmes, E.C., Ramos-Castañeda, J., 2010. Evolution of dengue virus in Mexico is characterized by frequent lineage replacement. Arch. Virol. 155, 1401–1412. https://doi.org/10.1007/s00705-010-0721-1.

Chao, D.Y., King, C.C., Wang, W.K., Chen, W.J., Wu, H.L., Chang, G.J.J., 2005. Strategically examining the full-genome of dengue virus type 3 in clinical isolates reveals its mutation spectra. Virol. J. 2 (72). https://doi.org/10.1186/1743-422X-2-72.

Chen, S., Huang, T., Zhou, Y., Han, Y., Xu, M., Gu, J., 2017. AfterQC: automatic filtering, trimming, error removing and quality control for fastq data. BMC Bioinformatics 18, 80. https://doi.org/10.1186/s12859-017-1469-3.

Díaz, F.J., Black IV, W.C., Farfán-Ale, J.A., Loroño-Pino, M.A., Olson, K.E., Beaty, B.J., 2006. Dengue virus circulation and evolution in Mexico: a phylogenetic perspective. Arch. Med. Res. 37, 760–773. https://doi.org/10.1016/j.arcmed.2006.02.004.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. https://doi.org/10.1093/molbev/mss075.

Dzul-Manzanilla, F., Martínez, N.E., Cruz-Nolasco, M., Gutiérrez-Castro, C., López-Damián, L., Ibarra-López, J., Martini-Jaimes, A., Bibiano-Marín, W., Tornez-Benitez, C., Vazquez-Prokopec, G.M., Manrique-Saide, P., 2016. Evidence of vertical transmission and co-circulation of chikungunya and dengue viruses in field populations of *Aedes aegypti* (L.) from Guerrero, Mexico. Trans. R. Soc. Trop. Med. Hyg. 110, 141–144. https://doi.org/10.1093/trstmh/trv106.

Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. https://doi.org/10.1093/nar/gkh340.

Gardella-Garcia, C.E., Perez-Ramirez, G., Navarrete-Espinosa, J., Cisneros, A., Jimenez-Rojas, F., Ramírez-Palacios, L.R., Rosado-Leon, R., Camacho-Nuez, M., Munoz, M.D.L., 2008. Specific genetic markers for detecting subtypes of dengue virus serotype-2 in isolates from the states of Oaxaca and Veracruz, Mexico. BMC Microbiol. 8, 1–16. https://doi.org/10.1186/1471-2180-8-117.

Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. https://doi.org/10.1093/molbev/msp259.

Holmes, E.C., Twiddy, S.S., 2003. The origin, emergence and evolutionary genetics of dengue virus. Infect. Genet. Evol. 3, 19–28. https://doi.org/10.1016/S1567-1348(03)00004-2.

Lambrechts, L., Lequime, S., 2016. Evolutionary dynamics of dengue virus populations within the mosquito vector. Curr. Opin. Virol. 21, 47–53. https://doi.org/10.1016/j.coviro.2016.07.013.

Lanciotti, R.S., 2003. Molecular amplification assays for the detection of flaviviruses. Adv. Virus Res. https://doi.org/10.1016/S0065-3527(03)61002-X.

Lequime, S., Fontaine, A., Ar Gouilh, M., Moltini-Conclois, I., Lambrechts, L., 2016. Genetic drift, purifying selection and vector genotype shape dengue virus intra-host genetic diversity in mosquitoes. PLoS Genet. 12, 1–24. https://doi.org/10.1371/journal.pgen.1006111.

Lequime, S., Richard, V., Cao-Lormeau, V.-M., Lambrechts, L., 2017. Full-genome dengue virus sequencing in mosquito saliva shows lack of convergent positive selection during transmission by *Aedes aegypti*. Virus Evol. 3. https://doi.org/10.1093/ve/vex031.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing

reads. EMBnet J. 17, 10. https://doi.org/10.14806/ej.17.1.200.
Okonechnikov, K., Conesa, A., García-Alcalde, F., 2015. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. https://doi.org/10.1093/bioinformatics/btv566.
Peng, Y., Leung, H.C.M., Yiu, S.M., Chin, F.Y.L., 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28, 1420–1428. https://doi.org/10.1093/bioinformatics/bts174.
Pickett, B.E., Greer, D.S., Zhang, Y., Stewart, L., Zhou, L., Sun, G., Gu, Z., Kumar, S., Zaremba, S., Larsen, C.N., Jen, W., Klem, E.B., Scheuermann, R.H., 2012. Virus pathogen Database and Analysis Resource (ViPR): a comprehensive bioinformatics Database and Analysis Resource for the Coronavirus research community. Viruses. https://doi.org/10.3390/v4113209.
Pickett, B.E., Liu, M., Sadat, E.L., Squires, R.B., Noronha, J.M., He, S., Jen, W., Zaremba, S., Gu, Z., Zhou, L., Larsen, C.N., Bosch, I., Gehrke, L., McGee, M., Klem, E.B., Scheuermann, R.H., 2013. Metadata-driven comparative analysis tool for sequences (meta-CATS): an automated process for identifying significant sequence variations that correlate with virus attributes. Virology 447, 45–51. https://doi.org/10.1016/j.virol.2013.08.021.
R Core Team, 2016. R.A. language and environment for statistical computing. R Found. Stat. Comput.https://doi.org/10.1007/978-3-540-74686-7.
Rambaut, A., Drummond, A., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using tracer 1.7. Syst. Biol. 67 (5), 901–904.
Rivera-Osorio, P., Vaughan, G., Ramírez-González, J.E., Fonseca-Coronado, S., Ruíz-Tovar, K., Cruz-Rivera, M.Y., Ruíz-Pacheco, J.A., Vázquez-Pichardo, M., Carpio-Pedroza, J.C., Cázares, F., Escobar-Gutiérrez, A., 2011. Molecular epidemiology of autochthonous dengue virus strains circulating in Mexico. J. Clin. Microbiol. 49, 3370–3374. https://doi.org/10.1128/JCM.00950-11.
Sessions, O.M., Wilm, A., Kamaraj, U.S., Choy, M.M., Chow, A., Chong, Y., Ong, X.M., Nagarajan, N., Cook, A.R., Ooi, E.E., 2015. Analysis of dengue virus genetic diversity during human and mosquito infection reveals genetic constraints. PLoS Negl. Trop. Dis. 9, 1–21. https://doi.org/10.1371/journal.pntd.0004044.
Sim, S., Aw, P.P.K., Wilm, A., Teoh, G., Hue, K.D.T., Nguyen, N.M., Nagarajan, N., Simmons, C.P., Hibberd, M.L., 2015. Tracking dengue virus intra-host genetic diversity during human-to-mosquito transmission. PLoS Negl. Trop. Dis. 9, 1–21. https://doi.org/10.1371/journal.pntd.0004052.
Suzuki, R., Shimodaira, H., 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics. https://doi.org/10.1093/bioinformatics/btl117.
Torres-Galicia, I., Cortés-Poza, D., Becker, I., 2014. Dengue en México: análisis de dos décadas Introducción. Gac. Med. Mex. 150, 122–127.
Vasilakis, N., Deardorff, E.R., Kenney, J.L., Rossi, S.L., Hanley, K.A., Weaver, S.C., 2009. Mosquitoes put the brake on arbovirus evolution: experimental evolution reveals slower mutation accumulation in mosquito than vertebrate cells. PLoS Pathog. 5 (6), e1000467. https://doi.org/10.1371/journal.ppat.1000467.
Velazquez-Salinas, L., Zarate, S., Eschbaumer, M., Pereira Lobo, F., Gladue, D.P., Arzt, J., Novella, I.S., Rodriguez, L.L., 2016. Selective factors associated with the evolution of codon usage in natural populations of arboviruses. PLoS One 11 (7), e0159943. https://doi.org/10.1371/journal.pone.0159943.
Verbist, B.M.P., Thys, K., Reumers, J., Wetzels, Y., Van Der Borght, K., Talloen, W., Aerssens, J., Clement, L., Thas, O., 2015. VirVarSeq: a low-frequency virus variant detection pipeline for Illumina sequencing using adaptive base-calling accuracy filtering. Bioinformatics. https://doi.org/10.1093/bioinformatics/btu587.
Weaver, S., Vasilakis, N., 2009. Molecular evolution of dengue viruses: contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease. Infect. Genet. Evol. 9 (4), 523–540.
Wilder-Smith, A., Byass, P., 2016. The elusive global burden of dengue. Lancet Infect. Dis. https://doi.org/10.1016/S1473-3099(16)00076-1.