AYDA SUSANA ORTIZ BAEZ

DIVERSIDADE INTRA-HOSPEDEIRO DO VÍRUS DA DENGUE TIPO 4 CIRCULANDO EM GUARUJÁ, SÃO PAULO

Dissertação apresentada ao programa de pós-graduação em Microbiologia do Instituto de Ciências Biomédicas da Universidade de São Paulo, para obtenção do Título de Mestre em Ciências.

São Paulo
2016

AYDA SUSANA ORTIZ BAEZ

INTRA-HOST GENETIC DIVERSITY OF DENGUE VIRUS TYPE 4 STRAINS FROM THE MUNICIPALITY OF GUARUJÁ, SÃO PAULO

Dissertation presented to the department of Microbiology, Graduate school of Biomedical Sciences at the University of São Paulo. In partial fulfillment of the requirements for the degree Master of Sciences.

São Paulo
2016

AYDA SUSANA ORTIZ BAEZ

DIVERSIDADE INTRA-HOSPEDEIRO DO VÍRUS DA DENGUE TIPO 4 CIRCULANDO EM GUARUJÁ, SÃO PAULO

Dissertação apresentada ao programa de pós-graduação em Microbiologia do Instituto de Ciências Biomédicas da Universidade de São Paulo, para obtenção do Título de Mestre em Ciências.

Área de concentração: Microbiologia

Orientador: Prof. Dr. Paolo Marinho de Andrade Zanotto

Versão original

São Paulo
2016

AYDA SUSANA ORTIZ BAEZ

INTRA-HOST GENETIC DIVERSITY OF DENGUE VIRUS TYPE 4 STRAINS FROM THE MUNICIPALITY OF GUARUJÁ, SÃO PAULO

Dissertation presented to the department of Microbiology, Graduate school of Biomedical Sciences at the University of São Paulo. In partial fulfillment of the requirements for the degree Master of Sciences.

Area of concentration: Microbiology

Advisor: Prof. Dr. Paolo Marinho de Andrade Zanotto

Original version

São Paulo
2016

CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
Serviço de Biblioteca e informação Biomédica
do Instituto de Ciências Biomédicas da Universidade de São Paulo

Ficha Catalográfica elaborada pelo(a) autor(a)

Ortiz Baez, Ayda Susana
Diversidade intra-hospedeiro do virus da dengue tipo 4 circulando em Guarujá, São Paulo / Ayda Susana Ortiz Baez; orientador Paolo Marinho de Andrade Zanotto. -- São Paulo, 2016.
83 p.

Dissertação (Mestrado)) -- Universidade de São Paulo, Instituto de Ciências Biomédicas.

1. Dengue . 2. Virus. 3. Intra-hospedeiro . 4. Diversidade. 5. São Paulo. I. de Andrade Zanotto, Paolo Marinho , orientador. II. Título.

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Candidato(a): Ayda Susana Ortiz Baez

Titulo da Dissertação: Diversidade Intra-hospedeiro do Vírus da Dengue Tipo 4 Circulando em Guarujá, São Paulo

Orientador: Paolo Marinho de Andrade Zanotto

A Comissão Julgadora dos trabalhos de Defesa da Dissertação de Mestrado, em sessão publica realizada a **...../...../.....**, considerou o(a) candidato(a):

( ) Aprovado(a) ( ) Reprovado(a)

Examinador(a): Assinatura: .....
Nome: .....
Instituição: .....

Examinador(a): Assinatura: .....
Nome: .....
Instituição: .....

Examinador(a): Assinatura: .....
Nome: .....
Instituição: .....

Presidente: Assinatura: .....
Nome: .....
Instituição: .....

RESEARCH ETHICS DISCLAIMER

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Cidade Universitária "Armando de Salles Oliveira"
Av. Prof. Lineu Prestes, 2415 - cep. 05508-000 São Paulo, SP - Brasil
Telefone :(55) (11) 3091.7733 telefax : (55) (11) 3091-8405
e-mail: cep@ icb.usp.br

São Paulo, 28 de agosto de 2014.

PARECER 1203/CEPSH

A Comissão de *Ética em Pesquisas em Seres Humanos* do ICB, nesta data, **APROVOU** o projeto intitulado: " *Intra-host genetic diversity of Dengue virus type 4 strains from the municipality of Guarujá São Paulo*" do Pesquisador Prof.Dr. **PAOLO MARINHO DE ANDRADE ZANOTTO e** aluna **AYDA SUSANA ORTIZ BAEZ.**

Cabe aos pesquisadores elaborar e apresentar a este Comitê, relatórios anuais (parciais e final), de acordo com a Resolução nº 466/12, item II, II.19 e II.20, do Conselho Nacional de Saúde, conforme modelo constante no site: **icb.usp.br.**

Aos pesquisadores cabe também finalizar o processo junto à Plataforma Brasil quando do encerramento deste.

O primeiro relatório deverá ser encaminhado à Secretaria deste CEP em **28.08.2015**.

Atenciosamente,

Profa. Dra. REGINA SCIVOLETTO
Coordenador da Comissão de Ética em
Pesquisas com Seres Humanos - ICB/USP

For my treasures: my mother and sister.
For my uncles.

ACKNOWLEDGMENTS

I wish to offer my gratitude to my supervisor, Dr. Paolo Marinho de Andrade Zanotto, whose expertise and support contributed to the development of this work. Thanks for the warm welcome into your research group and for giving me the opportunity to learn, grow, fail and succeed. Likewise, I would like to express my gratitude to Julian Villabona-Arenas, whose friendship, motivation, patience and expertise were fundamental to complete my research project and kept me going.

I would like to thank the LEMB members and learning partners, Anderson Pereira, Atila Iamarino, Caio Freire, Daniel Ferreira, Marielton dos Passos, Nicholas Di Paola, Norton Silva, and Shahab Zaki for the help they provided me and for all the moments we shared together. Many thanks go to Carla Braconi, I deeply appreciated her help and continuous support at different stages.

Special thanks go to Danielle Ferreira and whose help and guidance were very important over many different aspects. Likewise, A huge thank you to Victor Pimentel for his friendship, love, and helpful discussions.

I extend my gratitude to my ex-advisor Daniel R. Miranda-Esquivel for the unfailing support even from thousand of miles away, and because the many skills I develop under his mentorship were essential throughout this journey.

I would also like to thank my family for the endless love, support and encouragement through my entire life, and specially my mother who inspired me in everyday life. I want to thank my friends Adriana Maria Torres, Cinthy Jimenez, Jimmy Cabra, Los Adrianes, Luis Ladron, Maria Fernanda Carreño, Nathalia Quintero, Marco Rada, Steffania Gutierrez, Stella Rezende, and Suellen Herbster for the scientific and crazy debates, exchanges of knowledge, experiences, smiles and sometimes frustrations. Thanks also to one of my closest friends, Harold Rojas, who was always supporting and encouraging. Thanks for his help and confidence. High on this list is Jhon Ospina-Sarria, thanks for the support, for sharing with me your passion for science, and for many of my fondest memories in this amazing country.

I owe no less thanks to the professors for the assistance and willingness throughout my time as master student, and to the Department staff for all the help. An especial thanks goes to Gisele da Graça Santana for her kindness, advices and patience. I also owe thanks to Ursula de Oliveira for her tremendous help in the lab.

I would like to acknowledge and thank to the financial assistance of FAPESP (Grants 2013/25434-3), who supported my research and given me the opportunity to growth and contribute to knowledge. Finally, a special feeling of gratitude to Brazil and its people for welcoming me with open arms.

“The greatest enemy of knowledge

is not ignorance, it is the illusion of knowledge”

Stephen Hawking

RESUMO

Ortiz-Baez, AS. Diversidade intra-hospedeiro do vírus da dengue tipo 4 circulando em Guarujá, São Paulo. [Dissertação (Mestrado em Microbiologia)]. São Paulo: Instituto de Ciências Biomédicas, Universidade de São Paulo, 2016.

A caracterização da variabilidade genética intra-hospedeiro do vírus da dengue (DENV) é fundamental para a compreensão de sua evolução e dinâmica populacional no contexto atual como um importante patógeno viral humano. A diversidade viral acumulada em hospedeiros infectados influencia diretamente em outros aspectos como patogênese, transmissão e imunidade do hospedeiro. Contudo, apesar de existirem vários estudos sobre a diversidade genética intra-hospedeiro em dengue, nenhum tem sido relatado para DENV-4 até o presente momento. O ressurgimento e disseminação desse sorotipo foi associado com a sua co-circulação e o substituição dos sorotipos 1, 2 e 3 durante os recentes surtos no município do Guarujá-SP. Com base neste quadro epidemiológico, este estudo visou identificar a variação genética intra-hospedeiro do DENV-4 em amostras coletadas durante o surto de 2013, utilizando tecnologias de sequenciamento de nova geração. Portanto, nós caracterizamos a variabilidade genética de DENV-4 em diferentes níveis e as forcas evolutivas que afetam essa diversidade. Em adição, foram explorados os principais eventos na transmissão das variantes de DENV-4 identificadas. Nossos resultados revelaram uma baixa diversidade genética intra-hospedeiro para DENV-4. No entanto, mutação e pressões seletivas foram mecanismos importantes na variabilidade genética do vírus. A nível populacional, as variantes estão sujeitas á seleção natural negativa, não obstante identificamos seleção positiva atuando sob sítios específicos. Nenhuma evidência de recombinação foi detectada. Além disso, contra-intuitivamente, variantes de baixa frequência estão sendo transmitidas e contribuindo para diversidade genética do DENV-4 circulando em Guarujá. Nossos resultados fornecem novas evidências potencialmente úteis para futuros trabalhos focados em infecções mistas, escape imunológico, assim como o espalhamento e diversificação viral. Este estudo também e o primeiro trabalho a investigar a diversidade intra-hospedeiro do DENV-4.

Palavras-chave: Diversidade genética. Evolução molecular. Intra-hospedeiro. Vírus da Dengue tipo 4.

ABSTRACT

Ortiz-Baez, AS. Intra-host genetic diversity of dengue virus type 4 strains from the municipality of Guarujá, São Paulo. [Masters thesis (Microbiology)]. São Paulo: Instituto de Ciências Biomédicas, Universidade de São Paulo, 2016.

Characterizing intra-host genetic variability in dengue virus (DENV) virus is paramount for understanding its evolution and population dynamics in the context of its current status as a major human viral pathogen. The extent to which viral diversity accrues in infected host influences aspects such as pathogenesis, transmission, and host immunity. Although there are several studies about intra-host genetic diversity in dengue, so far nothing has been revealed about DENV-4. In the Guarujá municipality in the State of São Paulo, the reemergence and spread of this serotype was associated with its co-circulation and the displacement of serotypes 1, 2 and 3 during recent outbreaks. Based on this epidemiological framework, we seek to identify the intra-host genetic variation of DENV-4 strains from samples collected during the 2013 outbreak by using deep sequencing technologies. We characterized the genetic variability of DENV-4 at different levels, and the forces shaping this diversity. Likewise, we explored major transmission events among DENV-4 variants. Our results revealed a low intra-host genetic diversity for DENV-4. However, we found selective and mutational pressures contributing to genetic diversity, while recombination did not seem play an important role. We further identified purifying selection at population level but sites subject to potential diversifying selection. Additionally, we observed low frequency haplotypes being transmitted among hosts and contributing to the viral diversity of DENV-4 circulating in Guarujá. Our findings provide preliminary insights for future studies in mixed infections, drug resistance, virus variant spread and immune scape. This study is the first effort to investigate the intra-host diversity of DENV-4.

Keywords: Genetic Diversity. Molecular Evolution. Intra-host. Dengue virus type 4.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

Acronym	Meaning
aBSREL	Single likelihood ancestor counting
cDNA	Complementary DNA
CG	Guanine-cytosine content
CHIKV	Chikungunya virus
CSFV	Classical Swine Fever Virus
COA	Component analysis
DENV	Dengue virus
dN	Nonsynonymous site
dS	Synonymous site
ds-cDNA	Double strand complementary deoxyribonucleic acid
F	Female
FEL	Fixed effects likelihood
FUBAR	Fast unconstrained Bayesian approximation
GTR	Generalized time reversible model
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HDP	Highest posterior density
HIV	Human immunodeficiency virus
I	Invariants rate
K-mer	Short DNA sequence consisting of K bases
M	Male
MEME	Mixed Effects Model of Evolution
ML	Maximum Likelihood
mRNA	Messenger ribonucleic acid
NS	Non-structural
NGS	Next generation sequencing
NR	Non-reactive
PCA	Principal component analysis
PCR	Polymerase chain reaction
PV	Poliovirus
qPCR	Quantitative polymerase chain reaction
RE	Reactive
Ref-seq	Reference sequence
REL	Random effects likelihood
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
RSCU	Relative synonymous codon usage
S	Entropy
SLAC	Single likelihood ancestor counting
SNV	Single nucleotide variant
UTR	Untranslated region
WR	Weakly reactive
ZIKV	Zika virus
Γ	Gamma
ω	Omega

TABLE OF CONTENTS

1 INTRODUCTION

1.1 DENGUE VIRUS

Dengue virus (DENV) is a RNA virus, which belongs to the genus *Flavivirus* (family *Flaviviridae*) [1]. DENV structure consists of a 50 nm-diameter particle with an envelope of host-cell-derived lipids, associated with virus-derived proteins (envelope and membrane) (**Figure 1**). This structure encloses a ribonucleoprotein complex formed by the nucleocapsid protein and a single stranded, positive strand RNA of ≈11 kb in length [2]. Furthermore, the viral genome encodes a single open reading frame (ORF), flanked by untranslated regions (UTRs) at each end (95-450 nucleotides). The 5' UTR contain a cap structure while the 3' UTR lacks a poly (A) tail. Both 5'UTR and 3'UTR contain important elements critical for viral RNA replication and translation processes. Besides the UTRs, the viral RNA acts as an mRNA for the translation of ten viral proteins, starting with three structural proteins (C, prM, and E), and followed by seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) [3] (**Figure 2**).

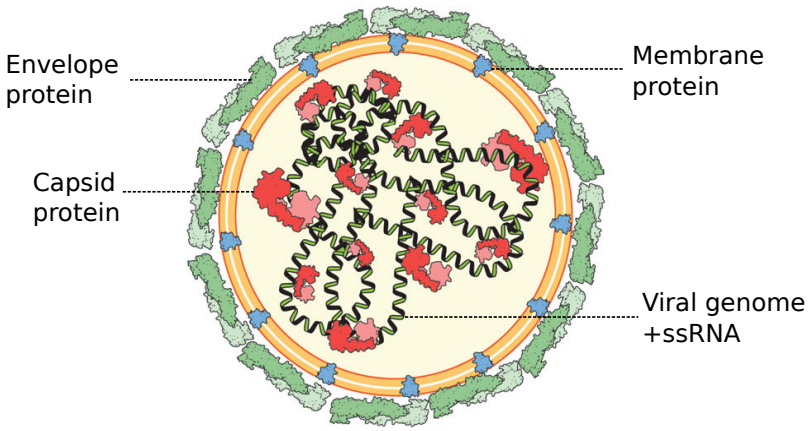


Figure 1 - Dengue virus structure. Schematic representation of dengue virus showing membrane and capsid structures organized around the viral genome. Figure adapted from: Cruz-Oliveira et al., 2015 [4].

The envelope glycoprotein (E) is considered as the major surface protein of DENV. It carries the main antigenic determinants that induces protective immunity, is essential for fusion activity and mediates receptor binding. Therefore, this protein presumably plays an important role in tropism, host range, and virulence [5]. A previous study also provided evidence of the

role of the E protein in assembly/disassembly process in flavivirus [6]. Crystal structures of the E protein in DENV have showed three structural and functional domains. Although antigenic regions have been reported for all the protein domains, it is possible to distinguish some domain-specific activities. Domain I (DI) is of particular importance for the organization of the protein structure and the DI-DII interface. Likewise, the domain II (DII) bears the fusion loop, providing an attachment point between the endosomal membrane and the viral membrane, thus triggering conformational changes in response to the acidic pH of the endosome. Finally, the domain III (DIII) participates in the recognition/ attachment to the cell surface, and harbors the receptor binding site [5,7,8] (**Figure 2**).

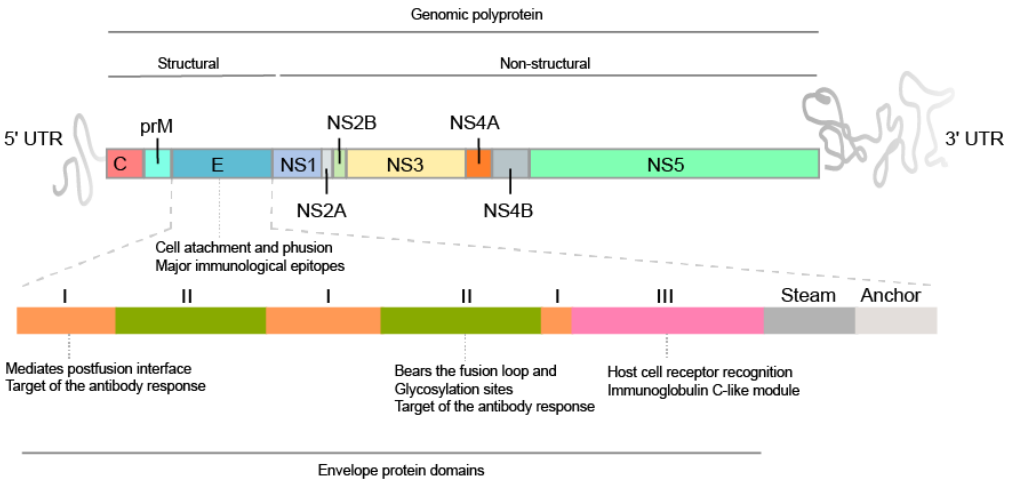


Figure 2 - Dengue virus genome. The viral genome is a positive-sense RNA that encodes ten genes flanked by the 5' and 3' untranslated regions. Genes encoding for structural proteins correspond to capside (C), pre-membrane (preM), and envelope (E). NS abbreviation represents genes encoding for non-stuctural proteins. The figure shows a zoom in the envelope protein, which is organized in three domains (DI-DIII), a stem and an anchor.

1.2 ECOLOGY AND TRANSMISSION IN BRIEF

Dengue viruses circulate in both human and wild animals (e.g. monkeys and bats). In sylvatic enzootic cycles, transmission among non-human primates is mediated by arboreal canopy-dwelling Aedes spp. [1,3]. It is hypothesized that viruses circulating among humans, emerged multiple times independently over the course of evolution, from sylvatic strains circulating in Africa or Southeast Asia forests. DENV Sylvatic strains continue to circulate in these regions, increasing the risk of newly emerging strains involved in sustained transmission chains. In human populations, DENV is responsible for the most common arthropod-borne viral disease, namelly dengue fever (DF) [9]. Mosquito vectors of genus Aedes mediate the transmission of this arbovirus in endemic/epidemic cycles. The *Aedes* (Stegomya) *aegypti* subsp. *aegypti* is the main anthropophilic vector in urban cycles

throughout the tropical and subtropical regions of the world [10,11]. The peridomestic *Aedes* (Stegomya) *albopictus* is the main responsible by the transmission in subtropics and peri-urban settings. The global distribution of both vectors defines the worldwide establishment of the four DENV serotypes, reaching more than 100 countries, and putting in risk nearly half of the world's population (**Figure 3**) [3,12].

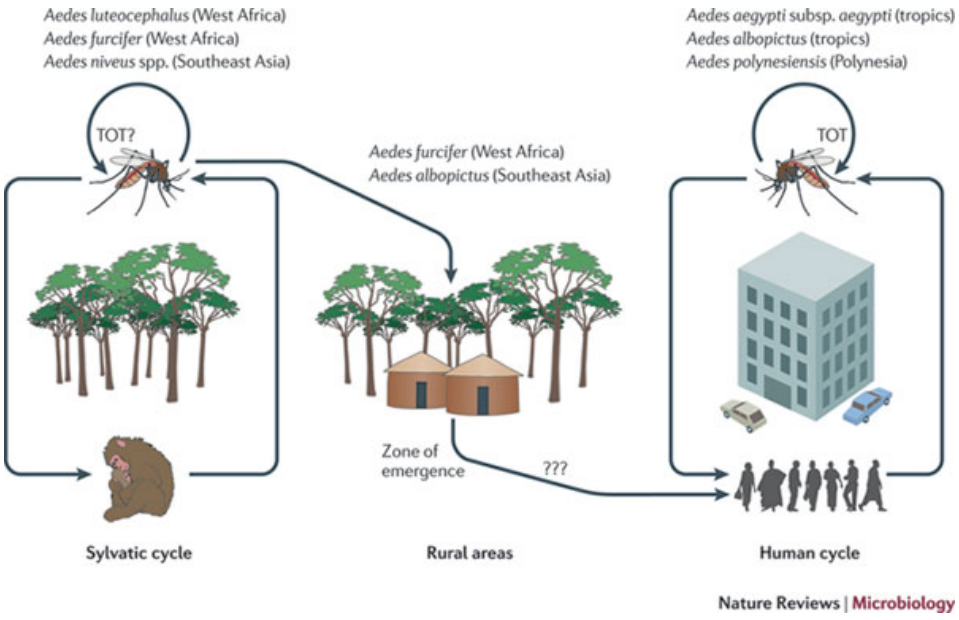


Figure 3 - Hypothetic origin and transmission of dengue virus in sylvatic, semi-urban and urban cycles. Source: Vasilakis et al., 2011 [3].

1.3 CLASSIFICATION

Four genetically and antigenically related serotypes of DENV are known (DENV 1-4) (**Figure 4**). Indeed, a recent study proposed the emergence of a fifth serotype circulating in a sylvatic transmission cycle [13]. The viral infection by any of them can result in a broad spectrum of clinical features, ranging from dengue without symptoms to overt severe manifestations, such as dengue haemorrhagic fever or shock syndrome [14]. Moreover, within each serotype it is possible to recognize a wide genetic diversity partitioned into monophyletic clusters of sequences termed genotypes, which also differ in their spatial-temporal distribution [15]. The absence of proof reading-repair activities contributes largely to the extensive genetic diversity within DENV populations [15,16].

Four major genotypes have been identified in DENV-4 [17,18]. Genotype I: Southeast Asia, genotype II: Southeast Asia and the Americas, genotype III: recent Thai strains isolated between 1997 and 2001 and, genotype IV: sylvatic DENV-4 strains isolated from sentinel monkeys in Malaysia. The genotype circulating in the Americas shows an additional spatiotemporal division, which segregates clade I (strains isolated in the Americas and Asian strains collected prior to 2000), from clade II (Asian strains collected after 2000) [1]. The

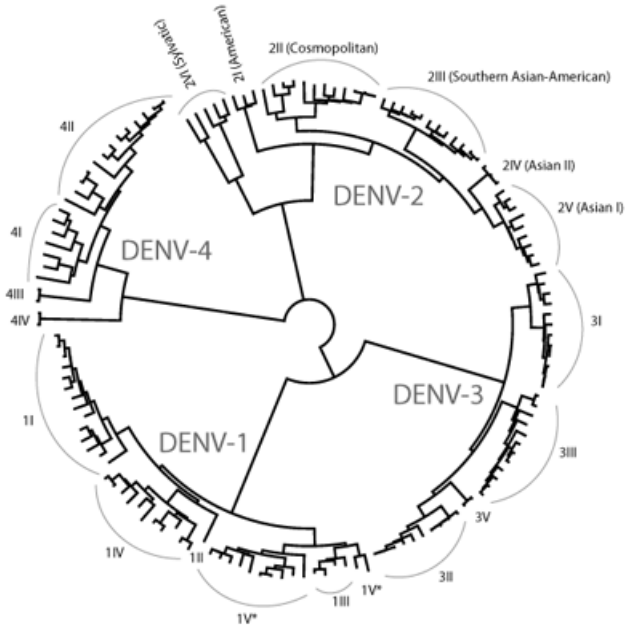


Figure 4 - Dengue virus phylogeny. The tree shows the classification of dengue viruses into serotypes and genotypes. Source: http://bioafrica2.mrc.ac.za.

Origin of DENV-4 in the Americas is currently unclear. Previous studies have suggested the introduction of this serotype (genotype II) from the Southeast Asia (French Polynesia) through the Caribbean [19,20], while other recent approach suggests the introduction through South America [21].

1.4 DENGUE IN THE WORLD AND BRAZIL

Currently, dengue is present in more than 125 countries in tropical and subtropical regions of the globe (**Figure 5**). More than 50% (3.6 billion people) of the global population is at risk, and an estimated 50 to over 200 million. DENV infections occur annually [22,23]. In Latin America, despite the implementation of an integrated management strategy for the prevention and control of dengue by the Pan American Health Organization (PAHO) [24], there have

been a progressive regional increase in the number of outbreaks and reported cases [9,25]. Co-circulation of multiple viral serotypes (hyperendemicity) and genotypes is considered the most common factor associated with the emergence of severe dengue, because of the heterologous DENV infections increase the probability of antibody-dependent enhancement (ADE), through enhancing the infection and replication of dengue virus in mononuclear cells [26–28].

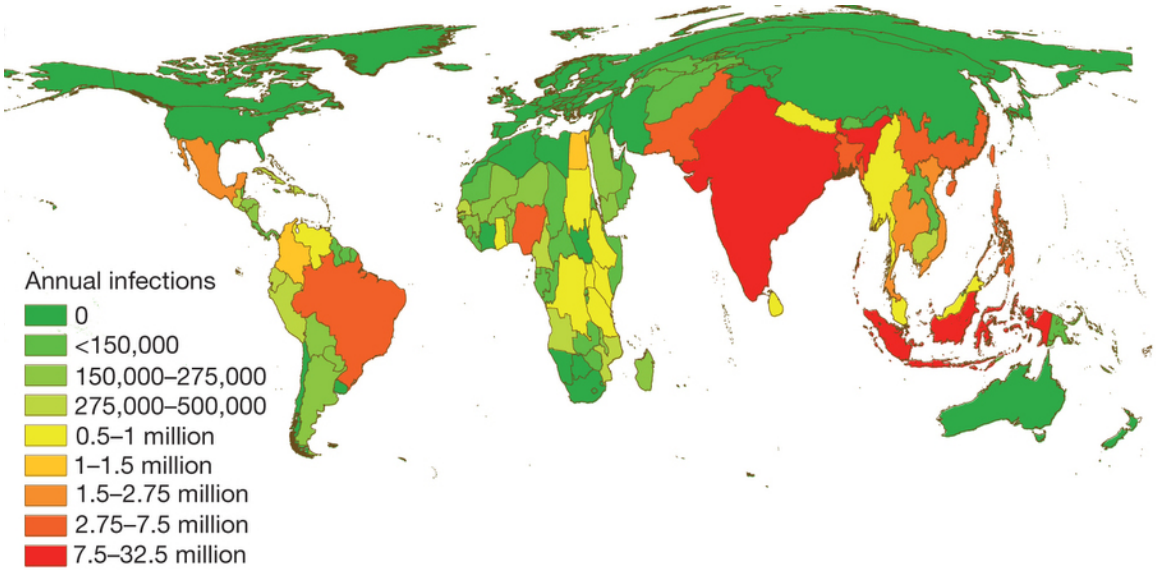


Figure 5 - Global distribution of dengue. World map representing the geographic distribution of dengue and the annual number of infections. Adapted from Bhatt et al., 2013.

Dengue has been a serious health problem in Brazil for more than 30 years [29]. The complex interplay of several factors such as hyperendemicity, deficiencies in the vector management, lack of public awareness, environmental factors, and the breakdown in the healthcare policies and procedures, imposes a tough challenge in the control, and prevention of the disease in the country. Furthermore, the costs deriving from disease-associated treatment have a major impact on financial resources and the welfare of both governments and communities [30,31]. According to the PAHO, since 2006 to 2015, Brazil has contributed with around US $ 450 million in the regional dengue program[1], while the national dengue cost estimated is US$ 310 million per year on average, excluding vector control, loss in productivity and prevention campaigns [32].

In 2013, over 1.450.000 clinical cases were reported in Brazil, 3.749 with severe manifestations, and 201 deaths [33], hence reaching the highest record of cases ever in the

country **(**data up to August 20 17, 2016**)** [34,35] **(Figure 6)**. Likewise, in 2013, 39.52% of notified cases were registered in the Southeast region of the country [36], being São Paulo among the states reporting the highest number of dengue cases. Guarujá, a coastal city located around 63 km away from São Paulo (**Figure 7**), and with a population of 290.752 inhabitants (IBGE index), was one of the worst affected areas in the southeast region during the 2013 epidemic. A recent study conducted in this municipality, revealed the co-circulation of the four serotypes in the early days of the dengue epidemic [33]. Nevertheless, a shift in serotype during the season showed that DENV-4 becomes the dominant serotype in Guarujá during the 2013 epidemic [33,37]. The spreading of DENV-4 and the population susceptibility triggered a major dengue emergency situation.

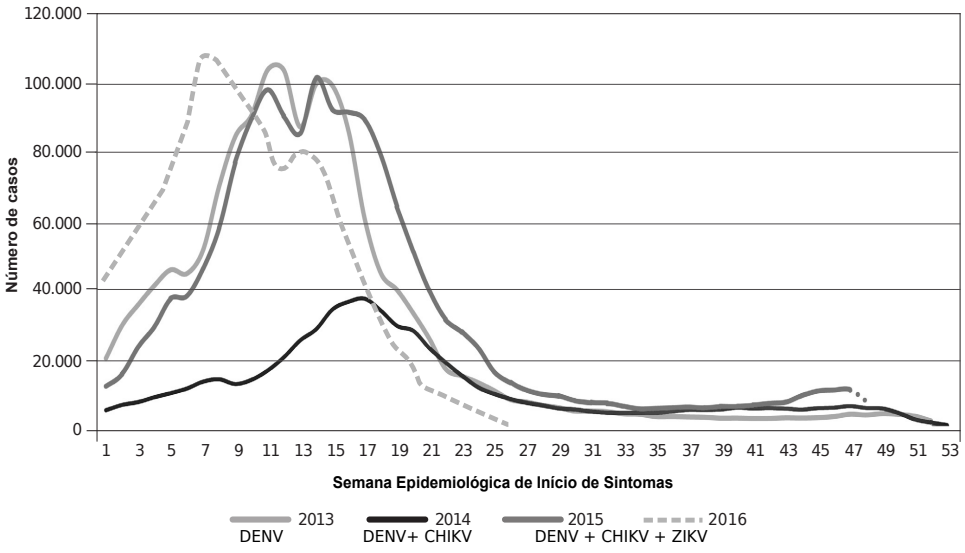


Figure 6 - Number of cases of CHIKV, DENV and ZIKV reported by epidemiological week. Adapted from Secretaria de vigilância em Saúde [34,38].

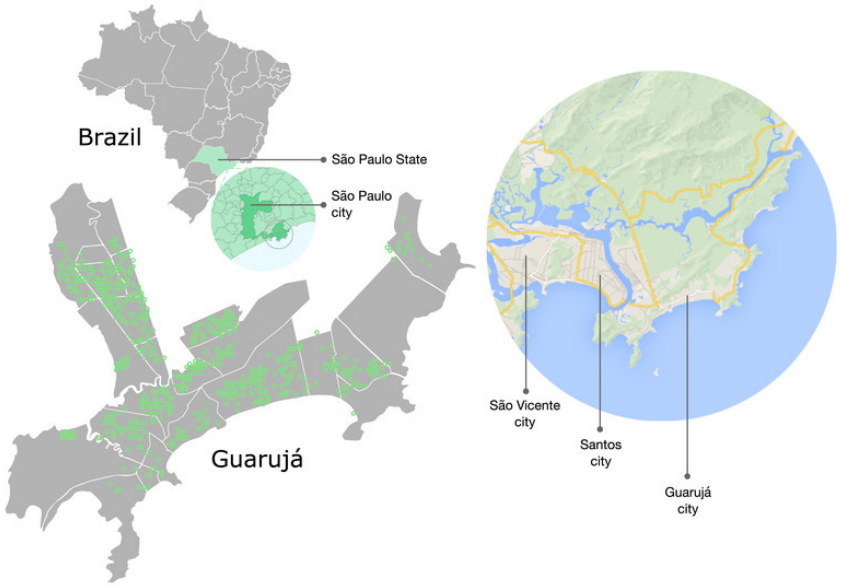


Figure 7 - Maps showing the geographic location of the municipality of Guarujá in São Paulo**.** Adapted from Villabona-Arenas et al., 2016 [37].

The early spread of DENV-4 started in 1953 in Philippines and Thailand. Then the virus spread to the French Polynesia and Southeast Asian countries, and was reported in the Americas until 1981 in several Caribbean countries and Brazil. Since then, the virus continued to expand its distribution, reaching several countries in South America and Asia (**Figure 8**) [39]. In Brazil, the reemergence of DENV-4 was documented in Boa Vista, Roraima State, 28 years after it was last detected in 1981 [40,41]. Afterwards, the virus was broadly associated with outbreaks in the Amazon, Bahia, Ceará, Goiás, Mato Grosso do Sul, Pernambuco, Pará, Piauí, Rio de Janeiro, Rio Grande do Sul, and São Paulo state. The spread of DENV-4 within Guarujá (São Paulo state) and the ensuing serotype turnover became into a dramatic epidemiological situation within a short time [33], providing an important opportunity to explore the intra-host diversity of DENV-4 into an urban population previously exposed to the remaining three serotypes.

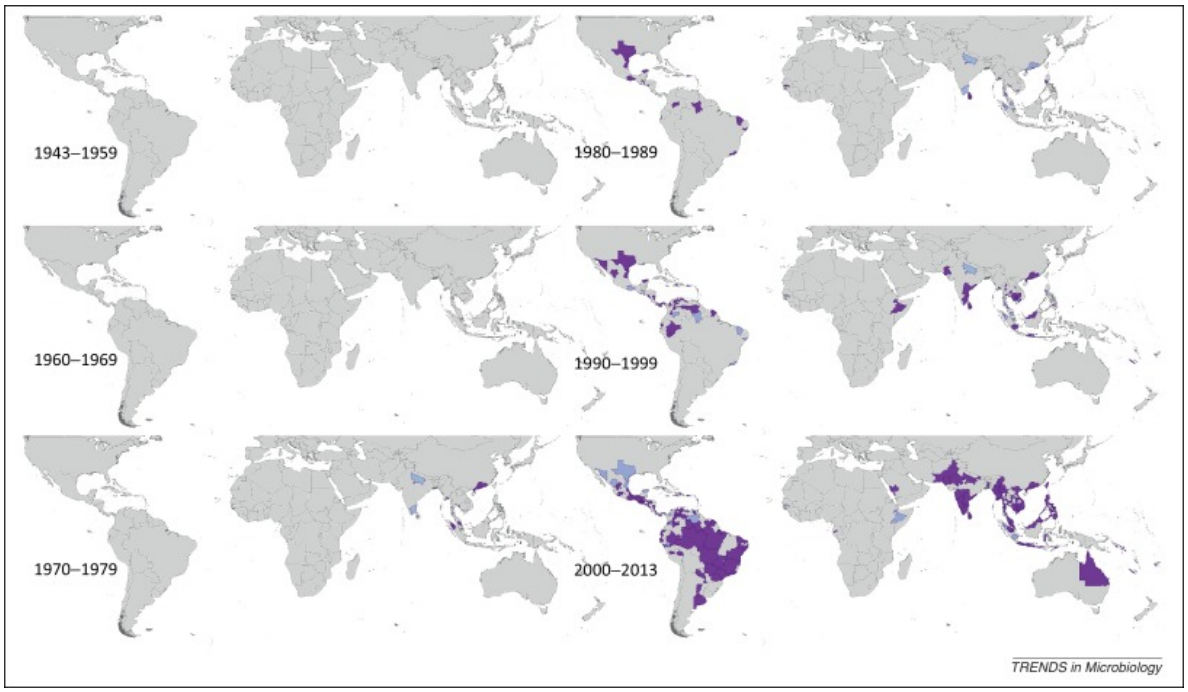


Figure 8 - Worldwide spread of DENV-4 from 1943 to 2013. Source: Messina et al., 2014 [39].

1.5 BASIC OF VIRAL INTRA-HOST DIVERSITY

Intra-host genetic variation in RNA viruses is ultimately determined by the error-prone nature of the RNA polymerase and the lack of proofreading mechanisms during replication [42–44]. As a result, an infected organism can harbor a collection of closely related variants or haplotypes which are subject to selection pressures imposed by the host environment [43,45]. (**Figure 9**). It is well known that genetic variation is the fuel for natural selection [46,47]. The occurrence of natural selection is important to promote advantageous mutations and purge deleterious changes in virus populations. Similarly, recombination, complementation and stochastic processes contribute to shape the heterogeneity in viral populations within a single host [48]. Therefore, in view of the underlying viral diversity, some variants in the pool of mutants will be potentially able to quickly adapt and confront changing environments.

Thai et al. (2011) suggested that intra-host virus diversity could be potentially increased by mixed infections, which would provide the raw material for intra-serotype recombination [49]. Moreover, mixed infections can be caused by different combinations of serotypes or

genotypes [50–53]. Importantly, hyperendemic transmission in endemic countries have increased the likelihood of mixed infections with different serotypes [54,55].

Heterogeneity in RNA virus populations can be accurately characterized using Next-Generation Sequencing (NGS) [56–59]. In general, NGS provides an opportunity to study viral diversity beyond the consensus sequence with cost-effective options, high throughput and resolution [60]. Moreover, the development of these technologies has showed a great potential for the study of re-emerging RNA viruses with sustained transmission cycles in humans, like dengue (DENV) [56,61–63]. For example, using the resolution enabled by next generation sequencing (NGS) technologies, a recent study explored the genetic diversity of DENV populations during human and mosquito infections [64]. Likewise, Lequime et al., [62] showed the evolutionary forces acting on DENV populations in the mosquito vector. The revolution of NGS in many fields of virology suggests the potential of these tools to investigate and unravel a wide range of questions in clinical virology.

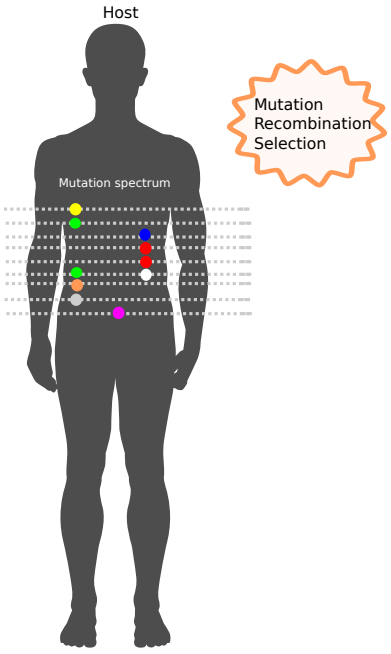


Figure 9 - Schematic representation of intra-host virus diversity and the most important factors shaping diversity in RNA viruses.

Understanding the intra-host genetic diversity of viruses is of critical importance to assess key aspects of disease pathogenesis, transmission, surveillance, and viral evolution in infected hosts. Despite, viral genetic diversity defines many evolutionary and epidemiological aspects

in DENV [65], there are not studies on the extent of the intra-host diversity of DENV-4. Related studies based on the remaining serotypes have focused on assessing the relationship between intra-host genetic variability and different issues such as disease severity [49,56,66,67], viral emergence [68], host's factors [69], clade identity [66], inter-host transmission [56,66,69], and selective pressures acting on viral populations [70–72]. Nevertheless, an important gap remains in our understanding of several aspects of DENV-4 evolution.

This work is part of an effort to assess the composition of DENV-4 populations at the intra-host level and improve our understanding about how this virus evolves and contributes to the extent of genetic diversity of DENV in Guarujá, which is ultimately a good indication of the viruses circulating near the largest metropolis of Brazil.

2 OBJECTIVES

The overall research goal of this study is to characterize the intra-host genetic diversity of dengue virus type 4 strains isolated during the 2013 outbreak in the municipality of Guarujá, São Paulo. The specific objectives are:

1) To identify the intra-host dengue virus variants using deep sequencing technologies to infer the evolutionary history of dengue lineages;

2) To evidence adaptive or purifying evolution over the virus sequences;

3) To assess the occurrence of mixed infections and its contribution to intra-host viral diversity and DENV recombination.

3 METHODS

3.1 PHYLOGEOGRAPHIC ANALYSIS

A total of 118 envelope sequences belonging to DENV-4 were obtained from the GenBank database. To construct the dataset, we included representative sequences from different localities of Brazil and other Latin American countries. We also included some sequences isolated in Asia in order to better document the introduction of DENV-4 in the Americas and its spreads to Brazil. All sequences were aligned using MUSCLE v3.8.31 [73]. Likewise, we further dereplicated the dataset using the function UCLUST implemented in USEARCH v7 [74]. Sequences were analyzed for recombination using RDP [75].

We assessed the phylogenetic content present in our dataset with TREE-PUZZLE v5.2 [76] using the likelihood-mapping algorithm [77]. Moreover, we explored the temporal structure, data integrity, and best-fit' root position for our data set with TempEst [78].

Spatial and temporal information were used in a Bayesian phylogenetic framework as implemented in BEAST v1.8.3 [79]. Analyses were performed under the GTR+I+Γ_4 nucleotide substitution model. This model was selected with the Akaike information criterion in jModelTest v2 [80]. The demographic history was inferred using the Bayesian skyride coalescent model [81] under a relaxed (uncorrelated log-normal) molecular clock model.

Furthermore, a non-reversible discrete phylogeography model was evaluated by implementing Bayesian stochastic search variable selection (BSSVS), with each locality used as a discrete state. Genealogy and parameters were estimated running MCMC chains for 200 millions states, and sampling every 200.000 states with 10% burn-in. Convergence and the Skyline-plot were examined using Tracer v1.6.0 [82]. Likewise, the maximum clade credibility tree was visualized and edited by using FigTree v1.4.2 [83].

3.2 VIRAL SAMPLES

In this study we used samples collected in 2013 during a previous work carried out in the municipality of Guarujá, São Paulo [33] (**Figure 7**). The samples corresponded to acute-phase

sera of clinically suspected dengue individuals. Infections were diagnosed using a qualitative serological test for dengue (**Table 1**). After collected, all the samples were stored at -80 °C.

Table 1 - Characteristics of samples included in this study.

Sample	Code	Collection date	NS1	IgM	IgG	Stage	Sex	Age	Locality
GU011	11-934298	2013/01/02	RE	NR	NR	P	M	37	Enseada
GU128	11-945877	2013/02/11	RE	NR	NR	P	F	40	NA
GU145	11-945688	2013/02/14	RE	NR	NR	P	F	17	Enseada
GU208	11-946520	2013/02/22	RE	WR	NR	P	F	57	Enseada
GU266	11-941507	2013/02/25	RE	NR	NR	P	M	50	Morrinhos
GU269	2-484595	2013/02/25	NR	RE	RE	S	M	53	Sta. Rosa
GU376	11-949038	2013/03/04	RE	NR	NR	P	M	11	NA
GU465	25-20876	2013/03/09	RE	NR	NR	P	M	14	C. Atlantica
GU530	11-942286	2013/03/13	RE	NR	WR	S	M	27	NA
GU544	11-953786	2013/03/13	RE	NR	NR	P	M	15	Enseada
GU568	11-953810	2013/03/14	RE	NR	NR	P	F	37	Mar e Céu
GU630	11-952974	2013/03/16	RE	NR	NR	P	M	33	B. Esperança
GU689	11-957589	2013/03/19	RE	NR	WR	S	F	NA	NA
GU727	11-957315	2013/03/19	RE	NR	NR	P	M	41	Perequé
GU792	11-958907	2013/03/21	RE	NR	WR	P	M	76	Pae-Cará
GU1128	11-977235	2013/04/22	RE	NR	NR	P	F	26	J. Progresso

P: Primary infection; S: Secondary infection

3.3 SAMPLE PREPARATION

Viral RNA was extracted with the QIAmp viral RNA mini kit (Qiagen,Venlo, Limburg, Netherlands) following manufacturer's instructions. The RNA extracted was purified using RNA Clean and Concentration kit (Zymo Research). Additionally, the genetic material was quantified with Qubit fluorometer and checked for quality with NanoDrop spectrophotometer. The number of amplifiable molecules present in the serum was quantified by qPCR to estimate viremia.

To confirm the presence of dengue virus in the samples, we performed a PCR using the SuperScript One-Step RT-PCR with Platinum taq (Invitrogen) and generic primers (**Table 2**) for the four dengue serotypes. The 511-bp products were purified using ExoSAP-IT (Affymetrix), and visualized in 1% agarose gel. In addition, some bands were extracted and sequenced by Sanger with ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems).

3.4 LIBRARY PREPARATION

Purified RNA was aliquoted for the construction of sequencing libraries using the Nextera XT DNA library preparation kit (Illumina). For the library construction, 10 µl of the purified RNA were reverse transcribed in two reactions using SuperScript® III Reverse Transcriptase (Invitrogen), a mix of random hexamers and primers specific for dengue, and RNAse inhibitor to avoid RNA degradation (RNAse OUT, Invitrogen). For double stranded DNA Synthesis, the RNA was cleavaged from the RNA:DNA hybrid with Rnase H (Invitrogen) according to the manufacturer's instructions. Next, the single stranded DNA was used as a template for double stranded DNA Synthesis using DNA Polymerase I, Large (Klenow) Fragment (Invitrogen) and T4 DNA ligase (Promega). The mix was incubated sequentially at 16 °C during 2 h and at 70 °C for 10 min.

For amplicon sequencing, the complementary DNA (cDNA) was used as template to generate the PCR products. The PCR was carried out using primers targeting the complete viral envelope gene (**Table 2**) along with the Phusion high-fidelity DNA polymerase (New England BioLabs). Finally, the amplicons and dsDNA samples were purified with DNA Clean and Concentrator kit (Zymo Reseach). The final concentration of the DNA was estimated with Qubit dsDNA HS Assay Kit Qubit fluorometer (Life Technologies), and the quality was checked with NanoDrop spectrophotometer.

Table 2 - List of primers used in different stages of this study.

Primer	Sequence	Reaction	Author
EHD1[a]	5 '-TCAATATGCTGAAACGCGCGAG AAACCG-3'	Diagnostic (DENV)	Lanciotti et al. 1992
EHD2[b]	5’-TTGCACCAACAGTCAATGTCTTC AGGTTC-3'	Diagnostic (DENV)	Lanciotti et al. 1992
D4-U486[a]	5'- CACGTATAAATGCCCCCTACTGGTC-3'	Envelope (DENV-4)	Bennet et al. 2013
D4-L2679[b]	5'- CCTTCACATCCCCAGCCACTACAGT-3'	Envelope (DENV-4)	Bennet et al. 2013

[a] Forward primer; [b] Reverse primer

Double stranded DNA from PCR amplicons and cDNA samples was used as the starting material to generate the Illumina library using NEXTERA XT kit, following manufacturer’s guidelines. All the samples were diluted to 0.2 ng/µL and 5µL/sample (1 ng total) were used

as input for the library preparation. Briefly, the DNA was fragmented and tagged using enzymatic treatment to add the adaptors. Subsequently, indexes and primers were added through an amplification step by PCR. Next, the library DNA was size selected and purified using AMPure XP beads (Beckman Coulter).

The resulting libraries, were diluted 1:1000 and quantified in triplicate data points by qPCR using KAPA SYBR FAST qPCR Kit (KAPA Biosystems) and DNA standards (six 10-fold dilutions). The average size distribution and quality of the libraries were checked using an Agilent High Sensitivity DNA kit (Agilent Technologies). The concentration of each library was calculated by performing a size adjustment using the following formula (**Equation 1**):

Equation 1.

$$\textbf{\textit{Library concentration}} = \text{average concentration x } \left(\frac{452}{\text{average fragment length}}\right) \text{x } 1000$$

The averages for the replicate data points were calculated considering at least two of the three replicates. Once we estimated the concentration for each library, all the samples were normalized based on the lowest concentration, in order to have the same representation into the pool. Next, the qPCR was repeated. Finally, equal volumes of normalized libraries were put together to prepare the pool and the resulting library was diluted in hybridization buffer prior to sequencing with MiSeq Regent Kit v2 500-cycles (Illumina, USA).

3.5 NEXT-GENERATION SEQUENCING DATA ANALYSIS

3.5.1 NGS data preprocessing

To avoid mismatches in the assembly, quality assessment was performed using FastQC [85]. From raw sequencing data, adapter sequences were removed. Then, reads were trimmed and those with low quality were discarded. Quality control checks were done after each preprocessing step (**Figure 10**). Reads filtered were used in downstream analyses.

3.5.2 NGS data processing

Filtered reads were de novo assembled into contigs from SPAdes v3.5.0 genome assembler [86] and with different k-mer values as input. From the contigs built, we selected the biggest contig as query sequence against BLAST database [87]. After confirming the taxonomic identity of our query sequence, we used it as reference to align the preprocessed reads using Bowtie v2.0.0-beta2 [88]. BAM files were filtered, sorted and duplicates were eliminated. Additionally, we mapped the reads against reference sequences from GenBank belonging to the four dengue serotypes. Amplicon-based NGS data was mapped to reference sequences generated using Sanger sequencing.

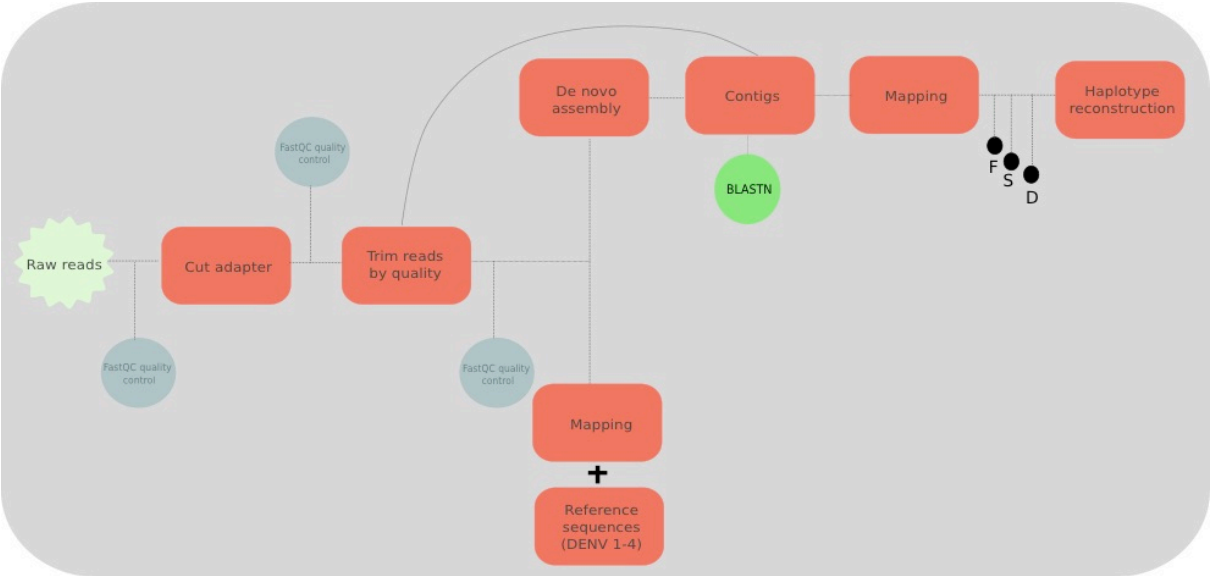


Figure 10 - Next generation sequencing workflow for assembly and haplotype reconstruction of dengue sequences. F: filter, S: sort, D: dereplicate.

3.6 SNVs DETECTION

We investigated the DENV-4 intra-host diversity present in 16 dengue patients. We were able of detecting genetic variation in the envelope (E) gene in 13 out of 16 samples, and changes through the whole-genome in the remaining three samples.

To adjust base-call error rates and improve the accuracy of variant calls, we recalibrated individual base quality scores using GATK (BQSR algorithm) [89], with default options.

After quality calibration step, we called single nucleotide variants (SNVs) with LoFreq v2.1.2 [90]. In order to reduce bias and filter variants, we defined some parameters in LoFreq, such as quality threshold (-Q 25), and minimum coverage (--min-cov 100). SNVs present in all the samples were mapped using the respective Sanger sequence as reference.

3.7 HAPLOTYPE RECONSTRUCTION

Haplotypes with frequencies were inferred with QuasiRecomb software [91]. Briefly, QuasiRecomb implements a probabilistic model based on an HMM which assume that the underlying diversity is generated by a few predicted generators through mutation and recombination. To estimate haplotype distribution, we took quality scores into account based on PHRED scores per base, and only haplotypes at a frequency > 1% were selected in order to reduce false-positive variants induced by experimental errors. Recombination, was tested with SplitsTree4 [92] and RDP [93].

3.8 TRANSMISSION NETWORK

In order to identify identical haplotypes at different points in time and estimate potential links among our samples in a transmission chain, we reconstructed a transmission network with haplotype sequences. The analysis was conducted in the R package Outbreaker [94].

3.9 GENETIC STRUCTURE

In order to explore the genetic-like pattern present in our samples, we performed a principal analysis component (PCA) by using the package outbreaker [94] for the R software [95]. In addition, we employed Neighbour-Joining tree and clustering approaches to identify major genetic groups in DENV-4 population.

3.10 CODON USAGE DEVIATION

To investigate if a codon is being used more or less frequently than expected in DENV-4 populations, we calculate the Relative Synonymous Codon Usage (RSCU) (**Equation 2**). This index takes advantage of the number of times a codon would be observed in absence of any codon usage bias to calculate the deviation in the codon usage. In order to reduce data

dimensionality and identify major trends, we employed a correspondence analysis (COA) of RSCU values. We further assessed the base composition by estimating the average content of GC at the first and second position (GC_{12}), versus the content at the third codon position (GC_3) in a neutrality plot. All the analyses were conducted using the software package CodonW v1.4.2 [96].

Equation 2.

$$\boldsymbol{RSCUx} = \frac{\text{Frequency of codonx}}{\text{Expected frequency of codon x if codon usage was uniform}}$$

3.11 SELECTION PRESSURES

Datasets were constructed using selected envelope and genome haplotypes (cutoff: frequency >1%). In order to identify inter-clade variation in substitution rates were included other sequences from Brazil and Latin America available from the GenBank database. Sequences were aligned with Muscle [73] using default parameters. To test evidence of recombination and the presence of chimeric sequences within each dataset, we used USEARCH v8.1.1861 [74,97], and RDP software [75]. We evaluated selection pressures acting on a site-by site basis by using four codon approaches: SLAC (counting substitutions), FEL (fixed effects), REL (random effects) and FUBAR (hierarchical Bayesian method). We included the MEME method to detect selection at individual sites following the developer recommendation. In addition, lineage-specific changes were tested using the aBSREL approach. All the analyses were performed with HyPhy software [98]and the Datamonkey web server (available at http://www.datamonkey.org/). Global omega (ω) or β/α, was also estimated. Phylogenetic trees were reconstructed using maximum likelihood criterion implemented in PhyML v3.0 [99]. For all datasets, the best nucleotide model was determined by jModelTest v2 [80].

3.12 DIVERSITY METRICS

Sequences were compared by using distance methods implemented in Mega v6.0 [100]. For this purpose, we estimated the proportion of nucleotide and amino acid differences. Furthermore, mean pair-wise distances were computed by sample and by month of isolation.

Global dN/dS ratios and Tajima's *D* statistic [101] were calculated with DnaSP v5.10.1 [102]. To evaluate the distribution of polymorphic sites along the alignments, we used 1000 replicates in a Monte Carlo test as implemented in the R package adegenet 2.0.0 [103].

4 RESULTS

4.1 BAYESIAN PHYLOGEOGRAPHY

Before attempting to document viral diversity within host, we reconstructed the spatiotemporal transmission pathway of DENV-4 (Genotype II) (**Figure S1**), in order to gain insight into the origin of strains circulating in Guarujá, São Paulo during the 2013 epidemic, and investigate the phylogenetic relationships with strains isolated from different Brazilian localities and the Americas.

The likelihood-mapping analysis suggested a high-moderated content of phylogenetic signal present in our dataset. Approximately, 71% of the sequences were placed into the distal areas and evolve on a completely resolved tree. No evidence of recombination was detected. Bayesian analysis was performed to reconstruct the evolutionary rate and spatial dynamics. Most parameters estimates showed ESS values >200. The overall evolutionary rate was 9.1827E-4 $\times 10^{-4}$ substitutions/site/year (95% HPD: 7.8057E-4 – 1.0847E-3), and the date of the most common ancestor (MRCA) was estimated to be 1972 (95% HPD: 1969–1975).

Our results revealed that the virus circulating in Guarujá corresponded to the lineage (genotype II) most widely distributed throughout Brazil. Phylogeographic analysis corroborated the introduction of the virus in the Americas through Puerto Rico and rapidly spread to others countries in the region. We recognized two main routes for genotype II, represented by well-supported monophyletic groups. The most recent event was an introduction from the Caribbean (Martinique) into the Para state, whereas the second event involved a transmission route throughout Venezuela, Colombia and Brazil. Finally, in Brazil the virus was first detected in Roraima state by 2005 and it was responsible for the spread into the country before emerging in Guarujá between 2010-2011 (**Figure 11**).

Additionally, we found sequences from Guarujá seem to diverge into two major clades, one consisting only of strains from Guarujá and the other comprising a strain from Guarujá along with strains from São José do Rio Preto and Mato Grosso state, thereby revealing a plausible route of introduction into this municipality.

Bayesian skyplot (BSP) for DENV-4 described a gradual increase in genetic diversity through time with a peak in the 90s, coinciding with the expansion of the genotype in the Americas (**Figure 12**). Nevertheless, we observed an abrupt downfall just prior to the extended transmission of the virus in the Brazil, followed by a moderate increase in DENV-4, probably as a result of the spread of the virus in the country.

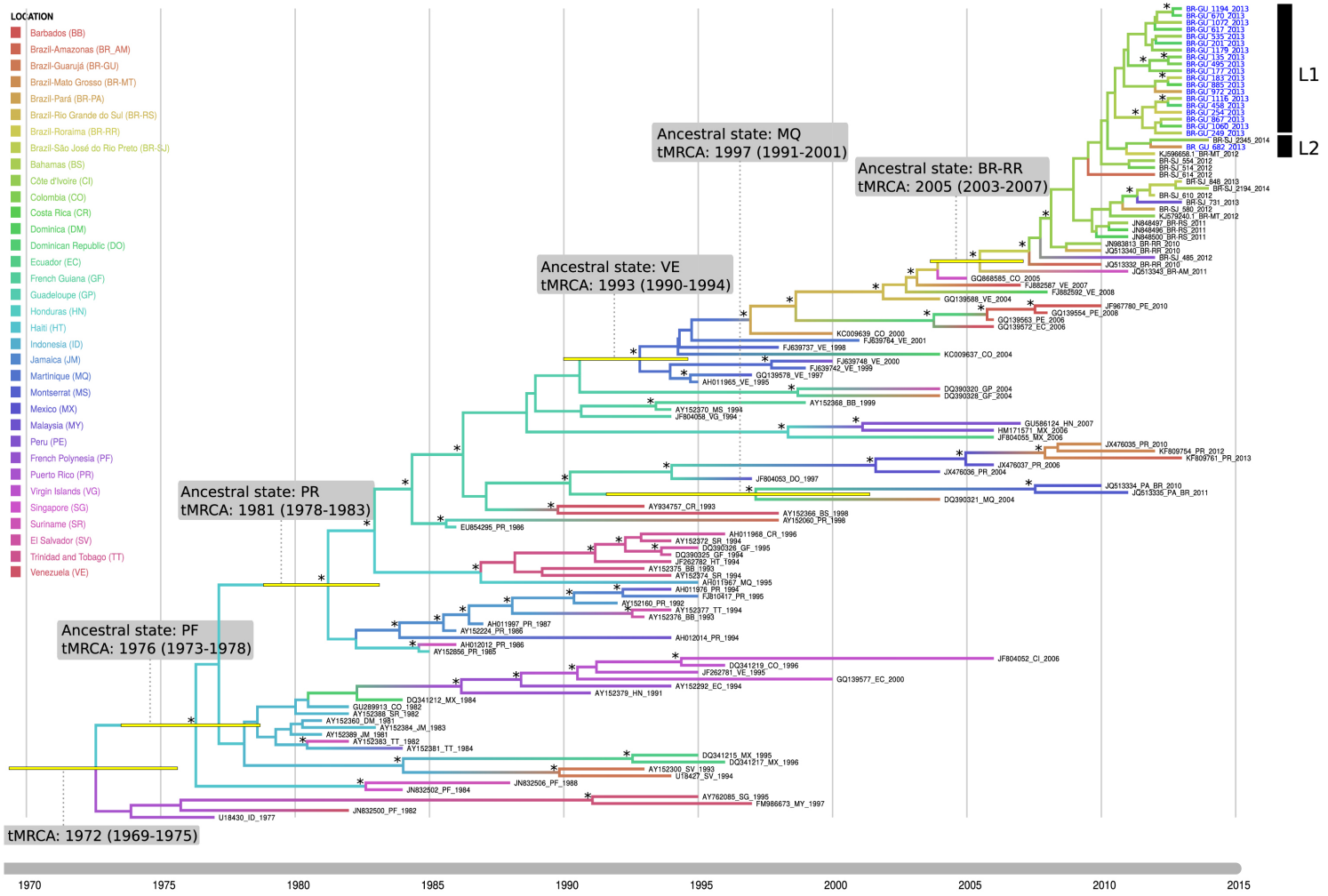


Figure 11- Bayesian phylogeography analysis of DENV-4 (genotype II). Locations are represented by color according to the legend on the left. The times of introduction are indicated in gray boxes. Asterisks represent well-supported nodes with posterior probabilities > 0.70. The timescale is depicted at the bottom on the x-axis. Lineages grouping sequences from Guarujá are represented by the L1 and L2 bars.

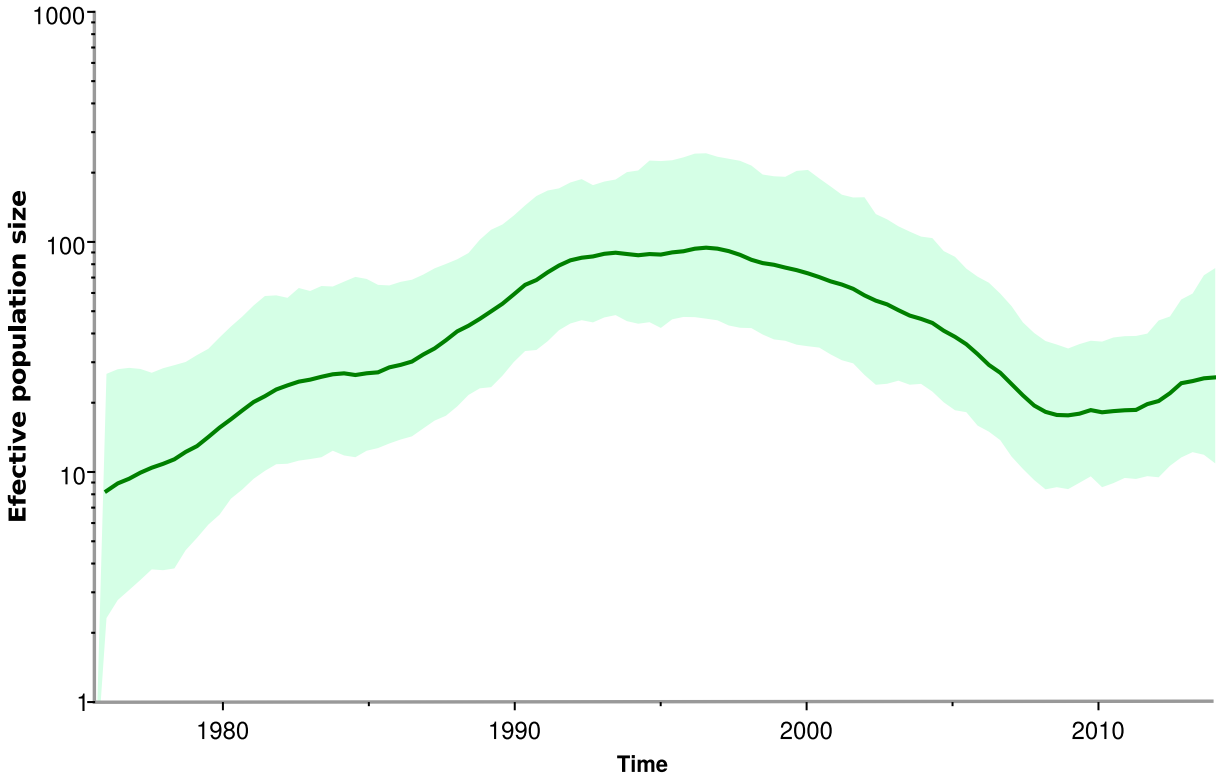


Figure 12. Bayesian skyride plot showing the effective population size trajectory of DENV-4 (genotype II) over time. The solid line represents the posterior median estimate and the shaded area shows the 95% HDP intervals.

4.2 LIBRARY CHARACTERIZATION

Because the envelope coding region is responsible for cell receptor binding and is the main target of neutralizing antibodies, amplicon based analysis of E gene sequences is a good strategy to detect relevant biological variants and study the genetic diversity of dengue virus populations [104,105]. In this study, we sequenced 16 samples (**Table 3**) corresponding to amplicon and cDNA libraries. In particular, gene amplicons showed an efficient representation into the library preparation and a better sequencing performance than cDNA samples (**Table 4**, **Figure 13**, **Figure 14**).

Despite the limited number of effective samples corresponding to ds-cDNA libraries (**Table S1**), we were able to obtain some genomes by sequencing ds-cDNA directly (**Table 4**). We observed the method is potentially efficient, however, we faced with some difficulties such as limited quantity, viral RNA degradation, and ribosomal RNA (rRNA) contamination, which biased the resolution of our results.

Table 3 - Processed samples available for library preparation and sequencing.

Sample	Number of extractions	Qubit (ng/µL)	Nanodrop (ng/µL)	Copies/ µL (dengue)	Source	One-Step
GU011	3	753,5	49,2	NA	Cultivated	Positive
GU128	12	1079,1	70,4	2.91 E+07	Serum	Positive
GU145	10	254,1	27,4	3.41 E+06	Serum	Positive
GU208	11	914	62,7	NA	Serum	Positive
GU266	6	305	21,5	NA	Serum	Positive
GU269	13	934	61	NA	Serum	Positive
GU376	12	677	45,8	NA	Serum	Positive
GU465	12	732	48,2	NA	Serum	Positive
GU530	8	353,2	26,6	NA	Serum	Positive
GU544	12	1016	60,8	NA	Serum	Positive
GU568	10	442,4	43	7.38 E+05	Serum	Positive
GU630	12	1130	80,2	NA	Serum	Positive
GU689	10	311,9	29,2	NA	Serum	Positive
GU727	12	631,9	45,1	NA	Serum	Positive
GU792	14	693.2	53,6	1.33 E+05	Serum	Positive
GU1128	12	86,3	34.7	NA	Serum	Positive

Even though we recovered a relatively low coverage for most ds-cDNA samples (**Table 4**), we got a high quality control during library preparation**.** All the sample concentrations were estimated between 20 and 0.2 pM (standards 1 and 3, respectively), further the libraries were free of primer or adaptor dimer contamination (**Figure S2**). To get an accurate quantification of the individual samples and the pool, we normalized the concentration using the average size distribution of the fragments (**Figure S3**). Despite we recovered some small DNA fragments, the peak ranged into the expected size. Likewise, this result was consistent across most samples included in the library preparation.

Table 4 - Summary of read composition obtained after Illumina sequencing.

Library	Sample	N° of reads	Avg. deep	% Trimmed	Mapped reads (%)	Non-mapped reads (%)	Avg. PHRED score
ds-cDNA	GU11	653253	455.54	98.23	5.04	94.96	>=30
ds-cDNA	GU128	787164	236.52	97.64	2.29	97.71	>=30
ds-cDNA	GU145	620067	151.95	97.28	2	98	>=30
Amplicon	GU208	148218	2,518.10	99.89	98.9	1.1	>=30
Amplicon	GU266	286406	4,715.61	99.75	98.99	1.01	>=30
Amplicon	GU269	163128	2,696.28	99.76	99.21	0.79	>=30
Amplicon	GU376	292334	4643.46	99.7	96.51	3.49	>=30
Amplicon	GU465	201241	2804.90	99.40	98.79	1.21	>=30
Amplicon	GU530	275941	4361.96	99.64	99.07	0.93	>=30
Amplicon	GU544	355387	5039.89	99.31	98.74	1.26	>=30
Amplicon	GU568	356110	5709.06	99.75	99.40	0.60	>=30
Amplicon	GU630	347022	5413.09	99.66	99.12	0.88	>=30
Amplicon	GU689	39486	630.05	99.96	96.92	3.08	>=30
Amplicon	GU727	433252	5920.16	99.25	98.75	1.25	>=30
Amplicon	GU792	338030	5460.18	99.7	99.14	0.86	>=30
Amplicon	GU1128	258538	4168.82	99.71	99.24	0.76	>=30

Based on the sequencing data (**Figure S4**), we got three samples nearly fully assembled (i.e. over 10.000 bp). In all the cases, the coding region was full-length reconstructed whereas untranslated regions were partially reconstructed. Hence, these sequences would represent the first genomes of the 2013 dengue outbreak (**Figure 13**, **Figure 15**). Moreover, we observed a high level of host RNA contamination (e.g. mRNA and rRNA) present in our libraries. Nonetheless, we got an average coverage >150 in the assembled genomes (**Table 4**). Despite the low proportion of mapped reads compared with the total reads, we identified two serotypes (**Figure 15**). Reads mapping against references sequences of DENV-1 and DENV-4 within the same patient suggested co-infection or superinfection events during the outbreak (**Table S2**).

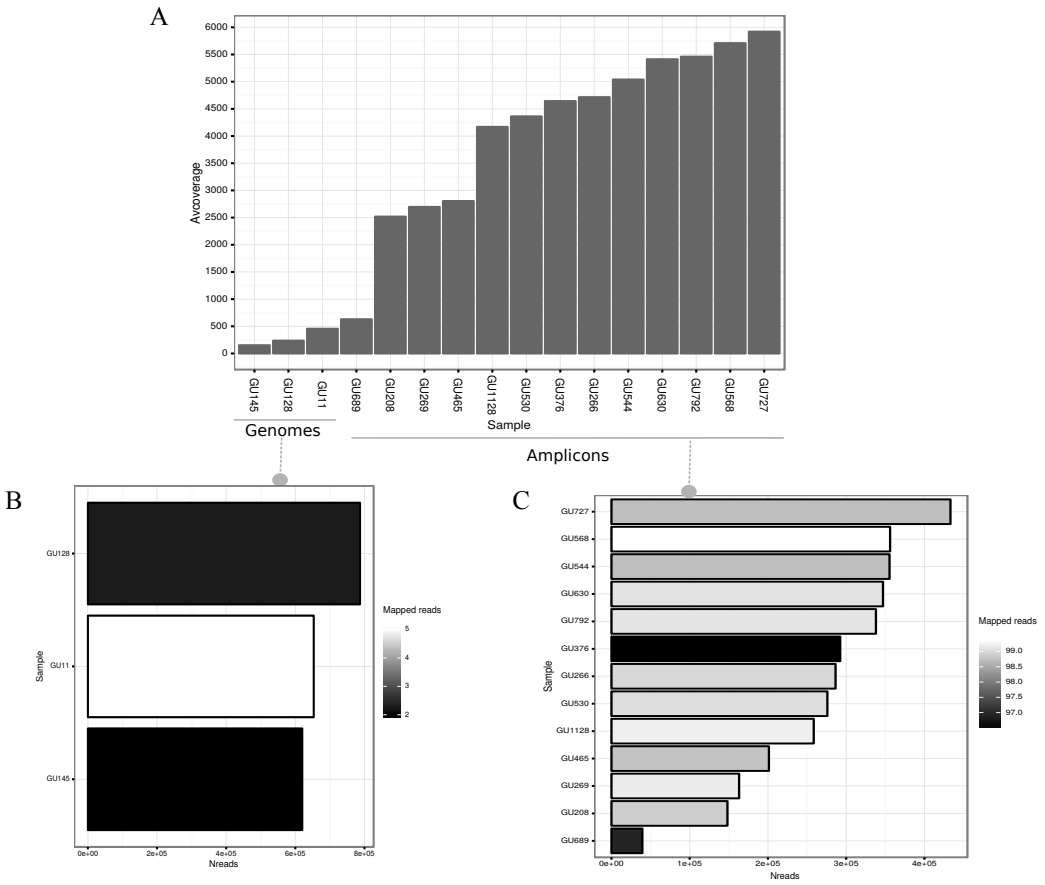


Figure 13 - Distribution of reads across the samples. (A) Total number of reads per sample. (B) Proportion of reads mapped to DENV genomes (C) Proportion of reads mapped to E gene amplicons.

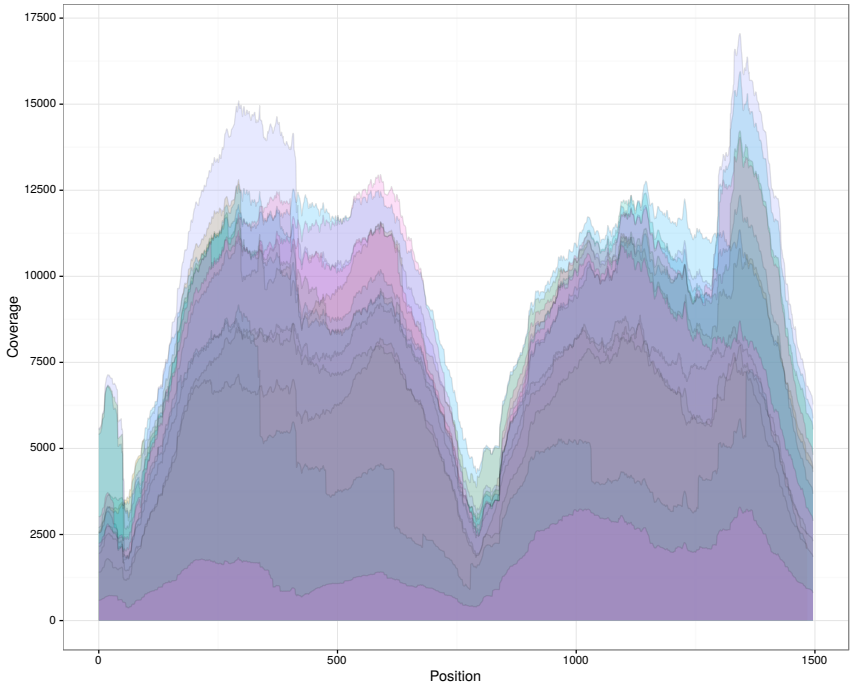


Figure 14 - Coverage of amplicons targeting the envelope gene of DENV-4.

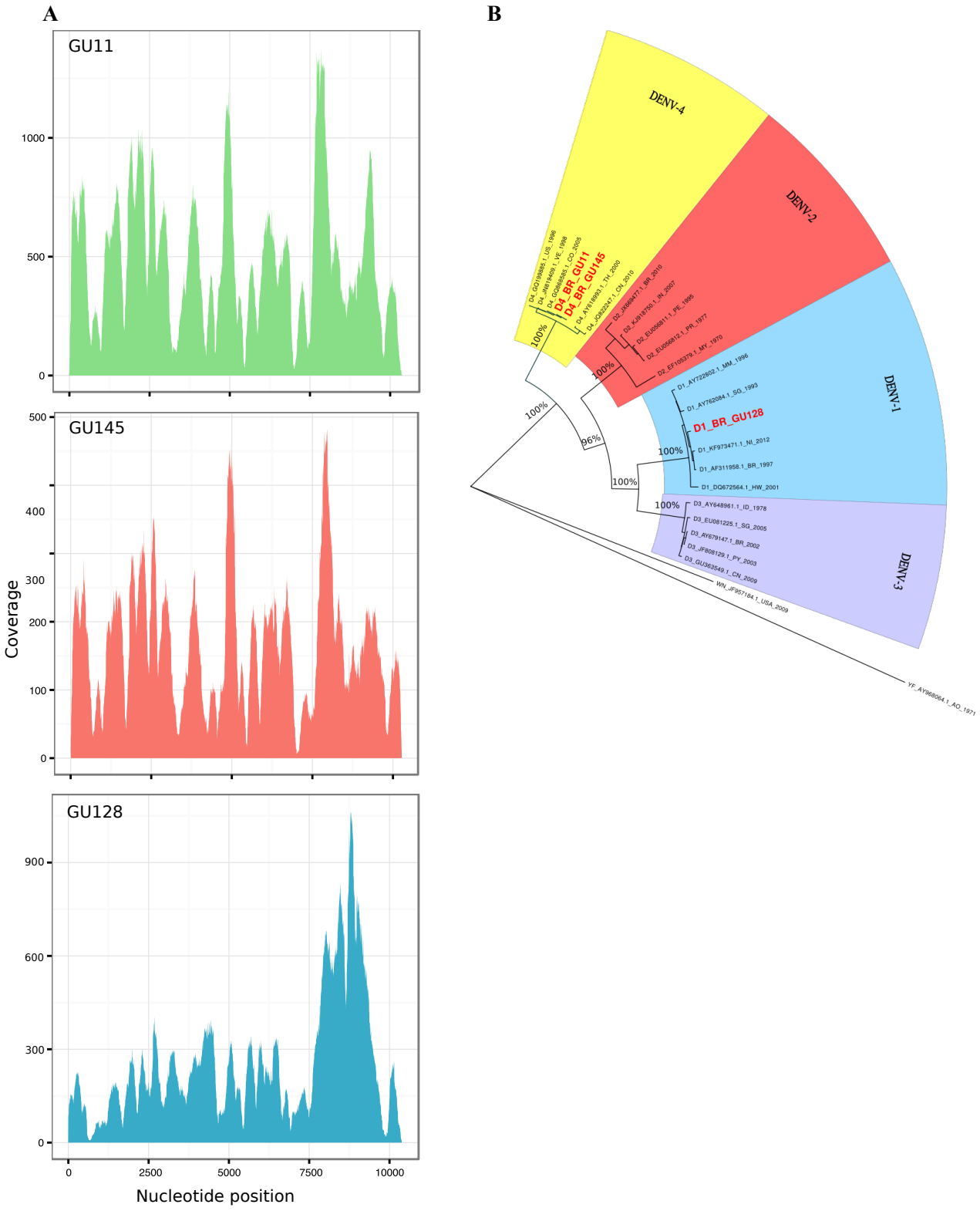


Figure 15 - Serotypes DENV-1 and DENV-4. (A) Distribution of depth coverage across the ds-cDNA samples. (B) Maximum-Likelihood (ML) phylogenetic tree showing the grouping of assembled genomes. Note that assembled genomes (red) grouped together with sequences belonging to serotypes DENV-1 and DENV-4.

4.3 DENV HAPLOTYPES

Because transmission is a stochastic process, identifying the composition of viral variants at intra-host level i.e., the collection of DENV variants present within-host, could provide a preliminary insight into the potential inter-host diversity. Our study comprised 60 envelope gene sequences and 29 genomes taken from 16 patients. No significant evidence of genetic recombination was detected ($p > 0.05$) (**Table 5**). Haplotype inference showed that most of the variants exist at low-frequencies (<1%) (**Figure 16A**), while only a few haplotypes occur at high frequencies within infected individuals. Despite stochastic events such us genetic drift and population bottlenecks impact on horizontal virus transmission, the most common variants seem to be transmitted more frequently than those less representatives (**Figure 16B**). However, we observed that haplotype frequencies can fluctuate among patients, and then major variants (more frequent) could turn into minor variants (less frequent) into different hosts.

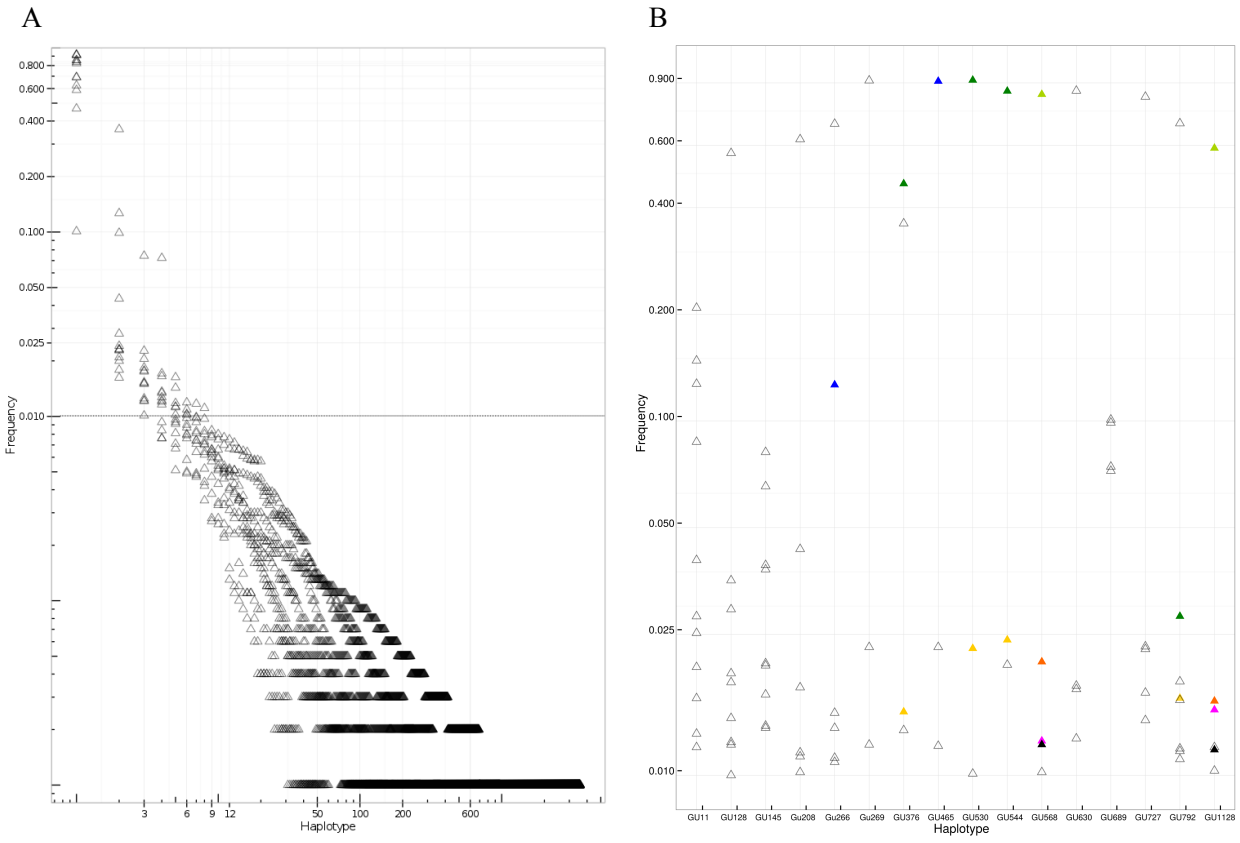


Figure 16 - Reconstructed haplotypes for DENV. (A) Haplotype distribution for all the samples. The cut-off value (1%) adopted for haplotype selection was represented by the dotted line. Plot on a x-axis log scale. (B) Frequency distribution of selected haplotypes. Color triangles represent the variants shared by different samples.

Table 5 - Total number of haplotypes selected per sample.

Sample	Haplotypes >1%	Phi P	DENV serotype	Target sequence
GU011	11	0.3079	DENV-4	Genome
GU128	9	0.4253	DENV-1	Genome
GU145	9	0.93	DENV-4	Genome
GU208	6	1	DENV-4	Envelope
GU266	6	0.97	DENV-4	Envelope
GU269	3	0.43	DENV-4	Envelope
GU376	4	0.92	DENV-4	Envelope
GU465	3	0.65	DENV-4	Envelope
GU530	3	0.09	DENV-4	Envelope
GU544	3	0.33	DENV-4	Envelope
GU568	5	1	DENV-4	Envelope
GU630	4	0.99	DENV-4	Envelope
GU689	4	0.89	DENV-4	Envelope
GU727	5	0.167	DENV-4	Envelope
GU792	8	1	DENV-4	Envelope
GU1128	6	1	DENV-4	Envelope

Furthermore, as a first effort to characterize the genetic variability of DENV-4 haplotypes in the envelope fragment, the entropy distribution (Shannon entropy, S) showed a similar profile across the samples but a heterogeneous variability across the envelope gene, with some positions exhibiting a high variability (lowly conserved) between positions 250-500 and 600-800 (**Figure S6A**). Similarly, when genomes were analyzed, we observed DENV-1 (GU128) have the lowest variability per position (0.01), whereas the DENV-4 genomes (GU11 and GU145) exhibited slightly high levels of variation (0.015 and 0.011, respectively) throughout the viral region (**Figure S6B**).

4.4 VARIANT CALLING

The base recalibration performed on the pre-processing stage evidenced an improvement in base quality scores. Therefore, more accurate bases were available for a better performance in SNVs calling (**Figure S5**). Moreover, SNVs detected from deep sequencing revealed a heterogeneous composition among samples, consequently limiting the identification common SNVs during transmission. Nonetheless, we observed that the most frequent substitution were transitions A>G and G>T followed by G>A and T>C. Transversions C>A were the most frequent followed by C>T into the full set of variants (**Figure 17**).

We further showed that the proportion of SNVs differs among domains, thus most nucleotide variants were located in domains II and III. Although haplotypes derived from DENV-4 samples showed a similar mutation profile, the composition and frequency of SNVs in the cell-culture derived sample was higher compared to that observed in the serum-derived sample. Likewise, clear differences in SNVs composition were detected between DENV-4 and DENV-1 samples, suggesting that the latter was more conservative.

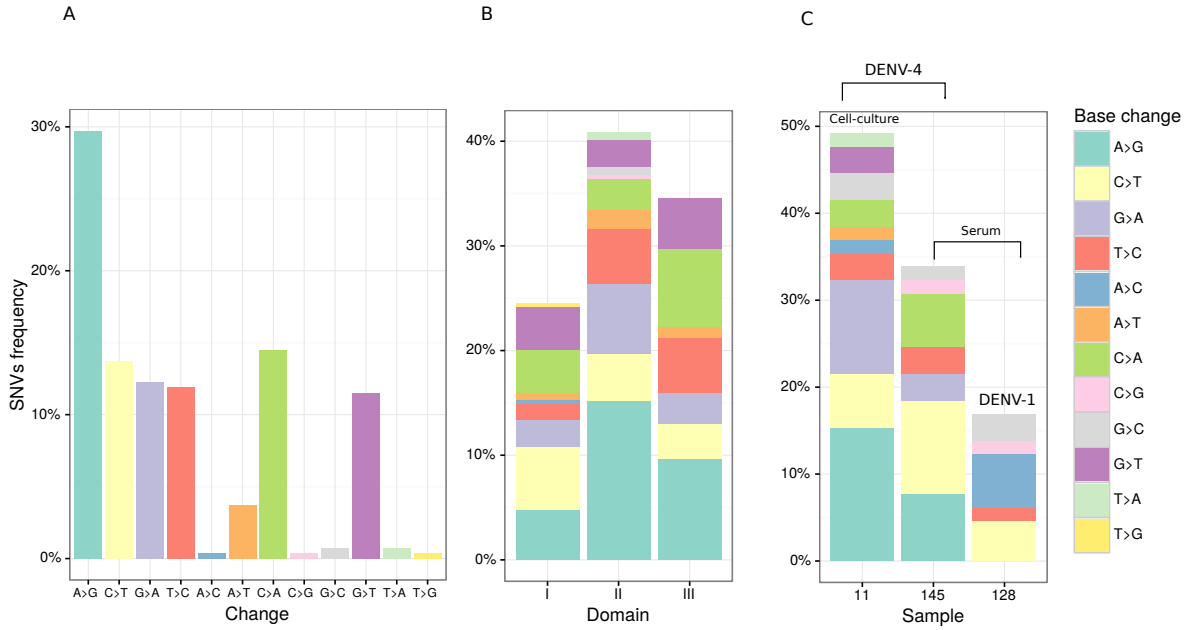


Figure 17 - Contribution of NGS-detected SNVs present in DENV samples. (A) Distribution of SNVS present in envelope-derived haplotypes. (B) Frequency of SNVS per domain. (C) Frequency of SNVS in genome-derived haplotypes.

4.5 CODON BIAS

RSCU values revealed an unequal usage of codons in the E gene of DENV-4 populations (**Figure 18**, **Table S3**). About 28% of the codons were over-represented (RSCU ≥ 1.1), likewise triplets such as AGA (**Arg**), GGA (**Gly**), UCA (**Ser**), and ACA (**Thr**) were the most frequent nucleotide combinations in codon composition, thus exhibiting a preference by A-ended codons. The triplets GCG (Ala) and CGA (Arg) were identified as the less abundant codons (RSCU ≤ 0.3), while CGG (Arg) was the only sense codon not present in the sequences. In addition, we found that GC_3 content was lower (45%) than the mean GC content (47%) and GC_{12} (48%). Similarly, after comparing the content of GC_{12} with GC_3 we did not identify a mutational bias influencing the codon usage in DENV-4. (r= 0.141, P>0.3164) (**Figure S7**).

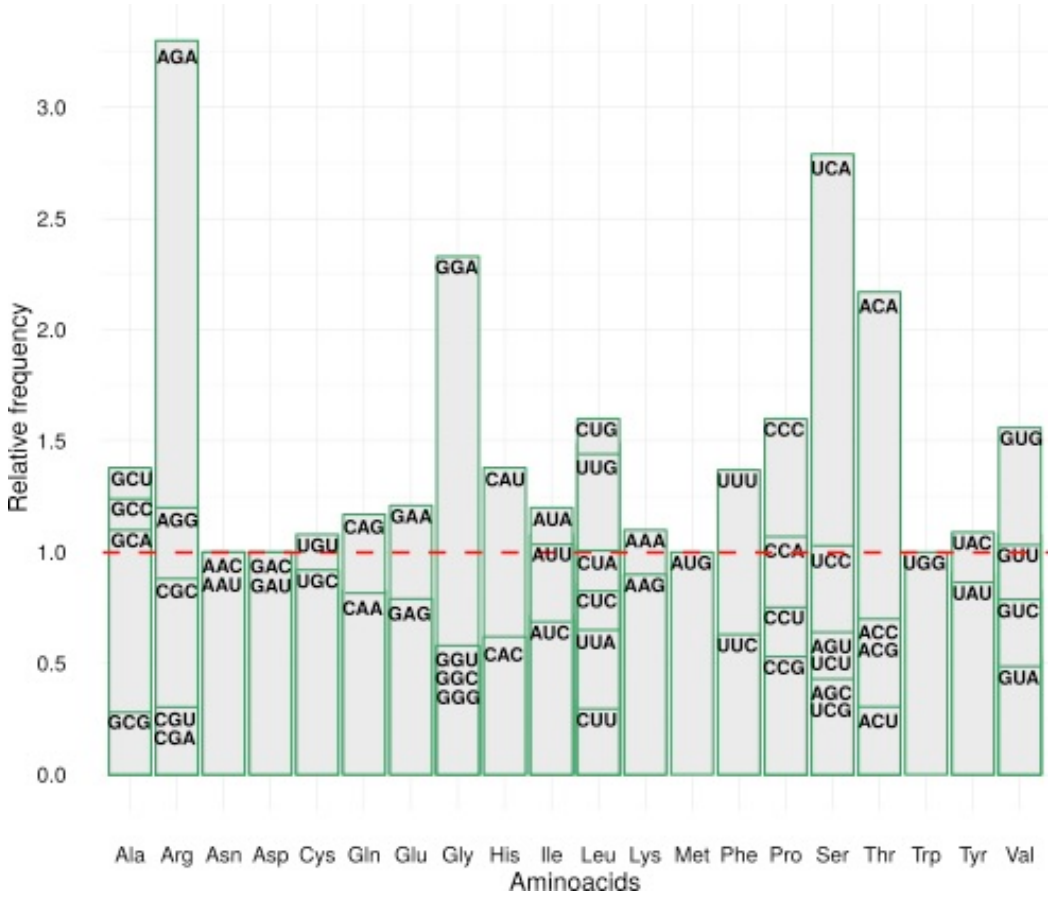


Figure 18 - Stacked bar chart showing codon usage in envelope for DENV-4 haplotypes. The relative synonymous codon usage (RSCU) is represented in the Y-axis. Values above the threshold (red dashed line) indicate a higher codon usage preference.

Trends in codon usage were observed in the correspondence analysis where the first principal axis accounted for 47% of the total variation, and the second most important factor explained about 19% of the remaining variation. The factor map based on RSCU values and amino acid usage showed a preference in the use of Lys contrasting with a reduction in the use of Asp, and Glu. We further observed a segregation of DENV-4 variants into several genetic groups located in the factorial plane. However, we did not recover a temporal structure (**Figure 19**).

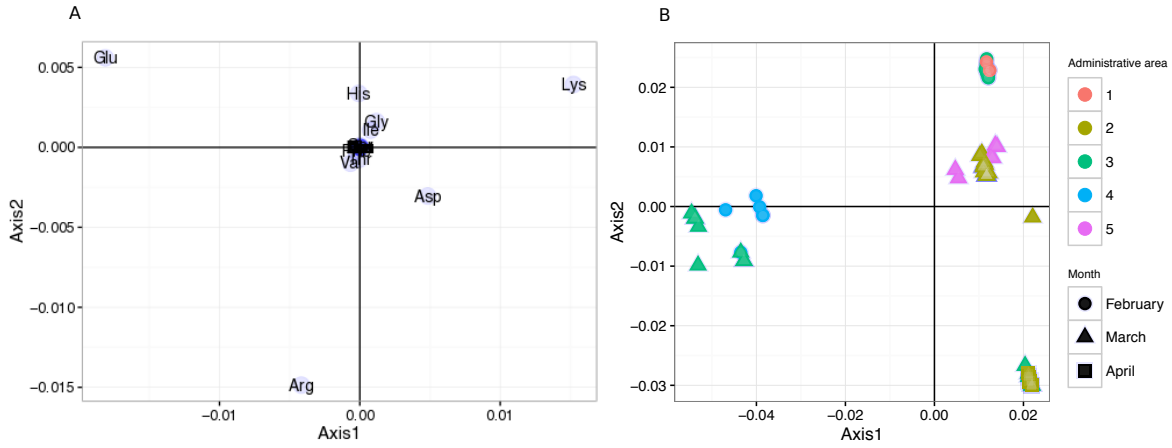


Figure 19 - Correspondence analysis. Ordination of amino acid (A) and codon usage (B) on the two principle correspondence analysis axes. Individual samples were grouped according to the Guarujá's administrative divisions and color-coded by annotation. Monthly distribution of DENV-4 isolates is represented using shapes.

4.6 INTRA-HOST GENETIC VARIABILITY

To establish the extent of genetic variation in DENV-4 haplotypes, we evaluated different diversity metrics (**Table 6**). Overall, we observed a low genetic diversity, with mean values of pair-wise nucleotide and amino acid distance of 0.0033 and 0.0021, respectively. Likewise, the percentage of polymorphic sites was 0.0189% for nucleotide and 0.0029% for amino acid sequences. The ratios of synonymous to nonsynonymous nucleotide substitutions (dN/dS) were less than unity for all the samples ($\omega<1$), with an average of 0.1992, suggesting strong negative selection pressures acting on DENV-4 populations. Distance metrics and selection pressure analysis performed across the whole coding region of DENV-4 variants showed similar values between the samples. Interestingly, we evidenced moderate negative selection acting on DENV-1 when comparing with DENV-4 population samples, independently of the source of virus isolation (**Table 7**).

Table 6 - Analysis of variability of envelope gene sequences of DENV-4 populations.

Sample	No. of variable nt	No. of variable aa	Π nt	Π aa	No. of Stop codons	Global dN/dS
Gu208	4 (0.27%)	3 (0.6%)	0.0011	0.002	-	0.480
Gu266	5 (0.33%)	2 (0.4%)	0.0011	-	-	0.213
Gu269	2 (0.13%)	-	0.0009	-	-	0
Gu376	2 (0.13%)	-	0.0009	-	-	0
Gu465	6 (0.4%)	-	0.0009	-	-	0
Gu530	2 (0.13%)	-	0.0009	-	-	0
Gu544	2 (0.13%)	-	0.0009	-	-	0
Gu568	4 (0.27%)	2 (0.4%)	0.0011	0.0016	-	0.32
Gu630	3 (0.2%)	1 (0.2%)	0.001	0.0010	-	0.159
Gu689	2 (0.13%)	-	0.0009	-	-	0
Gu727	4 (0.27%)	1 (0.2%)	0.0011	0.0008	1	0.32
Gu792	7 (0.44%)	4 (0.8%)	0.0012	0.002	-	0.426
Gu1128	9 (0.61%)	3 (0.6%)	0.0011	0.002	-	0.48

Pairwise distances among amino acid /nucleotide sequences.
d_N Nonsynonymous substitutions per site
d_S Synonymous substitutions per site

Table 7 - Analysis of variability of genome sequences of DENV populations.

Sample	No. of variable nt	No. of variable aa	Π nt	Π aa	No. of Stop codons	Global dN/dS
Gu11[a]	4 (0.04%)	1 (0.03%)	0.0002	0.0002	-	0.090
Gu128[bc]	8 (0.07%)	5 (0.15%)	0.0002	0.0003	-	0.050
Gu145[b]	4 (0.04%)	1 (0.03%)	0.0002	0.0001	-	0.085

[a] Virus grown in C6/36 cells
[b] Virus isolated from serum
[c] DENV-1 variants

For the envelope sequence, above 2.15% of all sites were polymorphic, whilst above 0.43% of the sites varied in the whole genome. All polymorphic sites corresponded to biallelic sites (**Figure S8**). Likewise, we observed a large number of fixed alleles and only a small number of rare alleles or loci showing intermediate frequencies. This result was also consistent with the Tajima's *D* test ($\theta = 0.0046$, $D = -0.905373$), which suggests natural selection removing genetic variation or a recent population expansion of DENV-4.

In order to identify hotspots in the distribution of polymorphic sites, we plotted the density of SNPs across the alignments. As we expected, the region corresponding to the envelope gene

was highly variable in DENV-4 (**Figure 20**). Furthermore, when polymorphisms are represented by codon position in the envelope gene, we observed a non-homogeneous distribution of polymorphic sites ($p < 0.05$), which were aggregated into at least two hotspots. Most SNP*s* were located within positions 200-500 and constrained on the second codon position.

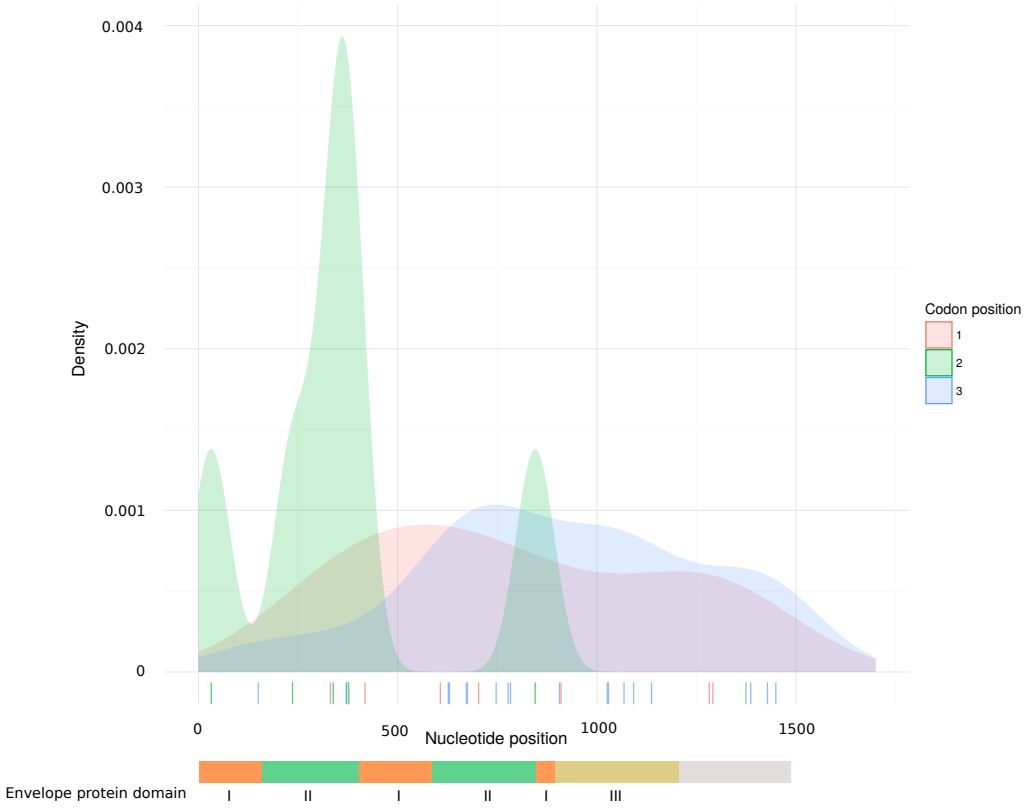


Figure 20 - Distribution of polymorphisms in the envelope gene of DENV-4. Linear representation of the DENV-4 envelope is showed at the bottom. Codon positions are represented by colors.

4.7 NATURAL SELECTION

No potential recombination events were found in our datasets using envelope or genome sequences. As previously noted, the estimates of selection pressures on coding sequences (global omega) indicated purifying (negative) selection pressure (**ω** <1) acting at the gene level (**Table 6, Table 7**). In contrast, when we investigated codon adaptation, differences in selection pressures were detected operating on individual codon positions. Hence, we found five codons being positively selected across the envelope gene (**Table 8**). However, position

235 associated with the envelope glycoprotein domain II, was the only residue consistently recovered by the different datasets and methods evaluated (**Figure 21**).

Furthermore, to test if natural selection acts uniform throughout time, we tested the distribution of dN/dS along both sites and branches. Nonetheless, not positively selected sites were identified using branch-site approaches (**Table 9**).

Table 8 - Individual sites under selection across envelope alignments using site-level approaches**.**

Method	Haplotypes		Haplotypes + Additional sequences [1]	
	Positive selected sites	Negative selected sites (%)	Positive selected sites	Negative selected sites (%)
FEL	**235**	8 (0.24)	64[†], **235**[†]	5 (1.01)
iFEL	NA	2 (0.40)	NA	5 (1.01)
REL	111, 113, 126, **235**	19 (3.83)	64, **235**, 350	NA
SLAC	NA	2 (0.40)	NA	8 (0.24)
FUBAR	**235**	5 (1.01)	**235**	35 (7.07)
MEME	113, 126, **235**[†]	NA	126, **235**	NA

Selected sites were detected at significance level set at $p \leq 0.05$ for FEL, iFEL, SLAC and MEME. Bayes factor was set at ≥ 50 for the REL and FUBAR method. [†] Marginally significant sites (>0.05-0.08 range).

Table 9 - Summary of selection analyses using a branch-site approach.

Method	aBSREL		
	Branch-wise ω variation (MG94) (*log* L)	Branch-site ω variation *(log* L)	p-value
Haplotypes	-2383.84	-2383.84	>0.05
Haplotypes + Additional sequences	-3662.59	-3659.80	

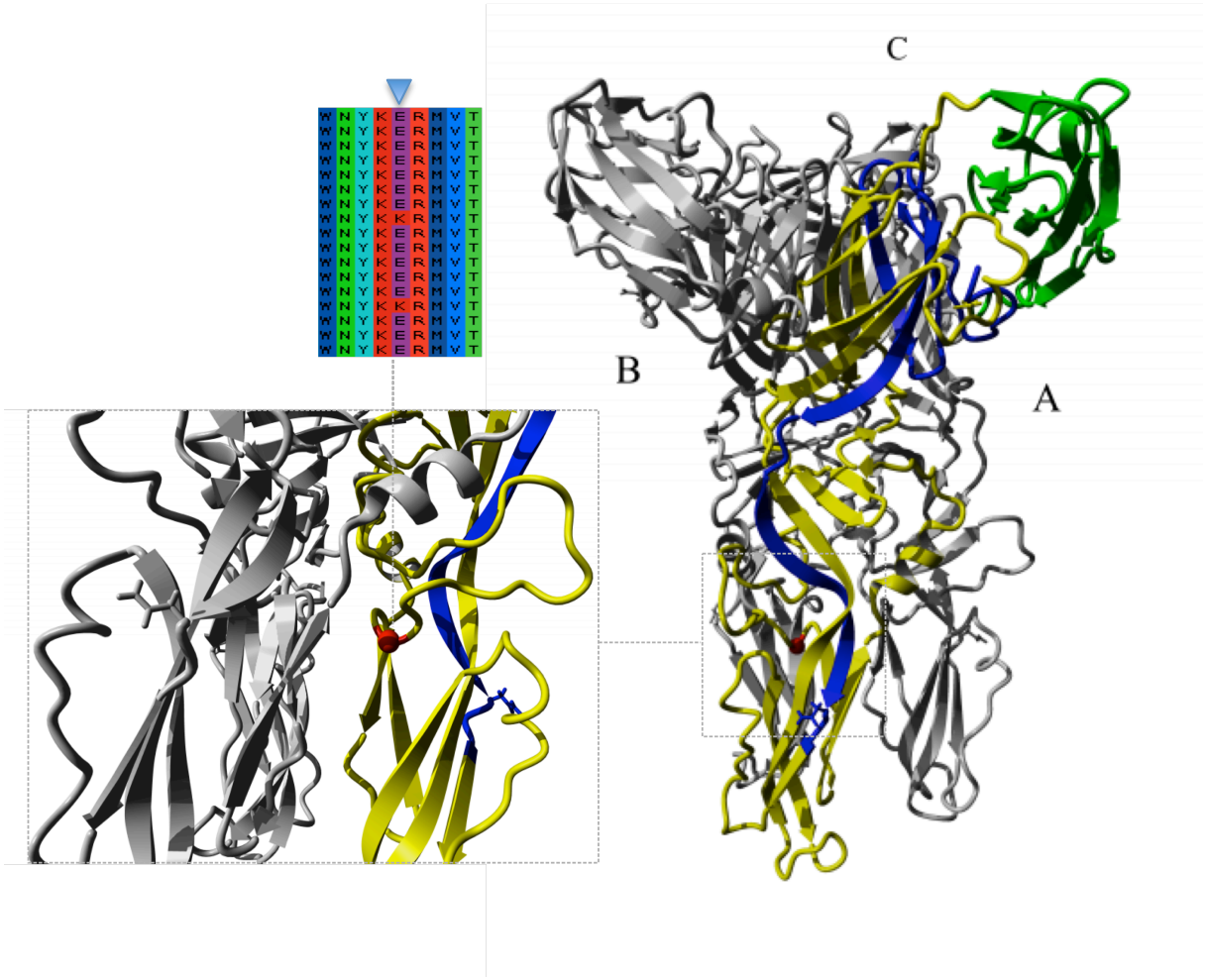


Figure 21 - Modeled envelope protein structure in its trimetric conformation for DENV-4. The positively selected residue 235 was indicated in red while domains I, II, and III are represented in blue, yellow and green, respectively. Monomers were represented with A-C letters.

After assessing selection pressures on genomes, we found four sites under positive selection for DENV-4 sequences and 21 sites for DENV-1 (**Table 10**). Although detected residues varied widely among methods, position 921 in DENV-1 alignment was conservative across most approaches, demonstrating positive selection acting on the non-structural (NS1) protein (**Table S3**).

As with the dataset based on viral envelope sequences, we did not detected positive selection on sites along branches using the aBSREL approach (**Table 11**). Likewise, all selected sites recovered by the MEME approach were located along terminal branches rather than on internal branches in the phylogenetic tree (**Figure S9**). Finally, no evidence for selection was detected through lineages representing DENV-4 populations isolated from serum and cell culture (**Figure S10**, **Table 11**). Nonetheless, we identified borderline significant variation of

selective pressures over branches across the serum group. Our results suggested relaxation of selection pressure rather than substantial positive selection (**Table 12**).

Table 10 - Individual sites under selection across polyprotein alignments using site-level approaches.

	Positive selected sites	**Negative selected sites (%)**	**Positive selected sites**	**Negative selected sites (%)**
FEL	NA	84 (2.48)	**921**	145 (4.27)
iFEL	NA	9 (0.26)	**921**, 1068	29 (0.85)
REL	630, 1505	3 (0.08)	95, 577, 618, 823, 868, **921**, 950, 1068, 1208, 1298, 1340, 1349, 2278, 3122	2 (0.06)
SLAC	NA	9 (0.26)	NA	31 (0.91)
FUBAR	NA	301 (8.88)	**921**	1192 (35.14)
MEME	248, 292	NA	495, 512, 2276, 2974, 3267, 3272	NA

Selected sites were detected at significance level set at $p \leq 0.05$ for FEL, iFEL, SLAC and MEME. Bayes factor was set at ≥ 50 for the REL and FUBAR method.

Table 11 - Summary of selection analyses using branch-site approaches with genomic datasets.

Method	**aBSREL**		
	Branch-wise ω variation (MG94) (*log* L)	Branch-site ω variation *(log* L)	p-value
DENV-4	-16004.79	-16004.79	>0.05
DENV-1	-20662.93	-20651.62	>0.05

Table 12 - Omega variation along branches representing different viral populations isolated in cell culture and serum.

Haplotypes (Foreground)	**Log likelihood H0: $\omega_0 = \omega_1$**	**Log likelihood H1: $\omega_0 \neq \omega_1$**	**p-value**	**Log likelihood H0: $\omega_1 \geq 1$**	**p-value**
Cell culture	-16094.67	-16092.09	0.16	NA	NA
Serum	-16094.67	-16091.13	0.07	16091.13	0.77

ω_0: Proportion of sites under selection on background branches. ω_1: Proportion of sites under selection on foreground branches.

4.8 INTER-HOST DYNAMICS

In order to detect the occurrence of genetic structure in our data set, we performed a principal component analysis (PCA). We identified three separate clusters arranged along the principal component, which accounted for 39% of the total variation. The second component accounted for 20% of the total genetic variability, suggesting a gradual similarity between some closely related variants (**Figure 22A**). As with the PCA method, we observed a consistent pattern of the genetic structure by performing a distance-based approach (**Figure 22B**). Hence, four clusters were identified, which also grouped variants sampled at different times, as previously noted.

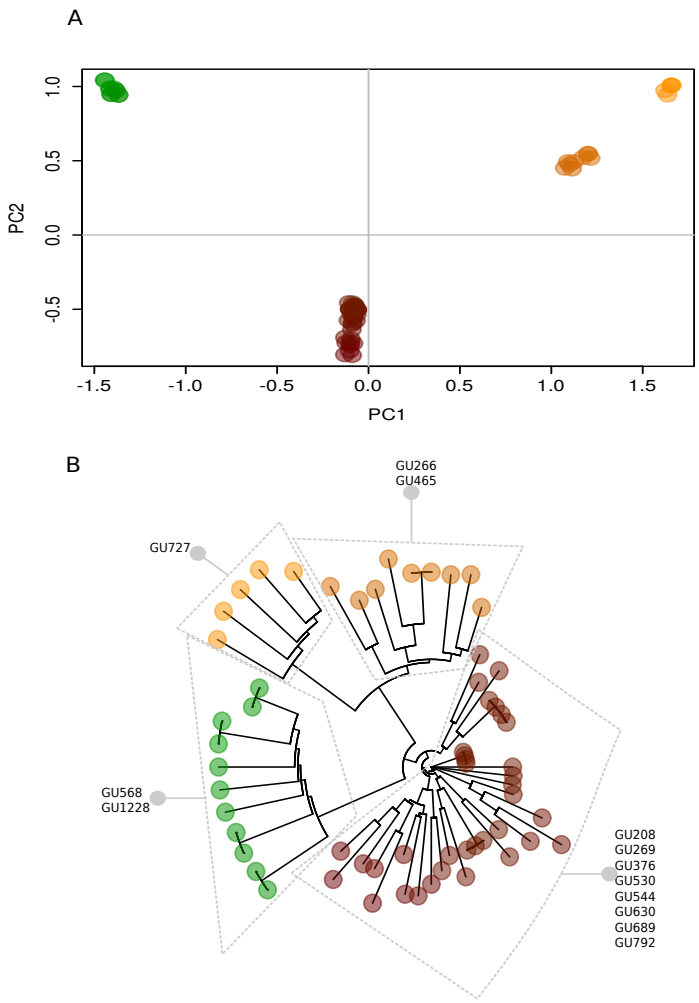


Figure 22 - Population genetic variation in DENV-4 samples. (A) PCA biplot of DENV-4 haplotypes. (B) Neighbor joining tree.

Despite our limited sample size, we attempted to elucidate major transmission events present in our sampling by integrating genetic and temporal data into a transmission tree. Genomes were not included in the analysis due to the small number of samples. The reconstruction exhibited several subtrees and some missing links that evidenced a biased sampling. However, we identified variants that appeared to be more successfully transmitted through time and were circulating in both early and late months into the outbreak. Overall, we observed that envelope gene haplotypes from samples GU208 and GU376 were the most common nodes in the transmission chain, which also set well-supported links with other samples (**Figure 23**).

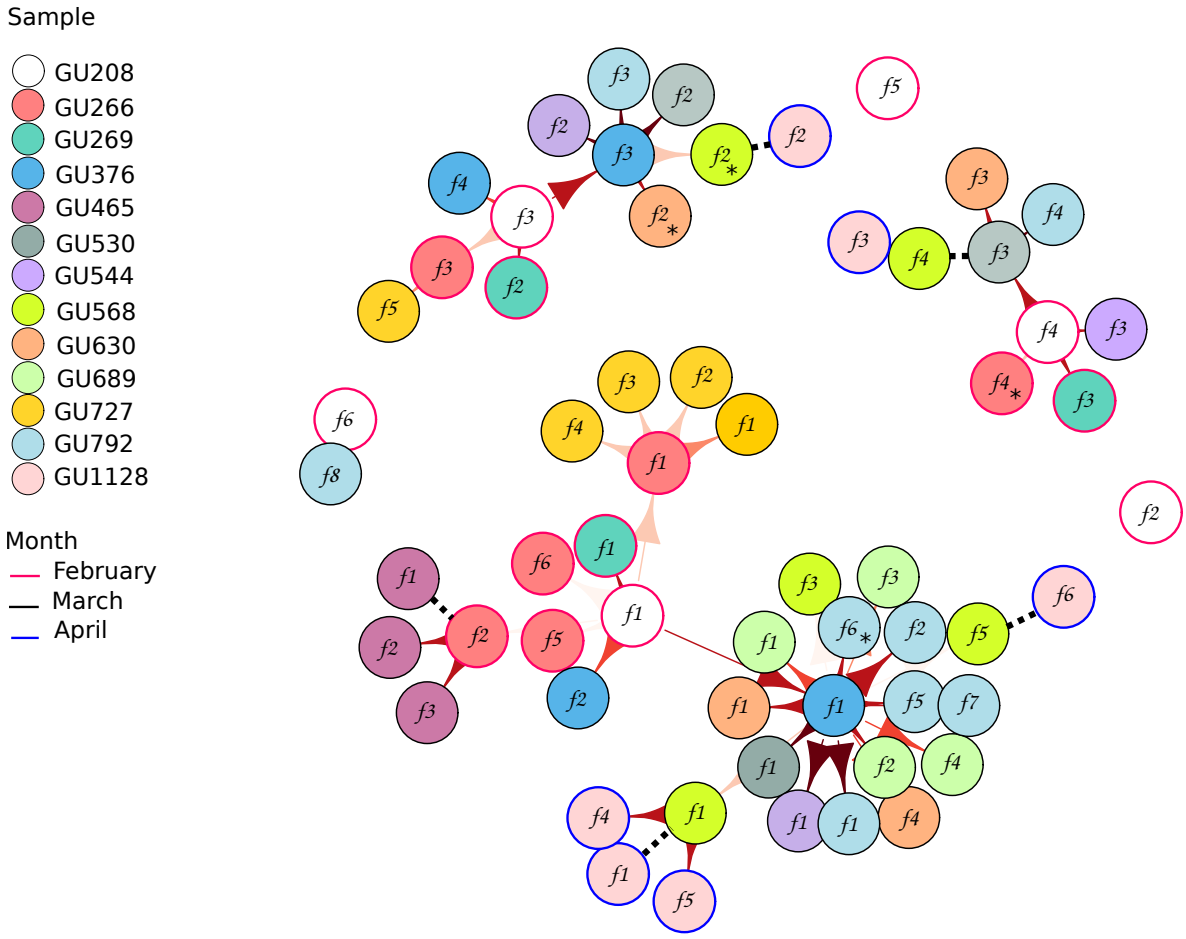


Figure 23 - Reconstruction of transmission events. Darker arrows correspond to well-supported transmission links. Haplotype frequencies f are represented in increasing order by host. Identical haplotypes are connected with a dashed line. Asterisks indicate those haplotypes bearing the mutation E235K. The month of sample collection was showed using border-colors.

5 DISCUSSION

5.1 PHYLOGEOGRAPHY

After the re-emergence of DENV-4 in Brazil, the spread of the virus across several Brazilian locations became a public health threat, reaching an unprecedented record of cases in the country [35,106]. Previous research studies have shown how inter and intra-host viral diversity continually feedback to one another [107,108]. Based on this point, we incorporated spatio-temporal information in a phylogeography analysis, in order to elucidate the origin, dispersal routes, and demography history of DENV-4 circulating in Guarujá, which ultimately provided us with insight into the epidemiological setting of the virus.

By including additional sampling (e. g. isolates from São Paulo), we recapitulated the phylogeography of DENV-4, genotype II in the Americas [109,110]. Although there is evidence of the circulation of genotype I in Brazil [111–113], all samples included from Guarujá and most of Brazilian localities belonged to genotype II; thereby, supporting the role of this genotype in boosting the recent circulation of DENV-4 in the country [33,37,114].

Similar to a recently published study, we identified multiple introductions of DENV-4 into Brazil [110]. However, the virus circulating in Guarujá corresponded to the most broadly distributed and well established lineage in Brazil, which also reached São José de Rio Preto in São Paulo, and other states such as Mato Grosso, Rio Grande do Sul, Roraima, and Amazonas. Remarkably, Northern South America (e.g. Colombia, Guyana, Venezuela) seems to be a center of viral diversity and a common route for the traffic of viral pathogens with Brazil [109,115–118].

Despite we included a more geographically diverse dataset, the estimated posterior mean substitution rate (9.1827E-4 $\times 10^{-4}$ substitutions/site/year) for DENV-4 was comparable with Allicock et al., 2012 [109]. Although we did not establish the local dispersal pattern of the virus in Brazil, we found a close relationship among strains isolated in São Paulo (Guarujá, São José de Rio Preto), and Mato Grosso, both corresponding to densely populated states. Furthermore, evidence for the circulation of DENV-4 in Goiás, Mato Grosso do Sul, Minas Gerais, and Rio de Janeiro between 2011 and 2012 has been published [119–123], thus

suggesting the wide dissemination of the virus across the region. However, samples from these states were not available to be included in our analysis.

As reported by Villabona-Arenas et al., 2016, we identified two lineages in Guarujá, which emerged and co-circulated over similar time-scales. Unfortunately, we recovered low support values in most nodes grouping Brazilian sequences, probably due to the high genetic similarity among these sequences.

Previous studies have demonstrated the re-emergence of DENV-4 in Roraima state and hypothesized its cryptic or imperceptible circulation before its detection [110,114]. This re-emergence and spread of DENV-4 toward different Brazilian locations, including Guarujá, coincided with the increase in the number of reported cases and the rising trend of genetic diversity observed in the Bayesian skyplot after 2010 (**Figure 12**). This epidemiological framework could be associated with the high number of susceptible host, which were not immunized against DENV-4 in Brazil [110,114,119,123].

5.2 SEQUENCING AND SEROTYPE DETECTION

Identifying the composition of viral populations at the intra-host level provides a preliminary insight into the potential inter-host diversity, and allows to better understand both the emergence and spread of DENV variants [56,66]. In this study, we analyzed 16 samples in order to characterize the intra-host diversity of DENV-4 for the first time. Here, we successfully obtained whole-genome haplotypes from three samples and envelope haplotypes from 13 samples.

In the last decade, the development of next-generation sequencing technologies have revolutionized the field of virus research by sequencing a variety of samples and allowing their exploration from different methodological approaches [124–126]. Targeted sequencing is a widely used strategy to construct DENV NGS libraries based on PCR-amplified fragments [56,62,64,127,128]. Although the main advantage of this approach is the specificity and the target sequence enrichment [129], we sequenced effectively three samples by implementing an alternative methodology based on the sequencing of ds-cDNA material. This approach has the potential to provide an enhanced detection of heterogeneous viral

populations and also it involves fewer steps in sample pre-processing, which reduces sequencing bias and increases efficiency [126].

Although cDNA-based library construction remained unexplored for high throughput sequencing of DENV, we obtained positive results but a limited performance, due to the presence of unwanted RNA in our samples. The large amount of host RNA contamination (i. e. mRNA and rRNA) together the potential risk of RNA degradation and decrease of viral RNA titers over time, represented the major technical challenges during sample preparation. Certainly, the high amount of host-derived RNA was evidenced in the high percentage of reads recovered by competition with the viral RNA [130]. Nonetheless, we were able to sequence three complete coding genomes that correspond to the first genomic sequences available for the outbreak and we provided additional evidence supporting the co-circulation of DENV-1 and DENV-4.

Our work support previous findings published by Villabona-Arenas et al., 2014 showing the emergence of DENV-4 and simultaneous circulation of all four dengue serotypes in late 2012-early 2013 in Guarujá, São Paulo [33]. Although most of our samples did not correspond to the early outbreak, we reiterated the co-circulation of DENV-1 and DENV-4 in Guarujá. A similar result was showed by Amâncio et al., 2014 in Minas Gerais state, where DENV-4 became the predominant serotype, following by the simultaneous circulation of DENV-1 [131]. Likewise, in the states of Mato Grosso and Mato Grosso do Sul was reported the co-circulation of both serotypes [119,132]. These results highlight the hyperendemic status of dengue in densely populated cities such as Guarujá, and promote an increase in surveillance efforts to control the disease.

Mixed infections contribute to the diversity of DENV populations at intra-host level [42,49,61]. By employing NGS we identified some samples with reads belonging to DENV-1 and DENV-4 serotypes, which was suggestive of potential mixed infections (i.e. super-infection, co-infection). However, given the low number of reads we cannot conclude that the number of reads is directly proportional to serotype abundance [133]. Evidence of co-infections between both serotypes has been previously documented. In São José de Rio Preto, São Paulo, Colombo et al., 2013, reported 359 positive samples in an outbreak headed by DENV-1. Of these, one sample was a co-infection between DENV-1 and DENV-4 [134]. Similarly, co-infections by both serotypes were also reported in Mato Grosso [132].

Nonetheless, co-infection cases between DENV-4 and other serotypes have been reported in Contagem, Minas Gerais [135]. Here, we reported the co-circulation of DENV-1 and DENV-4 serotypes in Guarujá and the potential co-infection by both serotypes in at least five patients. The presence of two serotypes within patients reveals the high transmission taking place near to the largest metropolis of Brazil.

Interestingly, according to the local public health authorities, after the burden attributable to DENV-4 in 2013, DENV-1 established itself as the dominant serotype (82%) followed by DENV-4 (16.1%) in 2014 [136]. Therefore, our results suggest an alternating epidemic pattern [137,138], where DENV-1 would be associated with low circulating levels in the population during the 2013 outbreak and then it peaked in prevalence for the next outbreak.

In the present study, we further investigated the intra-host diversity of DENV-4 based on the amplicon sequencing of the envelope gene region. Deep sequencing on a genome-wide level provides a comprehensive approach to explore variant diversity and get individual strains from a population [139,140]. However, in dengue, the study of the envelope region is widely used as an informative genetic target given its role in mediating receptor binding and membrane fusion [7,141]. Furthermore, the envelope gene is considered a mutational hotspot [64], responsible for coding the main antigenic determinants of the virus [142].

As expected, we recovered a higher coverage by using an amplicon-based sequencing method [126]. Nevertheless, we observed an heterogeneous depth of coverage across the envelope, with a sharp fall near position 800 (**Figure 14**), probably because sequencing and PCR amplification can be biased by the GC content and GC-rich motifs present in target sequences [143–145].

5.3 HAPLOTYPES

RNA viruses display high spontaneous mutation rates compared with DNA organisms [146]; however, distinguishing between errors and real variants, in order to capture the whole picture of viral diversity in a population, result a complex issue [147]. In this study we investigated the intra-host genetic diversity of DENV-4 populations by using high-throughput sequencing to reconstruct viral haplotypes. Although we characterized some whole genome haplotypes, we focused primarily on the inference of gene-specific trends in intra-host diversity, since this

gene encodes for the envelope protein which is highly immunogenic, and plays an important role in infectivity, viral membrane fusion, receptor binding and host antibody response, as highlighted in a previous report [66].

The influence of haplotype composition was reported for several viruses, including the CSFV, HBV, HCV, HIV, and PV with the purpose of investigating a wide range of topics such as pathogenesis, evolution, adaptability, and diversity [44,148]. In arboviral pathogens like dengue virus, the maintenance of intra-host variants have been mainly associated with its adaptability to the arthropod vector and vertebrate host [127].

Even though, our analyses include haplotypes being represented at ≥1% of the viral population, the vast majority of these haplotypes corresponded to low frequency variants. According to the quasispecies theory, mutants present at low frequency in viral populations contributed for adaptability and viral evolution. Therefore, minority variants may harbor mutations to overcome selective constraints imposed by factors such as immune response and antiviral agents [43,44]. For instance, the role of minority variants in disease outcome and drug-resistance has been widely documented in HIV [149–151]. Despite, the phenotypic role of minor variants in DENV evolution remains poor understood [45], a recent study with CHIKV suggested the association among low frequency variants with increased infectivity and transmissibility [152].

One intriguing question we tried to address along some sections in this study is whether DENV-4 haplotypes could be detected across different individuals. Interestingly, we found that haplotypes are being shared among host. Most importantly, haplotype frequencies can change among host, probably as a result of being subjected to different selection intensity [44]. Besides transmission of high-frequent variants, we identified some haplotypes, which were maintained even at low-frequencies across individuals. In this respect, two possible scenarios have been previously proposed: *i*) variants retain some level of replicative capability and efficiency of infection/transmission, and *ii*) transmission is helped by complementation [43,44]. Additionally, the occurrence of bottleneck events is considered a major factor associated with changes in the composition of haplotype assembles and fluctuations in viral fitness [44,47].

Recombination is one of the forces contributing to maintain genetic diversity in DENV [153]. Strand switching is the common proposed mechanism for recombination, which presupposes the concomitant infection with at least two different strains [153]. Nevertheless, little is known about recombination in DENV-4. Abubakar and colleagues [154], suggested the emergence of the genotype IIA in Malaysia as a result of intra-serotypic recombination between envelope sequences of DENV-4. Likewise, Woreobey and colleagues, revealed evidence for recombination between envelope sequences of DENV-4 [155]. Our results did not provide support for recombination events in DENV-4 haplotypes. Despite our findings suggested that recombination does not play a significant role in shaping genetic diversity of variants circulating at the intra-host level, it is worth noting that most of the target sequences corresponded to partial haplotypes, and thereby we mainly focus on detecting breakpoints in the envelope gene, which does not deny the possibility of recombination occurring across other genes. We also speculated that available methods for assessing recombination may lose sensitivity when analyze haplotype populations at the intra-host level.

5.4 INTRA-HOST GENETIC VARIABILITY

Since DENV-4 evolves as a heterogeneous population, the repertoire of DENV-4 sequences contributes to the variability within and between hosts. According to a previous report, structural proteins such as envelope represent the most hyper-variable region for studying quasispecies in DENV [156]. Despite this expected variability, and in agreement with previous estimates, the intra-host diversity of DENV-4 showed lower genetic variation and hence lower levels of amino acid diversity compared to those seen the other RNA viruses. However, levels of diversity were very similar with other serotypes [49]. This is relevant as DENV-4 was able of driving a high level epidemic situation, thus suggesting the role of the immune status in a susceptible population.

Negative selection is responsible by removing deleterious genetics variants in virus populations [71]. Current evidence supports that the lack o positive selection for the entire envelope gene in DENV is frequently associated to the highly immunogenic properties of its domains [49]. Likewise, several studies have suggested that the low intra-host genetic diversity of DENV in human host may be attributable to the effects of purifying selection on nonsynonymous variation.

Moreover, our analyses indicated that variability was not uniformly distributed across the envelope, whereas domains I and II exhibited the highest variability in the protein. Overall, these findings suggest that mutations across these regions may have important implications in structural features and antigenic properties [8,141,157].

Association among virulence and genetic variability is worthy of further investigation given that several studies seems to reach different conclusions [45,49,61,158]. However, we were not able to establish any association between genetic diversity and clinical outcome. Though, the lack of a pattern between genetic diversity and other epidemiological variables might underpin a small sample size, related studies found not significant correlation between intra-host genetic variability and age, sex, host immune status, and viraemia [49,159].

Premature translation stop codons result in non-functional proteins and therefore in defective viruses. As previously documented, when the frequency of stop mutations reaches high frequencies, the effect of bottlenecks is reduced to the point were a defective variant can become widespread [160]. According to the data presented here, we identified a haplotype bearing a mutation resulting in protein truncation, however, this haplotype did not appear to be successfully spread among susceptible hosts. Thai and colleagues [49], postulated complementation as the most probably mechanism explaining the presence of genome-defective DENV due to in-frame stop codons.

5.5 SNVS AND CODON BIAS

Identifying the true biological SNVs composition from technical artifacts represent a major challenge in deciphering NGS data [161,162]. In order to improve variant detection we applied different filtering strategies. However, due to the poor quality and low coverage of some bases, many reads were filtered out in the process, therefore, limiting the detection of common variants across individual samples and the feasible detection of changes in SNV frequency.

We explored intra-host diversity by mapping variants onto reference sequences in order to detect major tendencies. The accumulation of transitions and transversions across the samples, revealed a mutational bias towards A and T nucleotide variants. We hypothesized that the underlying substitution spectrum account for the codon usage, virus fitness and,

consequently, the intra-host genetic variation of DENV-4, as highlighted in other studies [163].

A mutational hotspot of SNVs in the envelope domains I and II of DENV from human derived samples. Likewise, coldspots have been associated with functionally important residues, which preserve a stable protein phenotype [127]. Other study reported that frequency and location of SNVs can vary among populations. Differences in the number of SNVs and the relatively lower intra-host diversity of DENV-4, could be associated with the distribution of SNVs across reads corresponding to different variant sequences instead of clustering together on one sequence, as have been previously suggested [127].

Mutational bias is considered the major determinant affecting the preferential usage of synonymous codons in RNA viruses [164,165]. Therefore, characterizing the bias in codon usage is important for understanding the evolutionary patterns and delineating the constraints that shape genetic diversity in DEN-4 [166].

Deviations in codon usage patterns occur widely in nature and can be detected even within a single gene [167,168]. Herein, we have presented evidence, which suggests codon bias in the envelope region of DENV-4. The observed bias in the nucleotide composition of the DENV-4 envelope showed a slight preference for A-ended codons, thus indicating that the A content at the third position of the codons influence the overall synonymous codons usage pattern. Moreover, the presence of four over-represented codons might be representing a translational efficiency and a relatively stable genetic composition in DENV-4, as has been showed in other RNA viruses [169].

Despite we evidenced a selective preference for codons in DENV-4 sequences, highly over-represented codons (RSCU > 1.8) were infrequent, while most of the remaining codons lacked strong codon preference. Mutational pressure and natural selection are frequently proposed as mechanistic hypothesis explaining the bias in codon usage among certain RNA viruses such as dengue. [168–171]. Furthermore, the alternating host replication between human and vector cells could be constraining the codon representation in DENV-4. Thus, a weak codon bias could result into a plastic response to replicate in different host cells [172].

Our results showed congruent findings with those of Behura and Severson by analyzing the genome wide genetic diversity in dengue [173]. Thereby, we corroborated a preferential codon usage for amino acids: Gly, Ser and Thr in dengue, but also we indicated a preference in codon choice for Arg in the envelope of DENV-4 [174]. Regions rich on these amino-acids have been associated with differences in host infectivity, binding affinity, and neutralization susceptibility [175–177].

In addition to codon usage analysis, we performed a correlation analysis between GC12 versus GC3 to determine if mutation pressure is the sole determinant of codon choice. However, a non-significant positive correlation was noticed for our dataset, suggesting that only mutational pressures did not govern patterns of base composition and codon usage.

Pal et al., 2014 [171] showed that mutational bias plays a major role in codon usage in DENV-4. Here, we posit that mutational pressure would have less influence than expected and some others factors such as natural selection could be playing an important role in shaping codon bias at the intra-host level, similar with some other reports for dengue and other flaviviruses [164,178]. Certainly, given the limited sample size, it would be interesting to re-evaluate this analysis by including additional samples and compare with other target regions in order to determine if this pattern is identical for all the genes in DENV-4.

Correspondence analysis is a useful multivariate technique to uncover trends in both codon and amino acid usage [96]. Therefore, by performing COA we examine amino acid and codon composition in DENV-4 envelope. A previous study found that codon usage in DENV-4 is apparently conservative between phylogenetically closely related strains [171]. However, we observed variants closely related, forming clusters, suggesting differences in codon usage at intra and inter-host level. Nonetheless, we did not observe clustering patterns related to the spatial or temporal origin, possibly because some of these variants were being transmitted through both scales. Our results also showed a discrepancy with a recent work which suggested that geographic origin can shape codon usage in DENV-1 [164]. This conclusion might be assumed as consequence of the spatial scale at which the analysis was performed by comparing codon usage bias by continental region.

Concerning the amino acid usage, it was observed that Arg, Glu and Lys amino acids had an important contribution to the amino acid usage. In this respect, these amino acids have been

associated with hydrogen-bond formation and stabilization of the domain-domain and envelope-antibody interfaces in dengue [8,179]. Thereby, suggesting that charged amino-acids residues could affect amino-acid usage.

5.6 SELECTION PRESSURES

Intra-host selection plays an important role in fitness, but also is the predominant force changing variant frequencies [180]. Evidence of natural selection acting on intra-host dengue populations has been previously observed [49,62,66,127,181]. Likewise, natural selection in the presence of variation has significant implications for transmissibility, replication, and immune response [16,182]. For example, a study with DENV-4 isolates collected from Puerto Rico over 20 years, showed that DENV-4 evolution and lineage turnover was driven by selection [183].

In agreement with previous reports [62,71,181], our results demonstrated that natural selection is predominantly negative in DENV-4, which means that natural selection eliminates the majority of newly arising mutants, presumably due to the high frequency of deleterious mutations estimated for dengue [71]. Despite the DENV-4 envelope gene operates under purifying selection, we also provided evidence of positive selection acting on a site-by-site basis in the gene, thereby highlighting the importance of assessing selection pressures at different levels beyond the average of dN/dS [184,185].

In a conservative mode, the most compelling evidence showed signatures of positive selection in the E235K amino acid position, which fell into the domain II of the envelope protein. Interestingly, this mutation was not fixed in the population during the outbreak (**Figure 24)**. Hence, this fact suggested that the E235K site was subject to diversifying selection. The advantage of maintaining this amino-acid diversity lies in an increased capacity of the virus population to confer phenotypic properties that could result advantageous in specific situations [186]. For example, in IVA (H1N1) and HIV, sites subject to diversifying selection are frequently associated with epitopes and responsible for immune evasion [187,188]. Additionally, in DENV a recent report showed diversifying selection associated with in-situ evolution of DENV-2 and DENV-3 in Pakistan [189]. Nonetheless, fixation of mutations after prolonged periods of time have been also reported for DENV [183].

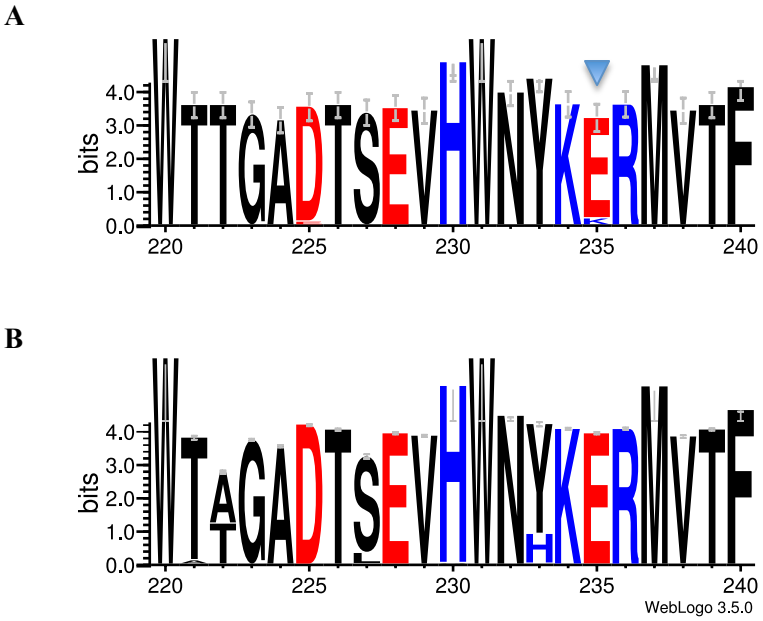


Figure 24. Sequence log representation of DENV-4 viruses. The scheme shows the frequency of the nucleotide position 235 in: (A) Intra-host level (B) Inter-host (consensus) level.

Because the E235K mutation was not detected at inter-host level using a Sanger-derived data set, we initially hypothesized that the mutation disappear at some point while cycling between vector and human hosts. In support of this hypothesis, a variety of studies suggest that different selection pressures are imposed on viral populations infecting vector and human hosts [42,127,190]. Therefore mutations favored in the vector can be selected against in the human host. However, it is important to consider that Sanger-sequences are weak indicatives of population heterogeneity given that only represent the predominant variant present in a sample [60].

Interestingly, the majority of samples included in this study carried very low-frequency haplotypes harboring the E235K mutation. Nonetheless, these haplotypes were filtered out from our analysis because they fell below the threshold. According to Borderia and colleagues [191], low-frequency mutants are selected along dominant mutants by group selection, in order to confer an increase in fitness. This may suggest that haplotypes carrying E235K mutation hold an intermediate fitness which would contribute to preserves the most fit sequence [47]. After detecting the E235K mutation present in four subjects at frequencies >1%, we postulated that this mutation is maintained at low frequencies in the population at intra and inter-host levels.

Domain II exhibits a high diversity and is believed to participate in the folding of the envelope, fusion, and provide a target for the host immune response [192]. In humans, the immune pressure is considered as the major driver of viral diversity even during acute infection [61,66]. Because domain II is highly immunogenic we speculated that the E235K residue might serve as an antigenic determinant. Conversely, after comparing with predicted and previously published epitopes, we identified that the E235K residue possessed non-neutralizing activity [193,194] (**Figure 25**). Nonetheless, the non-synonymous substitution from one negatively charged residue (Glu) to another positively charged (Lys) may be associated with an increase in the pH at which conformational changes occurs and lead to formation of fusion-competent trimmers [157]. Noticeably, this hypothesis deserves further investigation. However, the implications of changes in the pH of fusion have been well documented for other envelope viruses. There is evidence to suggest that pH-dependent viral fusion can affect viral tropism [195]. In influenza virus, the subunity HA2 of the envelope protein mediates the viral fusion. Alterations in its pH of fusion have been associated to the adaptation to different host and determine the establishment of the infection in different organs and tissues [196,197].

In addition, we detected signatures of positive selective pressures acting on NS1 in DENV-1. It is well known that NS1 is a non-structural gene, which is involved in immune evasion, virus replication and dengue infectivity. Given the critical role of the NS1 protein, this evidence may contribute to explain the displacement of DENV-4 and the predominance of DENV-1 in the 2014 and 2015 dengue outbreaks [38,136].

5.7 TRANSMISSION

Villabona-Arenas and colleagues [37,198] documented the epidemiological dynamics of the DENV-4 outbreak during 2013 in Guarujá, São Paulo. The study showed that the region maintained high levels of transmission and several imported cases. Throughout our analysis using haplotypes, we also suggested the circulation of dominant and minor haplotypes, which would be able to promote the high rate of inter-host transmission during the outbreak in Guarujá. Several studies have demonstrated that dominant and minor haplotypes have a sustainable transmission with implications on pathogenesis and treatment failures [149–151], adaptation [191], and population diversity [108,199]. In dengue, the maintenance of minority

haplotypes requires additional study, however evidence so far suggest the association between low-frequency variants and molecular memory [45,200].

As described previously in the haplotype section, stochastic processes influence inter-host transmission. Particularly, virus population bottlenecks during mosquito infection and transmission affect genetic diversity and contribute to fluctuations in haplotype frequencies. Therefore, low-frequency variants can potentially increase or decrease its frequency during transmission. Supporting these observations, Rodriguez-Roche and colleagues [201] showed that minor population became predominant during an epidemic with DENV-3. It is not clear, however, whether all the shared haplotypes result from transmission or eventually correspond to convergent de novo mutants, which arise multiple times during inter-host transmission [108].

Furthermore, our results demonstrated that some haplotypes spread efficiently through time and space scales, demonstrating the importance of intra-host surveillance over inter-host diversity and dynamics [108,199]. Villabona-Arenas and colleagues [198] postulated the neighborhood Pae-Cará as a key transmission loci during the outbreak in Guarujá. This fact could explain in part the widespread distribution of the haplotypes derived from sample GU792. Although, disconnected haplotypes could be reflecting a limited sampling, we identified four haplotypes bearing the mutation E235K, suggesting the effective spreading of the variant. Interestingly, high and low frequency variants shaped a genetic structure in DENV-4. Since multiple factors determine the genetic structure in viral populations, the observed distribution may result from the interplay between genetic drift and natural selection [202]

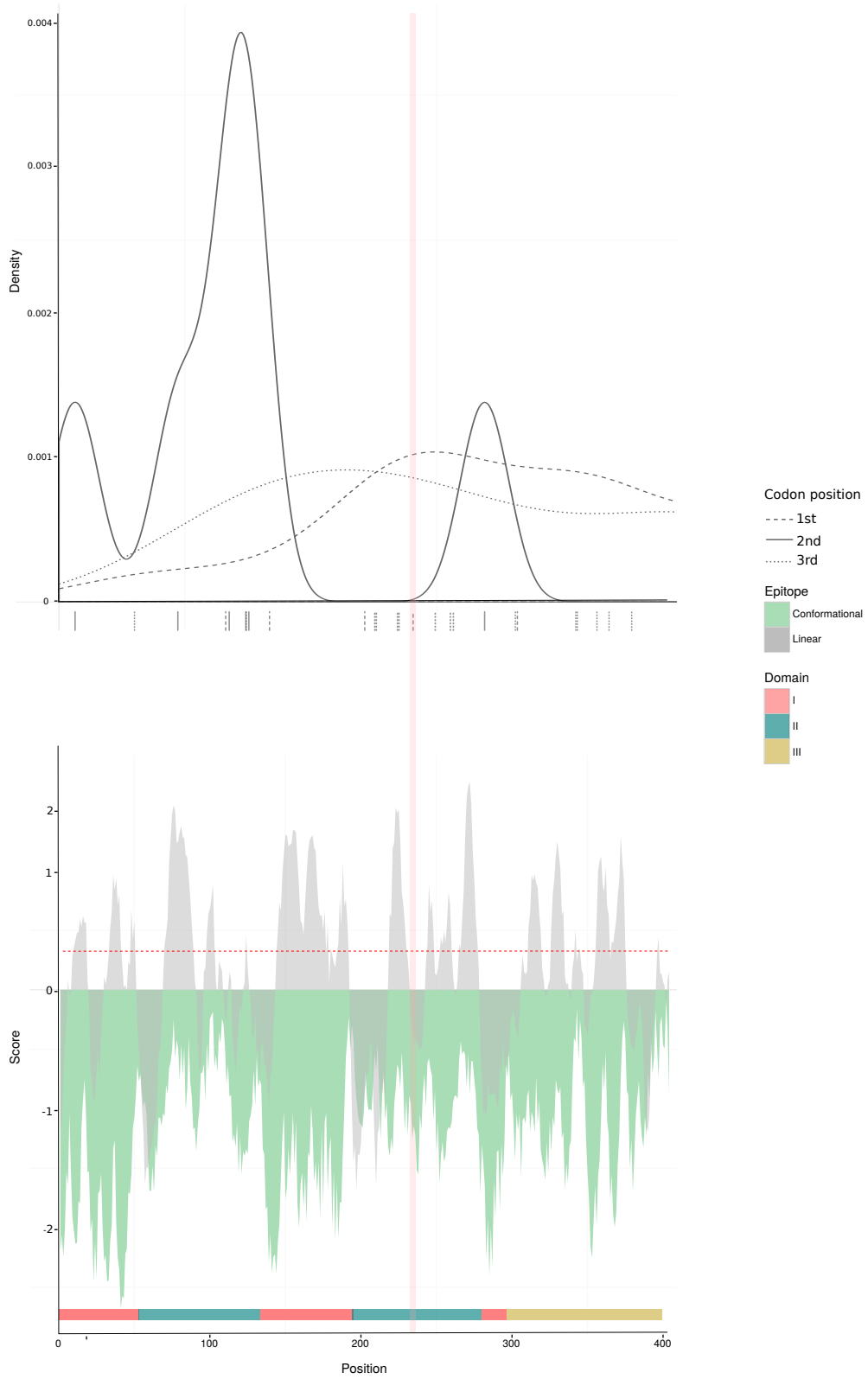


Figure 25 - Density plot of scores for antibody epitope prediction and distribution of polymorphic sites across the envelope gene. Predictions for linear epitopes are displayed in gray. Values above the threshold (red dashed line) correspond to positive predictions for linear epitopes. Predictions for conformational epitopes are displayed in green. Those values close to zero corresponded to positive predictions for conformational epitopes. The selected residue in domain II (position 235) is indicated by the red strip. Line coding represented the distribution of polymorphic sites per codon position.

6 CONCLUSIONS

1) DENV-4 circulating in Guarujá corresponded to the most broadly distributed lineage in Brazil, which was introduced in 2005 into northern Brazil from South American countries and spread widely across several Brazilian locations.

2) Guarujá sequences were grouped in two lineages. Additionally, we identified one isolate from Guarujá closely related with strains from São José de Rio Preto and Mato Grosso.

3) This study provides the first DENV-4 genomes available for the 2013 urban outbreak in Guarujá, São Paulo.

4) Our results corroborate the co-circulation of DENV-1 and DENV-4 and suggest the occurrence of potential mixed infections during the DENV-4 outbreak, in the city of Guarujá, São Paulo.

5) DENV-4 evolves into heterogeneous populations composed of a few dominant haplotypes and a large number of low-frequency haplotypes. Although some of these viral variants can be maintained at the inter-host level their frequencies fluctuate among individuals.

6) Evolution of intra-patient populations of DENV-4 lies on low diversity levels during infection. However, envelope domains I and II display a high variability.

7) We did not find any evidence for intra-host recombination in DENV-4 populations circulating in infected hosts during the 2013 epidemic.

8) Preferred synonymous codons were dominated by A-ended codons, whilst charged amino acids played a major contribution in amino-acid usage in DENV-4 at the intra-host level.

9) Mutational bias and natural selection contributes to the codon usage bias in the envelope of DENV-4 at the intra-host level.

10) Selection pressures imposed on DENV-4 populations at the intra-host level differ between the entire envelope gene and individual sites. Negative selection operates on the gene-wide

scale, while positive selection acting at a per-site level would be influencing pH-dependent postfusion conformational changes.

REFERENCES*

1. Chen R, Vasilakis N. Dengue--quo tu et quo vadis? Viruses. 2011 Sep;3(9):1562–608.

2. Gebhard LG, Filomatori C V, Gamarnik A V. Functional RNA Elements in the Dengue Virus Genome. Viruses. 2011;3:1739–56.

3. Vasilakis N, Cardosa J, Hanley KA, Holmes EC, Weaver SC. Fever from the forest: prospects for the continued emergence of sylvatic dengue virus and its impact on public health. Nat Rev Microbiol. 2011 Jun 13;9(7):532–41.

4. Cruz-Oliveira C, Freire JM, Conceição TM, Higa LM, Castanho MARB, Da Poian AT. Receptors and routes of dengue virus entry into the host cells. FEMS Microbiol Rev. 2015 Mar;39(2):155–70.

5. Crill WD, Chang G-JJ. Localization and Characterization of Flavivirus Envelope Glycoprotein Cross-Reactive Epitopes. J Virol. 2004 Dec 15;78(24):13975–86.

6. Heinz FX, Auer G, Stiasny K, Holzmann H, Mandl C, Guirakhoo F, et al. The interactions of the flavivirus envelope proteins: implications for virus entry and release. In: Positive-Strand RNA Viruses. Vienna: Springer Vienna; 1994. p. 339–48.

7. Modis Y, Ogata S, Clements D, Harrison SC. Structure of the dengue virus envelope protein after membrane fusion. Nature. 2004 Jan 22;427(6972):313–9.

8. Nayak V, Dessau M, Kucera K, Anthony K, Ledizet M, Modis Y. Crystal Structure of Dengue Virus Type 1 Envelope Protein in the Postfusion Conformation and Its Implications for Membrane Fusion. J Virol. 2009 May 1;83(9):4338–44.

9. Murray NEA, Mikkel B Q, Wilder-Smith A. Epidemiology of dengue: past, present and future prospects. Clin Epidemiol. 2013;5:299–309.

10. Weaver SC, Vasilakis N. Molecular evolution of dengue viruses: Contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease. Infect Genet Evol. 2009;9(4):523–40.

11. Raghwani J, Rambaut A, Holmes EC, Hang VT, Hien TT, Farrar J, et al. Endemic Dengue Associated with the Co-Circulation of Multiple Viral Lineages and Localized Density-Dependent Transmission. PLoS Pathog. 2011;7(6):e1002064.

12. Kraemer MUG, Sinka ME, Duda KA, Mylne AQN, Shearer FM, Barker CM, et al. The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. Elife. 2015;4:e08347.

*De acordo com: International Committee of Medical Journal Editors. [Internet]. Uniform requirements for manuscripts submitted to biomedical journals. [2011 Jul 15]. Available from: http://www.nlm.nih.gov/bsd/uniform_requirements.htlm

13. Mustafa MS, Rasotgi V, Jain S, Gupta V. Discovery of fifth serotype of dengue virus (DENV-5): A new public health dilemma in dengue control. Med journal, Armed Forces India. 2015 Jan;71(1):67–70.

14. WHO. Dengue guidelines for diagnosis, treatment, prevention and control. New. Geneva: World Health Organization; 2009. 147 p.

15. Holmes EC. The evolution and emergence of RNA viruses. Oxford University Press; 2009. 288 p. (Oxford series in ecology and evolution).

16. Holmes EC, Burch SS. The causes and consequences of genetic variation in dengue virus. Trends Microbiol. 2000;8:74–7.

17. Villabona-Arenas CJ, de Andrade Zanotto PM. Evolutionary history of Dengue virus type 4: Insights into genotype phylodynamics. Infect Genet Evol. 2011;11(5):878–85.

18. Vasilakis N, Weaver SC. Chapter 1 The History and Evolution of Human Dengue Emergence. In: Karl Maramorosch AJS, Murphy FA, editors. Academic Press; 2008. p. 1–76. (Advances in Virus Research; vol. 72).

19. Foster JE, Bennett SN, Vaughan H, Vorndam V, McMillan WO, Carrington CVF. Molecular evolution and phylogeny of dengue type 4 virus in the {C}aribbean. Virology. 2003;306:129–34.

20. Henchal EA, Repik PM, McCown JM, Brandt WE. Identification of an antigenic and genetic variant of dengue-4 virus from the Caribbean. Am J Trop Med Hyg. 1986 Mar;35(2):393–400.

21. Carrington CVF, Foster JE, Pybus OG, Bennett SN, Holmes EC. Invasion and Maintenance of Dengue Virus Type 2 and Type 4 in the Americas. J Virol. 2005 Dec;79(23):14680–7.

22. Gubler DJ. Dengue, Urbanization and Globalization: The Unholy Trinity of the 21(st) Century. Trop Med Heal. 2011;39(4 Suppl):3–11.

23. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature. 2013 Apr 25;496(7446):504–7.

24. JL SM, Brathwaite-Dick O. Integrated strategy for dengue prevention and control in the Region of the Americas. Rev Panam Salud Publica. 2007;21:55–63.

25. Mart'in S, Luis J, Brathwaite, Olivia, Zambrano, Betzana, et al. The Epidemiology of Dengue in the Americas Over the Last Three Decades: A Worrisome Reality. Am J Trop Med Hyg. 2010;82(1):128–35.

26. Anoop M, Issac A, Mathew T, Philip S, Kareem NA, Unnikrishnan R, et al. Genetic characterization of dengue virus serotypes causing concurrent infection in an outbreak in Ernakulam, Kerala, South India. Indian J Exp Biol. 2010;48:849–57.

27. Guzman MG, Halstead SB, Artsob H, Buchy P, Farrar J, Gubler DJ, et al. Dengue: a

continuing global threat. Nat Rev Microbiol. 2010 Dec;8:S7–16.

28. Halstead SB, O'Rourke EJ. Dengue viruses and mononuclear phagocytes. I. Infection enhancement by non-neutralizing antibody. J Exp Med. 1977;146(1):201–17.

29. Osanai CH, Travassos da Rosa AP, Tang AT, do Amaral RS, Passos AD TP. Dengue outbreak in Boa Vista, Roraima [in Portuguese]. Rev Inst Med Trop Sao Paulo. 1983;(25):53–4.

30. Guzmán MG, Kour G. Dengue diagnosis, advances and challenges. Int J Infect Dis. 2004;8(2):69–80.

31. PAHO W. State of the art in the Prevention and Control of Dengue in the Americas Meeting report. Washington, D.C.; 2014.

32. Instituto de Avaliação de Tecnologia em Saúde (IATS) [Internet]. 2015. Available from: http://www.iats.com.br/?p=noticias&id=388

33. Villabona-Arenas CJ, de Oliveira JL, Capra CDS, Balarini K, Loureiro M, Fonseca CRTP, et al. Detection of four dengue serotypes suggests rise in hyperendemicity in urban centers of Brazil. PLoS Negl Trop Dis. 2014 Mar;8(2):e2620.

34. Vigilância S De. Boletim epidemiológico. São Paulo, Brasil; 2016.

35. Brasil S. Casos de Dengue. Brasil, Regiões e Unidades Federadas, 1990 a 2014. 2015.

36. Organização Pan-Americana da Saúde - Brasil. Dados da dengue no Brasil, 2013.

37. Villabona-Arenas CJ, de Oliveira JL, de Sousa-Capra C, Balarini K, Pereira da Fonseca CRT, Zanotto PM de A. Epidemiological dynamics of an urban Dengue 4 outbreak in São Paulo, Brazil. PeerJ. 2016 Apr 5;4:e1892.

38. Vigilância S De. Boletim epidemiológico. Vol. 46. São Paulo, Brasil; 2015.

39. Messina JP, Brady OJ, Scott TW, Zou C, Pigott DM, Duda KA, et al. Global spread of dengue virus types: mapping the 70 year history. Trends Microbiol. 2014;22(3):138–46.

40. Pybus OG, Rambaut A. Evolutionary analysis of the dynamics of viral infectious disease. Nat Rev Genet. 2009 Aug;10(8):540–50.

41. Temporão JG, Penna GO, Carmo EH, Coelho GE, do Socorro Silva Azevedo R, Nunes MRT, et al. Dengue Virus Serotype 4, Roraima State, Brazil. Vol. 17, Emerg Infect Dis. 2011.

42. Sim S, Hibberd ML, Bhatt S, Gething P, Brady O, Messina J, et al. Genomic approaches for understanding dengue: insights from the virus, vector, and host. Genome Biol. 2016 Dec 2;17(1):38.

43. Andino R, Domingo E. Viral quasispecies. Virology. 2015 Mar 27;479:46–51.

44. Domingo E, Sheldon J, Perales C. Viral quasispecies evolution. Microbiol Mol Biol Rev. 2012 Jun;76(2):159–216.

45. Kurosu T. Quasispecies of dengue virus. 2011;39(4 Suppl):29–36.

46. Moya A, C.Holmes E, González-Candelas F. The population genetics and evolutionary epidemiology of rna viruses. Nat Rev Microbiol. 2004;2:1–11.

47. Lauring AS, Frydman J, Andino R. The role of mutational robustness in RNA virus evolution. Nat Rev Microbiol. 2013 May;11(5):327–36.

48. Andino R, Domingo E. Viral quasispecies. Virology. 2015 May;479–480:46–51.

49. Thai KTD, Henn MR, Zody MC, Tricou V, Nguyet NM, Charlebois P, et al. High-Resolution Analysis of Intrahost Genetic Diversity in Dengue Virus Serotype 1 Infection Identifies Mixed Infections. J Virol. 2012;86(2):835–43.

50. Gubler DJ, Kuno G, Sather GE, Waterman SH. A case of natural concurrent human infection with two dengue viruses. Am J Trop Med Hyg. 1985;34:170–3.

51. Laille M, Deubel V, Sainte-Marie FF. Demonstration of concurrent dengue 1 and dengue 3 infection in six patients by the polymerase chain reaction. J Med Virol. 1991;34(1):51–4.

52. Lorono-Pino MA, Cropp CB, Farfan JA, Vorndam A V, RodriGuez-Angulo EM, Rosado-Paredes EP, et al. Common Occurrence Of Concurrent Infections By Multiple Dengue Virus Serotypes. Am J Trop Med Hyg. 1999;61(5):725–30.

53. Fatima Z, Idrees M, Bajwa M, Tahir Z, Ullah O, Zia M, et al. Serotype and genotype analysis of dengue virus by sequencing followed by phylogenetic analysis using samples from three mini outbreaks-2007-2009 in Pakistan. BMC Microbiol. 2011;11(1):200.

54. VinodKumar CS, Kalapannavar NK, Basavarajappa KG, Sanjay D, Gowli C, Nadig NG, et al. Episode of coexisting infections with multiple dengue virus serotypes in central Karnataka, India. J Infect Public Health. 2013;6(4):302–6.

55. Wang WK, Chao DY, Lin SR, Chang SC. Concurrent infections by two dengue virus serotypes among dengue patients in Taiwan. J Microbiol Immunol Infect. 2003;36:89–95.

56. Romano CM, Lauck M, Salvador FS, Lima CR, Villas-Boas LS, Araújo ESA, et al. Inter- and Intra-Host Viral Diversity in a Large Seasonal DENV2 Outbreak. PLoS One. 2013;8(8):e70318.

57. Alfson KJ, Beadles MW, Griffiths A. A new approach to determining whole viral genomic sequences including termini using a single deep sequencing run. J Virol Methods. 2014 Nov;208:1–5.

58. Acevedo A, Andino R. Library preparation for highly accurate population sequencing

of RNA viruses. Nat Protoc. 2014 Jun;9(7):1760–9.

59. Archer J, Baillie G, Watson S, Kellam P, Rambaut A, Robertson D. Analysis of high-depth sequence data for studying viral diversity: a comparison of next generation sequencing platforms using Segminator II. BMC Bioinformatics. 2012;13(1):47.

60. Wright CF, Morelli MJ, Thebaud G, Knowles NJ, Herzyk P, Paton DJ, et al. Beyond the Consensus: Dissecting Within-Host Viral Population Diversity of Foot-and-Mouth Disease Virus by Using Next-Generation Genome Sequencing. J Virol. 2011 Mar 1;85(5):2266–75.

61. Hapuarachchi HC, Koo C, Kek R, Xu H, Lai YL, Liu L, et al. Intra-epidemic evolutionary dynamics of a Dengue virus type 1 population reveal mutant spectra that correlate with disease transmission. Sci Rep. 2016 Mar 4;6:22592.

62. Lequime S, Fontaine A, Ar Gouilh M, Moltini-Conclois I, Lambrechts L, Lauring A, et al. Genetic Drift, Purifying Selection and Vector Genotype Shape Dengue Virus Intra-host Genetic Diversity in Mosquitoes. Malik HS, editor. PLOS Genet. 2016 Jun 15;12(6):e1006111.

63. Aw P, de Sessions P, Wilm A, Hoang L, Nagarajan N, Sessions O, et al. Next-Generation Whole Genome Sequencing of Dengue Virus. In: Padmanabhan R, Vasudevan SG, editors. Dengue SE - 12. Springer New York; 2014. p. 175–95. (Methods in Molecular Biology; vol. 1138).

64. Sessions OM, Wilm A, Kamaraj US, Choy MM, Chow A, Chong Y, et al. Analysis of Dengue Virus Genetic Diversity during Human and Mosquito Infection Reveals Genetic Constraints. Armstrong PM, editor. PLoS Negl Trop Dis. 2015 Sep 1;9(9):e0004044.

65. Aaskov J, Buzacott K, Field E, Lowry K, Berlioz-Arthaud A, Holmes EC. Multiple recombinant dengue type 1 viruses in an isolate from a dengue patient. J Gen Virol. 2007;88(12):3334–40.

66. Parameswaran P, Charlebois P, Tellez Y, Nunez A, Ryan EM, Malboeuf CM, et al. Genome - Wide Patterns of Intrahuman Dengue Virus Diversity Reveal Associations with Viral Phylogenetic Clade and Interhost Diversity. J Virol. 2012;86(16):8546–58.

67. Descloux E, La Fuentez C, Roca Y, De Lamballerie X. Clinical significance of intra-host variability of Dengue-1 virus in venous and capillary blood. Clin Microbiol Infect. 2014 Mar;20(3):O167-75.

68. Ke R, Aaskov J, Holmes EC, Lloyd-Smith JO. Phylodynamic Analysis of the Emergence and Epidemiological Impact of Transmissible Defective Dengue Viruses. PLoS Pathog. 2013;9(2):e1003193.

69. Lin S-R, Hsieh S-C, Yueh Y-Y, Lin T-H, Chao D-Y, Chen W-J, et al. Study of Sequence Variation of Dengue Type 3 Virus in Naturally Infected Mosquitoes and Human Hosts: Implications for Transmission and Evolution. J Virol. 2004;78(22):12717–21.

70. Bennett SN, Holmes EC, Chirivella M, Rodriguez DM, Beltran M, Vorndam V, et al. Molecular evolution of dengue 2 virus in Puerto Rico: positive selection in the viral envelope accompanies clade reintroduction. J Gen Virol. 2006;87:885–93.

71. Holmes EC. Patterns of Intra- and Interhost Nonsynonymous Variation Reveal Strong Purifying Selection in Dengue Virus. J Virol. 2003;77(20):11296–8.

72. Descloux E, Cao-Lormeau V-M, Roche C, Lamballerie X De. Dengue 1 diversity and microevolution, French Polynesia 2001-2006: connection with epidemiology and clinics. {PLoS} Neglected Trop Dis. 2009;3(8):e493.

73. Edgar R. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5(1):113.

74. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460–1.

75. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics. 2010 Oct;26(19):2462–3.

76. Schmidt HA, Strimmer K, Vingron M, von Haeseler A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics. 2002 Mar 1;18(3):502–4.

77. Strimmer K, von Haeseler A. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc Natl Acad Sci U S A. 1997 Jun 24;94(13):6815–9.

78. Rambaut A, Lam TT, Carvalho LM, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016;2.

79. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012 Aug;29(8):1969–73.

80. Posada D. jModelTest: phylogenetic model averaging. Mol Biol Evol. 2008 Jul;25(7):1253–6.

81. Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol Biol Evol. 2008 Jul;25(7):1459–71.

82. Rambaut A, Drummond AJ. Tracer v1.4. Available from http://beast.bio.ed.ac.uk/Tracer. Available from http://beast.bio.ed.ac.uk/Tracer. 2007.

83. Rambaut A, Drummond A. FigTree. 2009.

84. Lanciotti RS, Calisher CH, Gubler DJ, Chang GJ, Vorndam A V. Rapid detection and typing of dengue viruses from clinical samples by using reverse transcriptase-

polymerase chain reaction. J Clin Microbiol. 1992 Mar;30(3):545–51.

85. Andrews S. FastQC: a quality control tool for high throughput sequence data. [Internet]. 2010. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

86. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol. 2012 May;19(5):455–77.

87. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct;215(3):403–10.

88. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Meth. 2012 Apr;9(4):357–9.

89. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Current Protocols in Bioinformatics. 2013. 1-33 p.

90. Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 2012 Dec;40(22):11189–201.

91. Töpfer A, Zagordi O, Prabhakaran S, Roth V, Halperin E, Beerenwinkel N. Probabilistic Inference of Viral Quasispecies Subject to Recombination. J Comput Biol. 2013 Feb 1;20(2):113–23.

92. Huson DH, Bryant D. Application of Phylogenetic Networks in Evolutionary Studies. Mol Biol Evol. 2006; 23(2):254-267.

93. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics. 2010;26(19):2462–3.

94. Jombart T, Cori A, Didelot X, Cauchemez S, Fraser C, Ferguson N, et al. Bayesian Reconstruction of Disease Outbreaks by Combining Epidemiologic and Genomic Data. Tanaka MM, editor. PLoS Comput Biol. 2014 Jan 23;10(1):e1003457.

95. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: The R foundation for Statistical Computing; 2011.

96. Peden JF. Analysis of Codon Usage. University of Nottingham; 1999.

97. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011 Aug 15;27(16):2194–200.

98. Pond SLK, Frost SDW, Muse S V. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 2005;21:676–9.

99. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010 May;59(3):307–21.

100. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 2007 Aug;24(8):1596–9.

101. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989 Nov;123(3):585–95.

102. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009 Jun 1;25(11):1451–2.

103. Jombart T, Bateman A. adegenet: a R package for the multivariate analysis of genetic markers. 2008;24(11):1403–5.

104. Amarilla AA, de Almeida FT, Jorge DM, Alfonso HL, de Castro-Jorge LA, Nogueira NA, et al. Genetic diversity of the E Protein of Dengue Type 3 Virus. Virol J. 2009;6(1):113.

105. Anderson R, King AD, Innis BL. Correlation of E protein binding with cell susceptibility to dengue 4 virus infection. J Gen Virol. 1992 Aug 1;73(8):2155–9.

106. Monitoramento dos casos de dengue e febre de chikungunya até a semana epidemiológica 15. 2015.

107. Lemey P, Rambaut A, Pybus OG. HIV evolutionary dynamics within and among hosts. AIDS Rev. 8(3):125–40.

108. Stack JC, Murcia PR, Grenfell BT, Wood JLN, Holmes EC, Chen R, et al. Inferring the inter-host transmission of influenza A virus using patterns of intra-host genetic variation. Proc Biol Sci. 2013 Jan 7;280(1750):20122173.

109. Allicock OM, Lemey P, Tatem AJ, Pybus OG, Bennett SN, Mueller BA, et al. Phylogeography and Population dynamics of Dengue viruses in the Americas. Mol Biol Evol. 2012;29(6):1533–43.

110. Teixeira Nunes RM, Nuno Rodrigues F, Vasconcelos Baldez H, Barbosa de Almeida Medeiros, Daniele; Pereira Silva de Lima, Clayton; Carvalho Lima V, Vieira Pinto da Silva E, Ferreira Cardoso, Jedson; costa Sousa Jr., Edivaldo; Nascimento Barbosa Nunes K, et al. Phylogeography of Dengue Virus Serotype 4, Brazil, 2010–2011. Emerg Infect Dis. 2012;18(November):1858–64.

111. de Melo FL, Romano CM, de Andrade Zanotto PM. Introduction of dengue virus 4 (DENV-4) genotype I into Brazil from Asia? PLoS Negl Trop Dis. 2009 Jan;3(4):e390.

112. de Figueiredo RMP, Naveca FG, de Souza Bastos M, do Nascimento Melo M, de Souza Viana S, Mourão MPG, et al. Dengue Virus Type 4, Manaus, Brazil. Emerg Infect Dis. 2008;14:667–9.

113. de Figueiredo MLG, Alfonso HL, Amarilla AA, Figueiredo LTM, Aquino VH, da Costa CA, et al. Detection of DENV-4 genotype I from mosquitoes collected in the city of Manaus, Brazil. Virol J. 2013;10(1):60.

114. de Souza RP, Rocco IM, Maeda AY, Spenassatto C, Bisordi I, Suzuki A, et al. Dengue Virus Type 4 Phylogenetics in {B}razil 2011: Looking beyond the Veil. PLoS Negl Trop Dis. 2011;5(12):e1439.

115. T Nunes MR, Palacios G, Rodrigues Faria N, Costa Sousa Jr E, Pantoja JA, Rodrigues SG, et al. Air Travel Is Associated with Intracontinental Spread of Dengue Virus Serotypes 1–3 in Brazil. PLoS Negl Trop Dis. 2014;8(4).

116. Carvalho LM, Faria NR, Perez AM, Suchard MA, Lemey P, Silveira W de C, et al. Spatio-temporal Dynamics of Foot-and-Mouth Disease Virus in South America. 2015.

117. Azevedo R do S da S, Oliveira CS, Vasconcelos PF da C, Azevedo R do S da S, Oliveira CS, Vasconcelos PF da C. Chikungunya risk for Brazil. Rev Saude Publica. 2015;49:1–6.

118. Araújo JMG de, Bello G, Romero H, Nogueira RMR. Origin and Evolution of Dengue Virus Type 3 in Brazil. PLoS Negl Trop Dis. 2012;6(9):e1784.

119. Bertolacci-Rocha LG, Cunha RV da, Lichs GG de C, Fabbro MMFJD, Motta-Castro ARC, Bertolacci-Rocha LG, et al. Introduction of the dengue virus type 4 in the State of Mato Grosso do Sul, Brazil. Cad Saude Publica. 2014 Aug;30(8):1789–92.

120. Superintendência de Vigilância em Saúde (SUVISA). Isolamento do vírus DENV 4 no estado de Goiás. Goiás; 2011.

121. Nogueira RM, Eppinghaus AL. Dengue virus type 4 arrives in the state of Rio de Janeiro: a challenge for epidemiological surveillance and control. Mem Inst Oswaldo Cruz. 2011 May;106(3):255–6.

122. Gonçalves De Castro M, Maria R, Nogueira R, Bispo De Filippis AM, Ferreira AA, Da M, et al. Dengue virus type 4 in Niterói, Rio de Janeiro: the role of molecular techniques in laboratory diagnosis and entomological surveillance. Mem Inst Oswaldo Cruz Rio Janeiro. 2012;107(7):940–5.

123. Amâncio FF, Pereira MA, Iani FC de M, D'anunciação L, Almeida JLC de, Soares JAS, et al. Fatal outcome of infection by dengue 4 in a patient with thrombocytopenic purpura as a comorbid condition in Brazil. Rev Inst Med Trop Sao Paulo. 2014 Jun;56(3):267–70.

124. Ho T, Tzanetakis IE. Development of a virus detection and discovery pipeline using next generation sequencing. Virology. 2014;471–473:54–60.

125. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016 May 17;17(6):333–51.

126. Marston D, McElhinney L, Ellis R, Horton D, Wise E, Leech S, et al. Next generation sequencing of viral RNA genomes. BMC Genomics. 2013;14(1):444.

127. Sim S, Aw PPK, Wilm A, Teoh G, Hue KDT, Nguyen NM, et al. Tracking Dengue Virus Intra-host Genetic Diversity during Human-to-Mosquito Transmission. PLoS Negl Trop Dis. 2015;9(9):e0004052.

128. Walker JM. Dengue: Methods and Protocols. Methods in Molecular Biology. 2014. 403 p.

129. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, et al. Target-enrichment strategies for next-generation sequencing. Nat Methods. 2010 Feb;7(2):111–8.

130. Djikeng A, Halpin R, Kuzmickas R, DePasse J, Feldblyum J, Sengamalay N, et al. Viral genome sequencing by random priming methods. BMC Genomics. 2008;9(1):5.

131. Amâncio FF, Ferraz ML, de M. Almeida MC, Pessanha JEM, Iani FCM, Fraga GL, et al. Dengue virus serotype 4 in a highly susceptible population in Southeast Brazil. J Infect Public Health. 2014;7(6):547–52.

132. Heinen LB da S, Zuchi N, Cardoso BF, dos Santos MAM, Nogueira ML, Dezengrini-Slhessarenko R. Dengue outbreak in Mato Grosso state, midwestern Brazil. Rev Inst Med Trop Sao Paulo. 2015 Dec;57(6):489–96.

133. Granberg F, Vicente-Rubiano M, Rubio-Guerri C, Karlsson OE, Kukielka D, Belák S, et al. Metagenomic Detection of Viral Pathogens in Spanish Honeybees: Co-Infection by Aphid Lethal Paralysis, Israel Acute Paralysis and Lake Sinai Viruses. Kapoor A, editor. PLoS One. 2013 Feb 27;8(2):e57459.

134. Colombo TE, Vedovello D, Mondini A, Reis AFN, Cury AAF, Oliveira FH de, et al. Co-infection of dengue virus by serotypes 1 and 4 in patient from medium sized city from Brazil. Rev Inst Med Trop Sao Paulo. 2013 Jul;55(4):275–81.

135. Andrade EHP, Figueiredo LB, Vilela APP, Rosa JCC, Oliveira JG, Zibaoui HM, et al. Spatial-Temporal Co-Circulation of Dengue Virus 1, 2, 3, and 4 Associated with Coinfection Cases in a Hyperendemic Area of Brazil: A 4-Week Survey. Am J Trop Med Hyg. 2016 May 4;94(5):1080–4.

136. Vigilância S De. Boletim epidemiológico. Vol. 45. São Paulo, Brasil; 2014.

137. Zhang C, Mammen Jr. MP, Chinnawirotpisan P, Klungthong C, Rodpradit P, Monkongdee P, et al. Clade Replacements in Dengue Virus Serotypes 1 and 3 Are Associated with Changing Serotype Prevalence. J Virol. 2005;79(24):15123–30.

138. Adams B, Holmes EC, Zhang C, M. P. Mammen J, Nimmannitya S, Kalayanarooj S, et al. Cross-protective immunity can account for the alternating epidemic pattern of dengue virus serotypes circulating in Bangkok. PNAS. 2006;103 (38):14234–9.

139. Kasibhatla SM, Waman VP, Kale MM, Kulkarni-Kale U. Analysis of Next-generation

Sequencing Data in Virology - Opportunities and Challenges. In: Next Generation Sequencing - Advances, Applications and Challenges. InTech; 2016.

140. Quiñones-Mateu ME, Avila S, Reyes-Teran G, Martinez MA, Holley RW, Apgar J, et al. Deep sequencing: becoming a critical tool in clinical virology. J Clin Virol. 2014 Sep;61(1):9–19.

141. Klein DE, Choi JL, Harrison SC. Structure of a Dengue Virus Envelope Protein Late-Stage Fusion Intermediate. J Virol. 2013 Feb 15;87(4):2287–93.

142. Rey FA. Dengue virus envelope glycoprotein structure: new insight into its interactions during viral entry. Proc Natl Acad Sci U S A. 2003 Jun 10;100(12):6899–901.

143. Chilamakuri CS, Lorenz S, Madoui M-A, Vodák D, Sun J, Hovig E, et al. Performance comparison of four exome capture systems for deep sequencing. BMC Genomics. 2014;15(1):449.

144. Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res. 2008 Sep;36(16):e105.

145. Van den Hoecke S, Verhelst J, Saelens X, Baets S De, Eckert N, Kim JI, et al. Illumina MiSeq sequencing disfavours a sequence motif in the GFP reporter gene. Sci Rep. 2016 May 19;6:26314.

146. Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral Mutation Rates. J Virol. 2010 Oct 1;84(19):9733–48.

147. Mangul S, Wu NC, Mancuso N, Zelikovsky A, Sun R, Eskin E. Accurate viral population assembly from ultra-deep sequencing data. Bioinformatics. 2014 Jun 15;30(12):i329–37.

148. Töpfer A, Höper D, Blome S, Beer M, Beerenwinkel N, Ruggli N, et al. Sequencing approach to analyze the role of quasispecies for classical swine fever. Virology. 2013 Mar 30;438(1):14–9.

149. Metzner KJ, Giulieri SG, Knoepfel SA, Rauch P, Burgisser P, Yerly S, et al. Minority Quasispecies of Drug-Resistant HIV-1 That Lead to Early Therapy Failure in Treatment-Naive and -Adherent Patients. Clin Infect Dis. 2009 Jan 15;48(2):239–47.

150. Capobianchi MR, Giombini E, Rozera G. Next-generation sequencing technology in clinical virology. Clin Microbiol Infect. 2013;19(1):15–22.

151. Chabria SB, Gupta S, Kozal MJ. GG15CH13-Kozal Deep Sequencing of HIV: Clinical and Research Applications. Annu Rev Genomics Hum Genet. 2014;15:295–325.

152. Forrester NL, Coffey LL, Weaver SC. Arboviral bottlenecks and challenges to maintaining diversity and fitness during mosquito transmission. Viruses. 2014 Oct;6(10):3991–4004.

153. Villabona-Arenas CJ, de Brito AF, de Andrade Zanotto PM. Genomic mosaicism in two strains of dengue virus type 3. Infect Genet Evol. 2013;18(0):202–12.

154. AbuBakar S, Wong P-F, Chan Y-F. Emergence of dengue virus type 4 genotype IIA in Malaysia. J Gen Virol. 2002;83(10):2437–42.

155. Worobey M, Rambaut A, Holmes EC. Widespread intraserotype recombination in natural populations of dengue virus. Proc Natl Acad Sci USA. 1999;96:7352–7.

156. Chao D-Y. Inter- and Intra-Host Evolution of Dengue Viruses and the Inference to the Pathogenesis. Viral Genomes - Mol Struct Divers Gene Expr Mech Host-Virus Interact. 2012;195–216.

157. Modis Y, Ogata S, Clements D, Harrison SC. A ligand-binding pocket in the dengue virus envelope glycoprotein. Proc Natl Acad Sci USA. 2003;10(100):6986–91.

158. Descloux E, Cao-Lormeau V-M, Roche C, De Lamballerie X. Dengue 1 Diversity and Microevolution, French Polynesia 2001–2006: Connection with Epidemiology and Clinics. PLoS Negl Trop Dis. 2009;3(8):e493.

159. Descloux E, Cao-Lormeau V-M, Roche C, De Lamballerie X. Dengue 1 diversity and microevolution, French Polynesia 2001-2006: connection with epidemiology and clinics. PLoS Negl Trop Dis. 2009;3(8):e493.

160. Aaskov J, Buzacott K, Thu HM, Lowry K, Holmes EC. Long-Term Transmission of Defective RNA Viruses in Humans and Aedes Mosquitoes. Science (80-). 2006 Jan;311:236–8.

161. Beerenwinkel N, Günthard HF, Roth V, Metzner KJ. Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. Front Microbiol. 2012;3(329).

162. Pirooznia M, Kramer M, Parla J, Goes FS, Potash JB, McCombie W, et al. Validation and assessment of variant calling pipelines for next-generation sequencing. Hum Genomics. 2014;8(1):14.

163. Lauring AS, Acevedo A, Cooper SB, Andino R. Codon Usage Determines the Mutational Robustness, Evolutionary Capacity, and Virulence of an RNA Virus. Cell Host Microbe. 2012;12(5):623–32.

164. Lara-Ramírez EE, Salazar MI, López-López M de J, Salas-Benito JS, Sánchez-Varela A, Guo X, et al. Large-scale genomic analysis of codon usage in dengue virus and evaluation of its phylogenetic dependence. Biomed Res Int. 2014;2014:851425.

165. Schubert AM, Putonti C. Evolution of the sequence composition of Flaviviruses. Infect Genet Evol. 2010;10(1):129–36.

166. Xiang H, Zhang R, Butler RR, Liu T, Zhang L, Pombert J-F, et al. Comparative Analysis of Codon Usage Bias Patterns in Microsporidian Genomes. Ling E, editor. PLoS One. 2015 Jun 9;10(6):e0129223.

167. Carbone A. Codon Bias is a Major Factor Explaining Phage Evolution in Translationally Biased Hosts. J Mol Evol. 2008;66(3):210–23.

168. Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet. 2011 Jan;12(1):32–42.

169. Butt AM, Nasrullah I, Tong Y. Genome-wide analysis of codon usage and influencing factors in chikungunya viruses. PLoS One. 2014;9(3):e90905.

170. Jenkins GM, Holmes EC. The extent of codon usage bias in human RNA viruses and its evolutionary origin. Virus Res. 2003;92:1–7.

171. Pal S, Banerjee A, Das PS, Pati BR, Mandal KC, Chakraborty B, et al. Dengue virus type 4 evolution and genomics: A bioinformatic approach. Indian J Biotechnol. 2014;13:330–6.

172. Shi S-L, Jiang Y-R, Liu Y-Q, Xia R-X, Qin L. Selective pressure dominates the synonymous codon usage in parvoviridae. Virus Genes. 2013 Feb 21;46(1):10–9.

173. Behura SK, Severson DW, Kyle J, Harris E, Gubler D, Ramanathan M, et al. Nucleotide substitutions in dengue virus serotypes from Asian and American countries: insights into intracodon recombination and purifying selection. BMC Microbiol. 2013;13(1):37.

174. Zhou J, Zhang J, Sun D, Ma Q, Chen H, Ma L, et al. The distribution of synonymous codon choice in the translation initiation region of dengue virus. PLoS One. 2013;8(10):e77239.

175. Das S, Nandy A. Phylogenetic and Genetic Analysis of Envelope Gene of the Prevalent Dengue Serotypes in India in Recent Years. Proceedings. 2015;1(Section B):1–9.

176. Behura S, Severson D. Nucleotide substitutions in dengue virus serotypes from Asian and American countries: insights into intracodon recombination and purifying selection. BMC Microbiol. 2013;13(1):37.

177. Dowd KA, DeMaso CR, Pierson TC. Genotypic Differences in Dengue Virus Neutralization Are Explained by a Single Amino Acid Mutation That Modulates Virus Breathing. MBio. 2015 Dec 31;6(6):e01559-15.

178. Schuh AJ, Tesh RB, Barrett ADT. Genetic characterization of Japanese encephalitis virus genotype II strains isolated from 1951 to 1978. J Gen Virol. 2011 Mar;92(Pt 3):516–27.

179. Shah M, Wadood A, Rahman Z, Husnain T, Erum K, Mehreen K, et al. Interaction and Inhibition of Dengue Envelope Glycoprotein with Mammalian Receptor DC-Sign, an In-Silico Approach. Agarwal PK, editor. PLoS One. 2013 Mar 18;8(3):e59211.

180. Grubaugh ND, Weger-Lucarelli J, Murrieta RA, Fauver JR, Garcia-Luna SM, Prasad AN, et al. Genetic Drift during Systemic Arbovirus Infection of Mosquito Vectors

Leads to Decreased Relative Fitness during Host Switching. Cell Host Microbe. 2016 Apr;19(4):481–92.

181. Romano CM, de Matos AM, Araújo ESA, Villas-Boas LS, da Silva WC, Oliveira OMNPF, et al. Characterization of Dengue Virus Type 2: New Insights on the 2010 Brazilian Epidemic. PLoS One. 2010;5(7):e11811.

182. Twiddy SS, Farrar JJ, Vinh Chau N, Wills B, Gould EA, Gritsun T, et al. Phylogenetic relationships and differential selection pressures among genotypes of dengue-2 virus. Virology. 2002;298(1):63–72.

183. Bennett SN, Holmes EC, Chirivella M, Rodriguez DM, Beltran M, Vorndam V, et al. Selection-driven evolution of emergent dengue virus. Mol Biol Evol. 2003;20(10):1650–8.

184. Kosakovsky Pond SL, Frost SDW. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. Mol Biol Evol. 2005 Feb 9;22(5):1208–22.

185. Sironi M, Cagliani R, Forni D, Clerici M. Evolutionary insights into host–pathogen interactions from mammalian sequence data. Nat Rev Genet. 2015 Mar 18;16(4):224–36.

186. Pond SLK, Frost SDW, Muse S V. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 2005 Mar 1;21(5):676–9.

187. Lee AJ, Das SR, Wang W, Fitzgerald T, Pickett BE, Aevermann BD, et al. Diversifying Selection Analysis Predicts Antigenic Evolution of 2009 Pandemic H1N1 Influenza A Virus in Humans. García-Sastre A, editor. J Virol. 2015 May 15;89(10):5427–40.

188. Ross HA, Rodrigo AG. Immune-Mediated Positive Selection Drives Human Immunodeficiency Virus Type 1 Molecular Variation and Predicts Disease Duration. J Virol. 2002 Nov 15;76(22):11715–20.

189. Koo C, Nasir A, Hapuarachchi H, Lee K-S, Hasan Z, Ng L-C, et al. Evolution and heterogeneity of multiple serotypes of Dengue virus in Pakistan, 2006–2011. Virol J. 2013;10(1):275.

190. Sessions OM, Barrows NJ, Souza-Neto JA, Robinson TJ, Hershey CL, Rodgers MA, et al. Discovery of insect and human dengue virus host factors. Nature. 2009;458(7241):1047–50.

191. Bordería A V, Isakov O, Moratorio G, Henningsson R, Agüera-González S, Organtini L, et al. Group Selection and Contribution of Minority Variants during Virus Adaptation Determines Virus Fitness and Phenotype. PLoS Pathog. 2015 May;11(5):e1004838.

192. Choudhury MA, Lott WB, Banu S, Cheng AY, Teo Y-Y, Twee R, et al. Nature and Extent of Genetic Diversity of Dengue Viruses Determined by 454 Pyrosequencing.

PLoS One. 2015;10(11):e0142473.

193. Tang C-T, Liao M-Y, Chiu C-Y, Shen W-F, Chiu C-Y, Cheng P-C, et al. Generation of Monoclonal Antibodies against Dengue Virus Type 4 and Identification of Enhancing Epitopes on Envelope Protein. Mantis NJ, editor. PLoS One. 2015 Aug 26;10(8):e0136328.

194. Howard CB. Identification of epitopes on the Dengue virus type 4 envelope glycoprotein involved in neutralisation by antibodies. [Ph. D. thesis]. Fac Sci Technol Sch Life Sci. Queensland University of Technology; 2006.

195. Morizono K, Chen ISY. Receptors and tropisms of envelope viruses. Curr Opin Virol. 2011;1(1):13–8.

196. Mair CM, Ludwig K, Herrmann A, Sieben C. Receptor binding and pH stability - How influenza A virus hemagglutinin affects host-specific virus infection. Biochim Biophys Acta - Biomembr. 2014;1838(4):1153–68.

197. Schrauwen EJ, Fouchier RA. Host adaptation and transmission of influenza A viruses in mammals. Emerg Microbes Infect. 2014;3(0):e9.

198. Villabona-Arenas C. Phylogeography of the 2013 urban outbreak of dengue virus in Guarujá, São Paulo. São Paulo. [Ph. D. thesis (Microbiology)]. São Paulo: Instituto de Ciências Biomédicas, Universidade de São Paulo; 2014.

199. Bull RA, Eden J-S, Luciani F, McElroy K, Rawlinson WD, White PA. Contribution of Intra- and Interhost Dynamics to Norovirus Evolution. J Virol. 2012 Mar 15;86(6):3219–29.

200. Kinoshita H, Mathenge EGM, Hung NT, Huong VTQ, Kumatori A, Yu F, et al. Isolation and characterization of two phenotypically distinct dengue type-2 virus isolates from the same dengue hemorrhagic Fever patient. Jpn J Infect Dis. 2009 Sep;62(5):343–50.

201. Rodriguez-Roche R, Blanc H, Bordería A V, Díaz G, Gonzalez D, Santana E, et al. Increasing clinical severity during a dengue virus type 3 Cuban epidemic: deep 1 sequencing of evolving viral populations 2 Running title: Evolving DENV-3 populations, Cuba, 2001-2002 3. J Virol. 2016;14(90):4320–33.

202. García-Arenal F, Fraile A, Malpica J. Variability and genetic structure of plant virus populations. Annu rev Phytopathol. 2001;39:157–86.

APPENDIXES

FIGURES

Figure S1 - ML phylogenetic tree showing the grouping of DENV-4 sequences from Guarujá into genotype II. Asterisk indicates a support value >70% for genotype II. The phylogenetic tree was reconstructed based on the envelope gene.

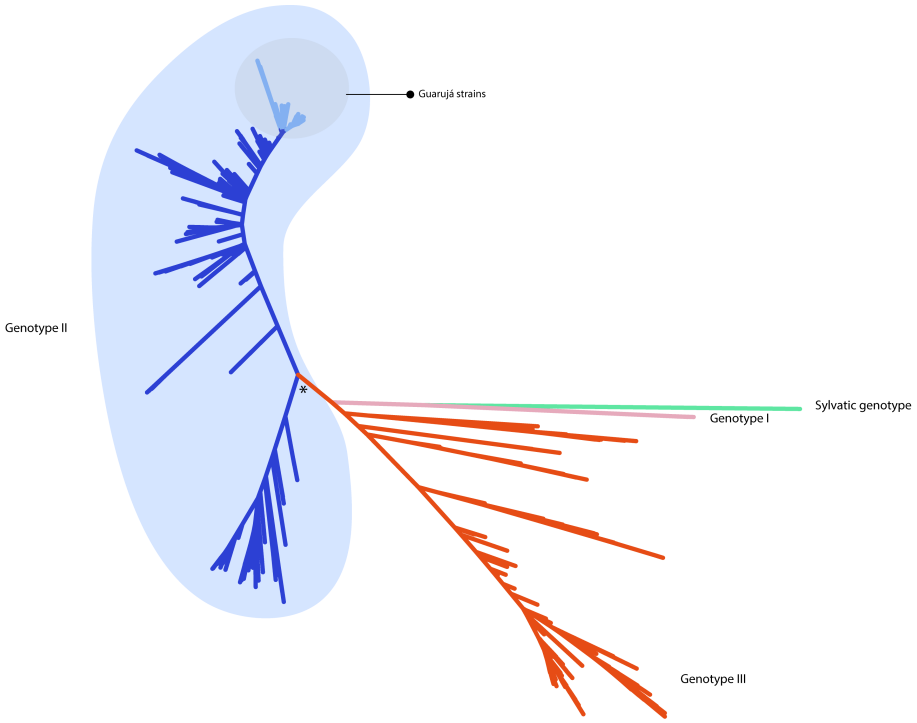


Figure S2 - NGS library quantification with qPCR. (A) Amplification curves using six plasmid DNA standards (10-fold dilutions); all our samples fell between the first and third plasmid standards (B) Dissociation analysis (melt curve). (C) Standard curve.

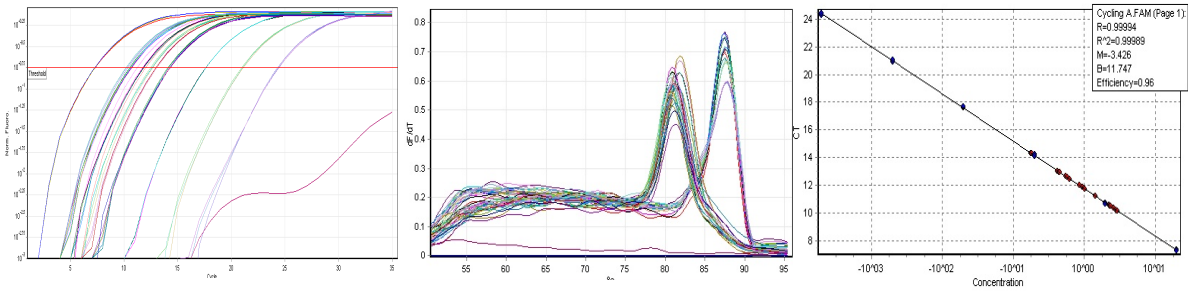


Figure S3 - Electropherograms of genetic material used as input in Nextera XT library preparation. A) Fragmentation of cDNA samples B) Fragmentation of amplicons.

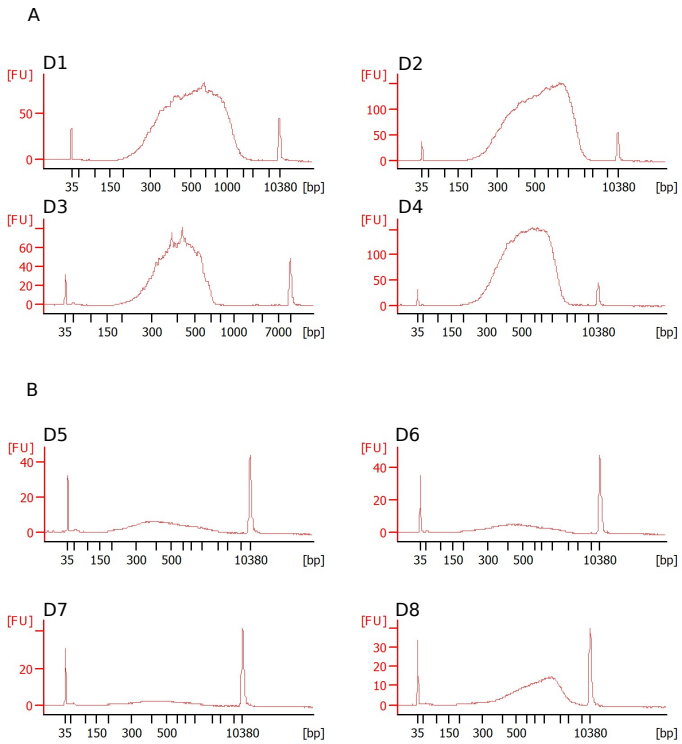


Figure S4 - General profiles of per base sequence quality scores for DENV sequences.

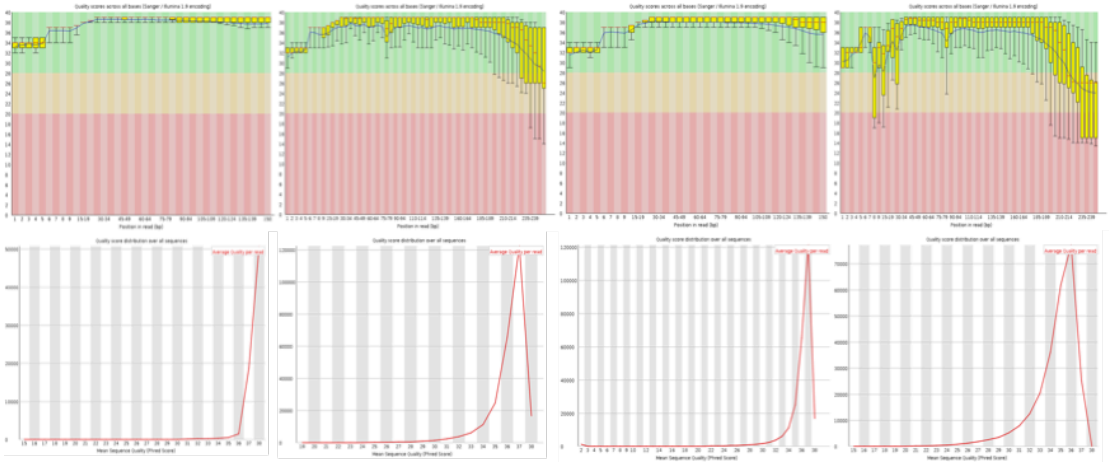


Figure S5 - General profile of base recalibration in DENV-4 sequences.

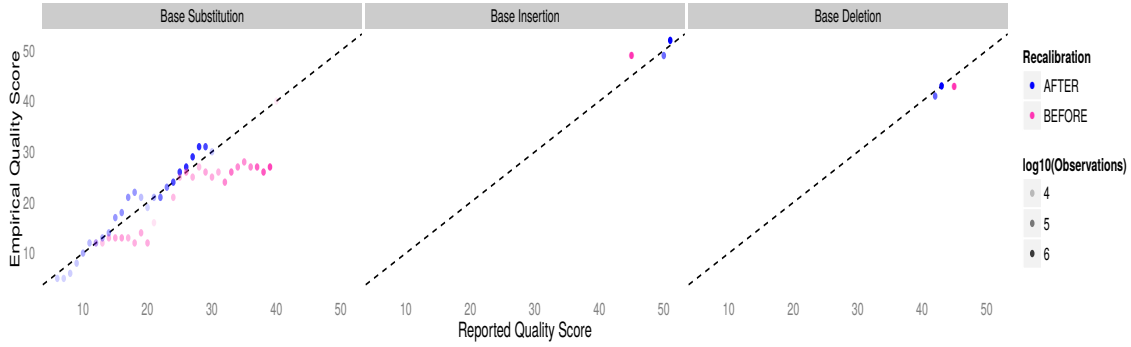
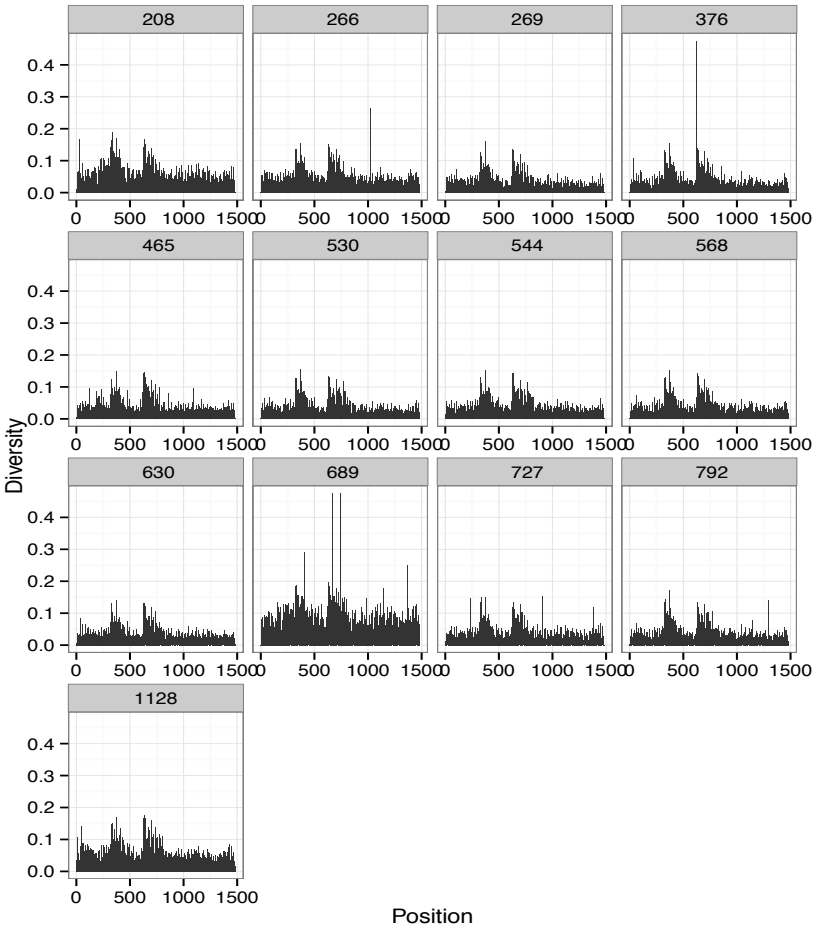


Figure S6 - Diversity per nucleotide position based on Shannon entropy (H). (A) Envelope gene sequences. (B) Genomes sequences. Bombers on top of each plot represent the sample code.

A



B

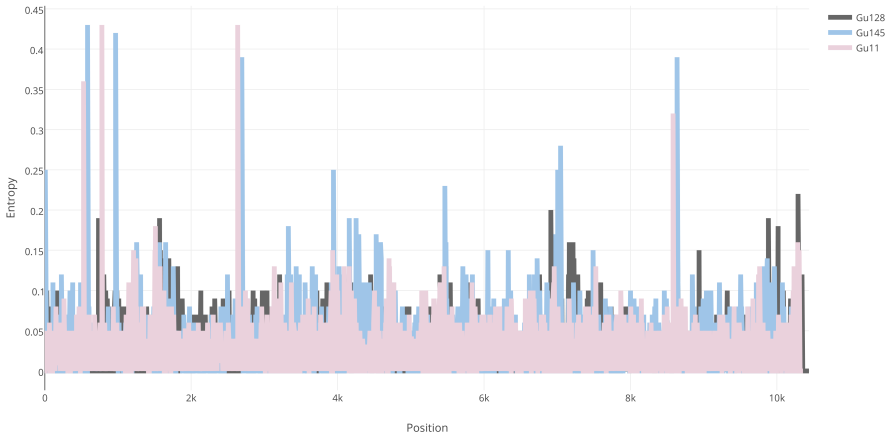


Figure S7 - Hexagonal binning plot representing the relationship between the GC content of first-second and third codon position.

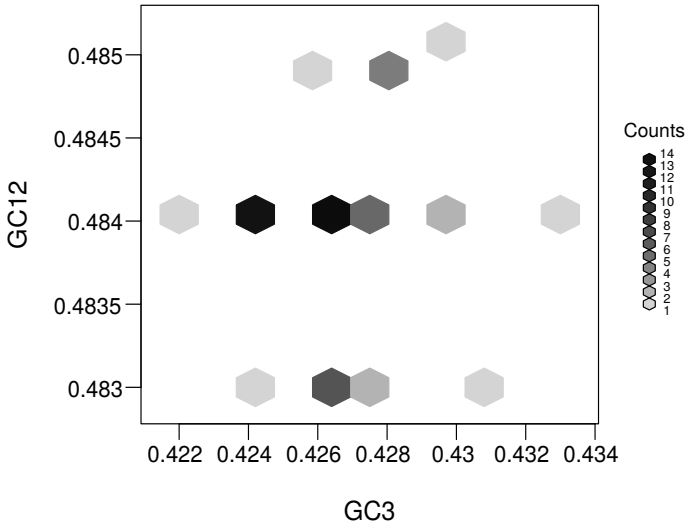


Figure S8 - Proportion of monomorphic and polymorphic sites present in DENV sequences included in this study.

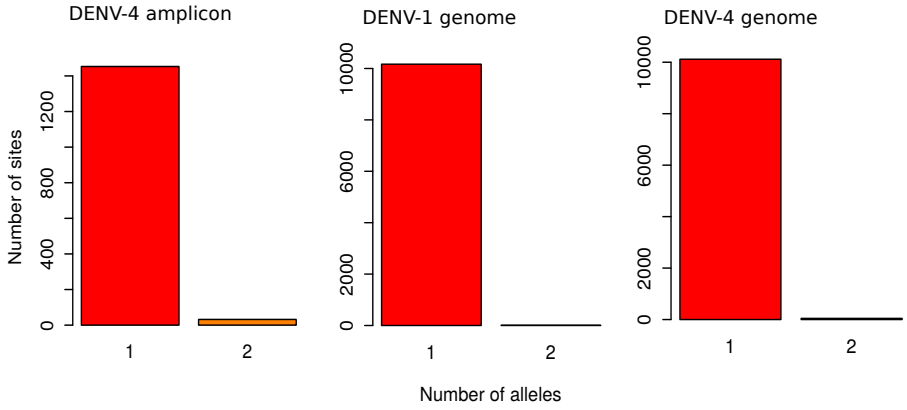


Figure S9 - ML DENV-1 phylogenetic tree showing branches and sites under positive selection pressure**.** The tree was constructed using genomic sequences of DENV-1.

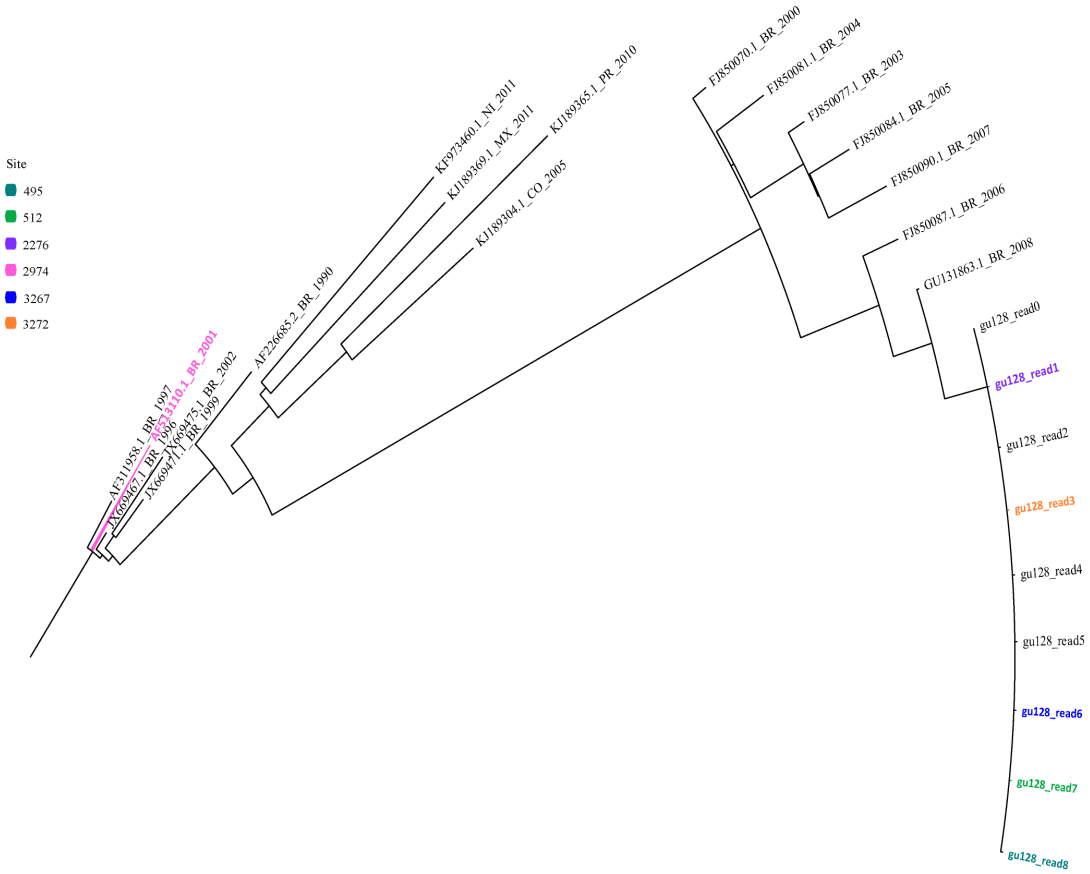
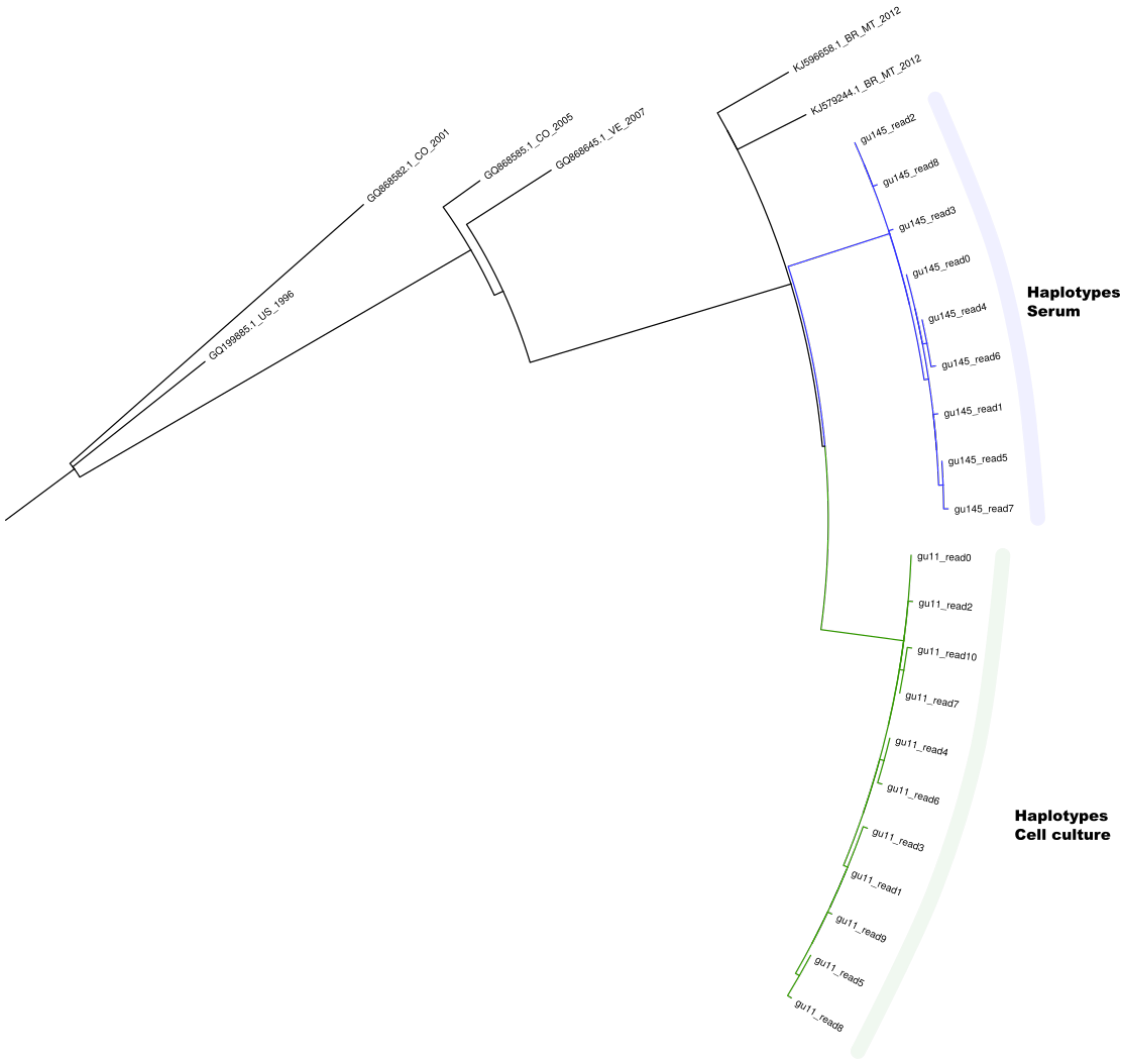


Figure S10 - ML phylogenetic tree depicting lineages isolated from serum and cell culture. The tree was constructed using genomic sequences of DENV-4.



TABLES

Table S1. List of samples included for ds-cDNA library preparation.

Sample	**Number of extractions**	**Qubit (ng/µL)**	**Nanodrop (ng/µL)**	**Copies/ µL (dengue)**	**Source**	**One-Step**
GU006	1	337	32	NA	Serum	Positive
GU011	3	753,5	49,2	NA	Cultivated	Positive
GU128	12	1079,1	70,4	2.91 E+07	Serum	Positive
GU142	17	1550,1	102	5.87 E+05	Serum	Positive
GU144	8	1719,4	105	1.24 E+06	Serum	Positive
GU145	10	254,1	27,4	3.41 E+06	Serum	Positive
GU795	14	690	43,6	1.33 E+05	Serum	Positive
GU797	8	1497,9	95,2	3.79 E+06	Serum	Positive
GU882	7	985,7	53,4	2.32 E+05	Serum	Positive
GU1153	11	28,3	17	1.24 E+05	Serum	Positive

Table S2. List of samples with reads mapping DENV-1 and DENV-4 reference sequences.

Sample	DENV-1	DENV-4	Mapped reads	Ref-Seq DENV-1	Ref-Seq DENV-4
GU128	17.347	9	17.354	97.8%	8.3%
GU142	2	61	63	2.8%	38%
GU145	9	12.255	12.662	7.1%	99.1%
GU882	4	7	11	3.2%	5%
GU1153	4	2	6	1.2%	0.5%

Table S3. Relative synonymous codon usage in envelope E gene of DENV-4 populations. Values in parenthesis correspond to codon frequencies.

AA	Codon	RSCU (N)	AA	Codon	RSCU (N)
Phe	**UUU**	**1.37** (39)	Ser	UCU	0.64 (9)
	UUC	0.63 (18)		**UCC**	**1.07** (15)
Leu	UUA	0.69 (12)		**UCA**	**2.79** (39)
	UUG	**1.49** (26)		UCG	0.43 (6)
	CUU	0.34 (6)	Pro	CCU	0.80 (9)
	CUC	0.86 (15)		**CCC**	**1.60** (18)
	CUA	**1.03** (18)		**CCA**	**1.07** (12)
	CUG	**1.60** (28)		CCG	0.53 (6)
Ile	**AUU**	**1.08** (27)	Thr	ACU	0.35 (12)
	AUC	0.72 (18)		ACC	0.78 (27)
	AUA	**1.20** (30)		**ACA**	**2.17** (75)
Met	AUG	1.00 (54)		ACG	0.70 (24)
Val	**GUU**	**1.08** (39)	Ala	**GCU**	**1.38** (30)
	GUC	0.83 (30)		**GCC**	**1.24** (27)
	GUA	0.53 (19)		**GCA**	**1.10** (24)
	GUG	**1.56** (56)		GCG	0.28 (6)
Tyr	UAU	0.91 (15)	Cys	**UGU**	**1.08** (21)
	UAC	**1.09** (18)		UGC	0.92 (18)
TER	UAA	0.00 (0)	TER	**UGA**	0.00 (0)
	UAG	0.00 (0)	Trp	UGG	1.00 (30)
His	**CAU**	**1.38** (27)	Arg	CGU	0.30 (3)
	CAC	0.62 (12)		CGC	0.90 (9)
Gln	CAA	0.83 (15)		CGA	0.30 (3)
	CAG	**1.17** (21)		CGG	0.00 (0)
Asn	AAU	1.00 (27)	Ser	AGU	0.64 (9)
	AAC	1.00 (27)		AGC	0.43 (6)
Lys	**AAA**	**1.10** (50)	Arg	**AGA**	**3.30** (33)
	AAG	0.90 (41)		**AGG**	**1.20** (12)
Asp	GAU	1.00 (30)	Gly	GGU	0.58 (24)
	GAC	1.00 (30)		GGC	0.51 (21)
Glu	**GAA**	**1.21** (54)		**GGA**	**2.33** (96)
	GAG	0.79 (35)		GGG	0.58 (24)