Universidade do Minho

School of Medicine

Francisco Javier Cuevas Araujo

Genomic and immunoinformatic analysis of *Dengue virus*

Master dissertation

Master in Health Sciences

Advisors

Nuno Miguel Sampaio Osório, PhD

Maria Isabel Mendes Veiga, PhD

June 2017

The work presented in this thesis was done in the Microbiology and Infection Research Domain of the Life and Health Science Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal. ICVS is a member of the ICVS/3B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal.

Table of Contents

Illustration Index

Index of Tables

Acknowledgments

I want to thank all the people who during these academic years here in Portugal gave me their support and guided me throughout the development of this work:

Nuno Osorio and Isabel Veiga, for your guidance, and for share yours experience in the field of science. I appreciate your hospitality and friendship.

Pedro, Carlos, and Vera, thank you for your support and friendship, also for letting me be part of the dry-lab team.

To all the people of the malaria team, for your support and your teachings in the laboratory part, thank you for being patient with me.

To thank in general all the people that work in the ICVS and the School of Medicine of the University of Minho, especially to the students and researchers in the Microbiology and infection research domain (MIRD).

In the Dominican Republic, I would like to thank the team of the Institute of Microbiology and Parasitology of the Autonomous University of Santo Domingo, especially to Modesto Cruz for giving me the opportunity to be part of the IMPA family.

Last but not least to thank my family and friends in the Dominican Republic especially my mother Francisca Araujo and my daughter Shannon for being so patient and give me all your support from so far.

Abstract

Dengue is a vector-borne disease caused by the *Dengue virus* (DENV) and is currently considered as a major public health problem. It is estimated that approximately four billion people are at risk of DENV infection and 400 million are infected every year. DENV can be phylogenetically divided into four serotypes and several genotypes. The pathology caused by DENV ranges from mild flu-like symptoms to potentially lethal complications that are often associated with infection with different serotypes. Taking advantage of 2177 publically available whole-genome sequences we explored, by phylogenetic analysis, the global genetic population structure of DENV and identified serotype/genotype-associated genetic diversity. The results support that serotype 4 is the most genetically distinct, while serotype 2 and 3 are the closest. The analysis of the phylogenetic patterns of DENV also highlighted 7 sequences that are likely to belong to previously unidentified genotypes. In addition, we mapped 1613 experimentally validated T cell epitopes to the studied sequences and identified clusters of serotype and genotype specific epitopes. Most interestingly, we found that 17 of the 1613 epitopes studied were hyperconserved and could be combined to achieve above 90% predicted coverage in the human population. Overall, this work supports that the diversity of DENV could be even larger than what was considered as new genotypes are likely to exist. This raises the importance of detailed molecular surveillance of DENV. Despite the high diversity, it was possible to find T cell epitopes with high levels of conservation. These epitopes could be useful for the development of effective immunotherapy strategies to prevent severe forms of dengue.

Keywords: *Dengue virus*, genomics, immunoinformatics.

Resumo

A dengue é uma doença transmitida por vetores causada pelo vírus da Dengue (DENV) e atualmente é considerada um importante problema de saúde pública. Estima-se que aproximadamente quatro bilhões de pessoas estejam em risco de infeção pelo DENV e 400 milhões sejam infetadas *de novo* a cada ano. O DENV pode ser filogeneticamente dividido em quatro serotipos e vários genótipos. A patologia causada por DENV varia entre sintomas moderados (semelhantes aos da gripe) até complicações potencialmente letais que são frequentemente associadas a infeções com diferentes serotipos. Partindo de 2177 sequências de genoma completo publicamente disponíveis, exploramos, por análise filogenética, a estrutura genética global do DENV e identificamos a diversidade genética associada a serotipos e genótipos. Os resultados suportam que o serotipo 4 é o mais geneticamente distinto, enquanto os serotipos 2 e 3 são os mais próximos. A análise dos padrões filogenéticos destas sequências de DENV também destacou 7 sequências que provavelmente pertencerão a genótipos previamente não identificados. Mapeamos ao total das sequências analisadas 1613 epítopos de células T validados experimentalmente e, assim, identificamos clusters de epítopos específicos de serotipo e genótipo. De forma muito relevante, descobrimos que 17 dos 1613 epítopos estudados são hiperconservados e quando combinados apresentam um resposta estimada em mais de 90% na população humana. Em síntese, com a potencial descoberta de sequencias de novos genótipos, este trabalho evidenciou uma maior diversidade de DENV do que a que era conhecida. Estes dados reforçam a importância da vigilância molecular detalhada do DENV. Apesar da elevada diversidade foi possível encontrar epítopos de células T com altos níveis de conservação. Estes epítopos podem ser úteis para o desenvolvimento de estratégias de imunoterapia eficazes para prevenir formas graves de dengue.

Palavras-chave: Vírus da Dengue; genómica; imunoinformática.

Abbreviations

DENV *Dengue virus*

RNA Ribonucleic acid

WHO World Health Organization

DSS Dengue Shock Syndrome

ADE Antibody-Dependent Enhancement

WGS Whole genome sequence

NGS Next generation sequence

HLA Human leucocyte antigen

CYD-TDV - Chimeric Yellow fever and Dengue tetravalent dengue vaccine

ER Endoplasmic reticulum

SNP Single Nucleotide Polymorphism

PAML Phylogenetic analysis by maximum likelihood

NCBI National Center for Biotechnology Information

VIPR Virus Pathogen Resource

NNI Nearest Neighbor Interchanges

GTR Generalized time reversible

MAFFT Multiple Alignment using Fast Fourier Transform

IEDB Immune Epitope Database and Analysis Resource

MHC Major Histocompatibility Complex

Introduction

Dengue: an emerging disease and public health problem in XXI century

Dengue is the most important disease transmitted by arthropods. Approximately four billion people in more than one hundred tropical and subtropical countries predominantly in the regions of Asia, Latin America, and Africa are at risk of being affected [1]. Only in 2013, there 3.9 million estimated cases of dengue (Figure 1) resulting in nine thousand reported deaths in low- middle-income areas [2]. Furthermore, it is estimated that in the past five decades the incidence of dengue increases 30 fold becoming one of the major public health problems worldwide.

The origin of dengue is elusive, but records from ancient China (992 A.D) already present clinical descriptions with symptoms very similar to dengue. It is believed that the epidemic of Dengue was developed from ancestral lineages of the virus more than two thousand years ago in Asia and/or Africa and then spread in others regions [3], [4]. Some studies pinpoint an event of enzootic transmission between non-human primates to humans to around 300 years ago [5]. However, it was not until the 19th century that Dengue reached strong significance as a cause of disease outbreaks in South-East Asia, Africa, Western Pacific, America and the Caribbean [6].

Dengue virus (DENV) is a lipid-enveloped virion belonging to the *Flavivirus* genus of the *Flaviviridae* family and transmitted to humans by *Aedes* mosquitoes (primarily *Aedes aegipty and Aedes albopictus*) [7]. So far four distinct but antigenically related serotypes (DENV1, DENV2, DENV3 and DENV4) have been described with a phylogenetic variation within them defined as genotypes [8]. Moreover, the distribution of the 4 serotypes and their different genetic variants have dramatically changed since the 1970s, and nowadays the spread of the 4 lineages in all world regions is a reality [9], [10] (Figure 2).

The incidence of different serotypes circulating in the same geographic area may increase the risk of an outbreak, since the interaction of several serotypes in the same host may result in a complex immune response known as ADE (Antibody-Dependent Enhancement) [11], [12]. This theory explains why severe dengue is often associated with heterotopic or secondary infection [13]–[15]. Even so host immune responses can result in a broad range of clinical manifestations that range from asymptomatic or mild-flu febrile illness, headache, and retro-orbital pain to a severe haemodynamic compromise and Dengue Shock Syndrome (DSS). This last phase known as severe Dengue has a case fatality rate of 0.2 - 5%, increasing considerably if the patient is not treated with extreme care in an intensive care unit [16], [17].

The diagnosis of dengue is based mainly on the medical history and in some clinical settings, laboratory tests to confirm the infection are also used. The most common tests are based on serology, IgM antibody-capture enzyme-linked immunosorbent assay (MAC-ELISA) and IgG ELISA or plaque reduction and neutralization test (PRNT) [18]. Also, molecular methods such as Reverse transcriptase polymerase chain reaction (RT-PCR) and nonstructural protein 1 detection (NS1) are also available [19], [20]. Despite the availability of diagnostic methods, a large percentage of cases are under-reported due to the lack of resources in the affected areas. In low income endemic countries epidemiology reports of dengue are solely based on clinic history of the patient

in addition to the WHO guidelines classifications. These strategies are useful in guiding the primary clinical response but have no effect to monitor and prevent, at a molecular epidemiology level, the spread of more transmissible DENV genotypes [19].

There is no antiretroviral treatment for DENV and dengue therapy is based on fluid maintenance and the use drugs to control the major symptoms. Most of non-steroidal anti-inflammatory drugs (NSAIDs) are banned because of increased hemorrhagic risk [16]. The major efforts for dengue prevention are still focused on the control of vectors by the elimination of breeding sites, by the use of insecticides and mosquito nets. These are the most significant public health strategies in endemic regions to avoid vector growth and transmission. For almost 50 years the challenge for developing an effective vaccine has been unmet. Nevertheless, substantial advances have been made in recent years. Live attenuated vaccines are currently in clinical trials and more recently the Chimeric Yellow Fever-Dengue - Tetravalent Dengue Vaccine (CYD-TDV) from Sanofi Pasteur company was recently licensed and approved for use in endemic countries since 2016 [21]–[23]. Other types of vaccines, not based on live attenuated virus, such as DENVax or TDV (Takeda Dengue Vaccine) from Takeda Inc. are also progressing through final phases [24]. These are considered future alternatives that will help control this disease. However, multiple factors such as world population growth, uncontrolled urbanization in developing countries and increased air transport remain crucial in promoting the spread of the virus in the coming years. Dengue is likely to continue to have a very strong socio-economical impact in endemic countries as well as other diseases such as malaria and tuberculosis [16]. For instance in Puerto Rico, a tropical country where dengue is endemic, an economical study estimated that during the period of 2002-2010 the costs for prevention and control the disease summed more than $400 million USD [25]. This example reinforces the necessity to create new health policies to foster continuous research and innovation programs to address this problem [6], [10]. From another point of view, the exponential increase of DENV infections in the past decades make the development of new strategies to fight dengue ever more urgent to. For instance, innovation based on the use of molecular biology, next generation sequence (NGS) and bioinformatics analysis are likely to become an essential part in understanding the pathophysiology and following the evolutionary path of the virus to prevent new outbreaks [26], [27].

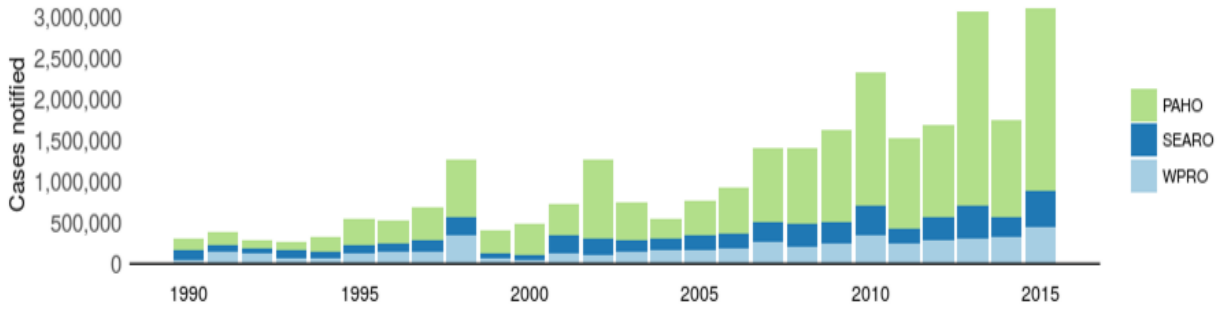


Figure .1 Dengue cases reported by the regional offices of WHO from 1990 to 2015. Bar chart represents the numbers of cases suspected or diagnosed by dengue laboratory in the different regions. Pan American Health Organization (PAHO), South-East Asia Regional Office (SEARO) and Western Pacific Regional Office (WPRO). Adapted from: [28].

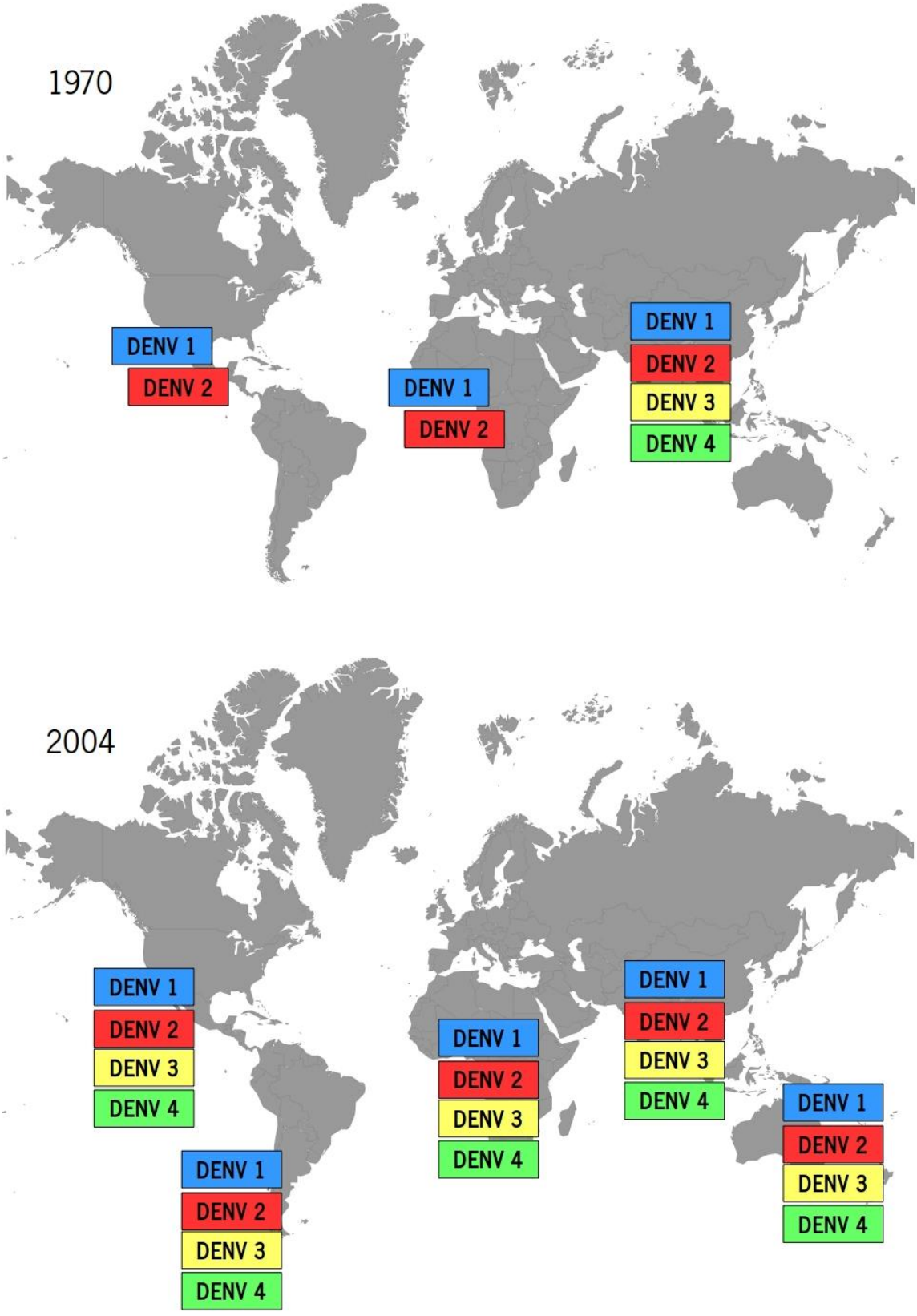


Figure 2. Dynamics of the *Dengue virus* serotypes. The upper figure shows the distribution of *Dengue virus* serotypes in 1970 and the bottom figure show the distribution in 2004. Adapted from [6], [10]. Reproduced by permission of Nature Publishing Group, license number 4096581281699.

Genetics and diversity of *Dengue virus*

As mentioned earlier DENV is a member of *Flaviviruses* family, which is related to many important pathogens like, Yellow fever virus, West Nile virus and Zika virus [7], [29]. The causal agent of Dengue is a single strand RNA virus of ~11kb in length enclosed in a spherical virion particle of 40-50nm of diameter [10]. The genome organization comprises three structural proteins (C, prM, E) and seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5) which are flanked by 5' and 3' untranslated regions [7], [30](Figure 3). These regions contain more than 60 percent of similarity within serotypes and 6-8 percent of dissimilarity between genotypes [31].

Genotypes of DENV, Zika, Chikungunya and West Nile Virus can be classified from genome sequences using a computational algorithm that has been shown to be an accurate method for genetic classification [18], [32]. In case of DENV the envelope (E) with 1485bp is the gene used by this tool for genotype classification [18]. So far there is no consensus on the nomenclature used to denote genotypes a numeric short nomenclature will be used in this thesis (Table.1)

As expected for an RNA viruses, DENV demonstrates significant genetic variability thanks to the fast replication and high mutation and recombination rates but also to others factors like the natural selection imposed by the large host/vector range during transmission cycles and [3], [8]. This genetic variability of the virus is likely to be related with the spread of the pandemics in the last decades [3] due to the introduction of new genotypes or the interchange of genotypes between geographic regions [33]–[35].

The substitution rate for DENV has already been reported, and its estimated to be in the range of $4.55x10^{-4}$ to $9.01x10^{-4}$ nucleotide substitutions per site/year, that is in line with the substitution rates for other RNA viruses with similar transmission cycles [34], [36]. Several polymorphisms described in DENV genome seem to be associated with either vulnerability or protective effects to infection [37]. For instance, mutations of the envelope (E) protein in the position W101 decreased the reactivity of polyclonal human serum by afected the binging of antibodies in the fusion loop epitope of the protein. Moreover, another E protein polymorphism characteristic of DENV2 genotypes (E390) has been associated with virulence in experimental studies [15], [38]. Other studies suggest that regions conserved in the envelope protein (250-270) of DENV influence viral replication in human cells, [39].

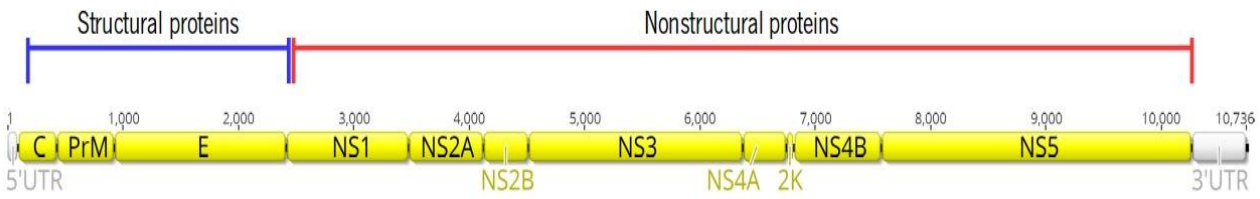


Figure 3. *Dengue virus* genome. RNA protein-coding regions and genome organization. Viral RNA of 10,756 kb length that translated three structural proteins Capsid (C), Pre membrane (prM), and Envelope (E) and seven non-structural proteins (NS). Flanked by 5' and 3' untranslated region. Adapted with permission of: [7].

Table 1. Genetic diversity of *Dengue virus* (DENV). Serotypes and genotypes according to different nomenclatures. Adapted from: [32], [40].

Serotype	Genotype	*Short Nomenclature
DENV1	Genotype I	1I
	Genotype II	1II
	Genotype III	1III
	Genotype IV	1IV
	Genotype V	1V
DENV2	Genotype I - American	2I
	Genotype II - Cosmopolitan	2II
	Genotype III – Southern Asian American	2III
	Genotype IV – Asian II	2IV
	Genotype V – Asian I	2V
	Genotype VI – Sylvatic	2VI
DENV3	Genotype I	3I
	Genotype II	3II
	Genotype III	3III
	Genotype IV	3IV
	Genotype V	3V
	Genotype VI	3VI
DENV4	Genotype I	4I
	Genotype II	4II
	Genotype III	4III
	Genotype IV - Sylvatic	4IV

*Nomenclature used for the development of this thesis

Immunopathology and T cell recognition of *Dengue virus*

A mosquito infected by DENV can spread the virus in the host when feeding, being possible to infect more than one host when trying to complete feeding [41], [42]. Many glycoproteins are proposed to mediated the viral entry into the host cell. These include the manose receptor, heparin sulfate receptor and others. It is also well accepted that the E protein also affects the entry of the virion and the host receptor binding [41]. After penetrating the cell the virus begins replicating its genome and the translation of viral proteins is also initiated (Figure 4a) [43]. The immune response to the virus is primarily mediated by T and B lymphocytes. Neutralizing antibodies are generated after the presentation of viral epitopes by dendritic cells [44] on Major Histocompatilblity Complex (MHC) I and MHC class II receptors. More specifically, B and $CD4^+$ T cells were

described to induce a direct response predominantly to the E protein of the virus, while the CD8+ T cells are involved in responses to nonstructural proteins such as NS3 and NS5 [44], [45].

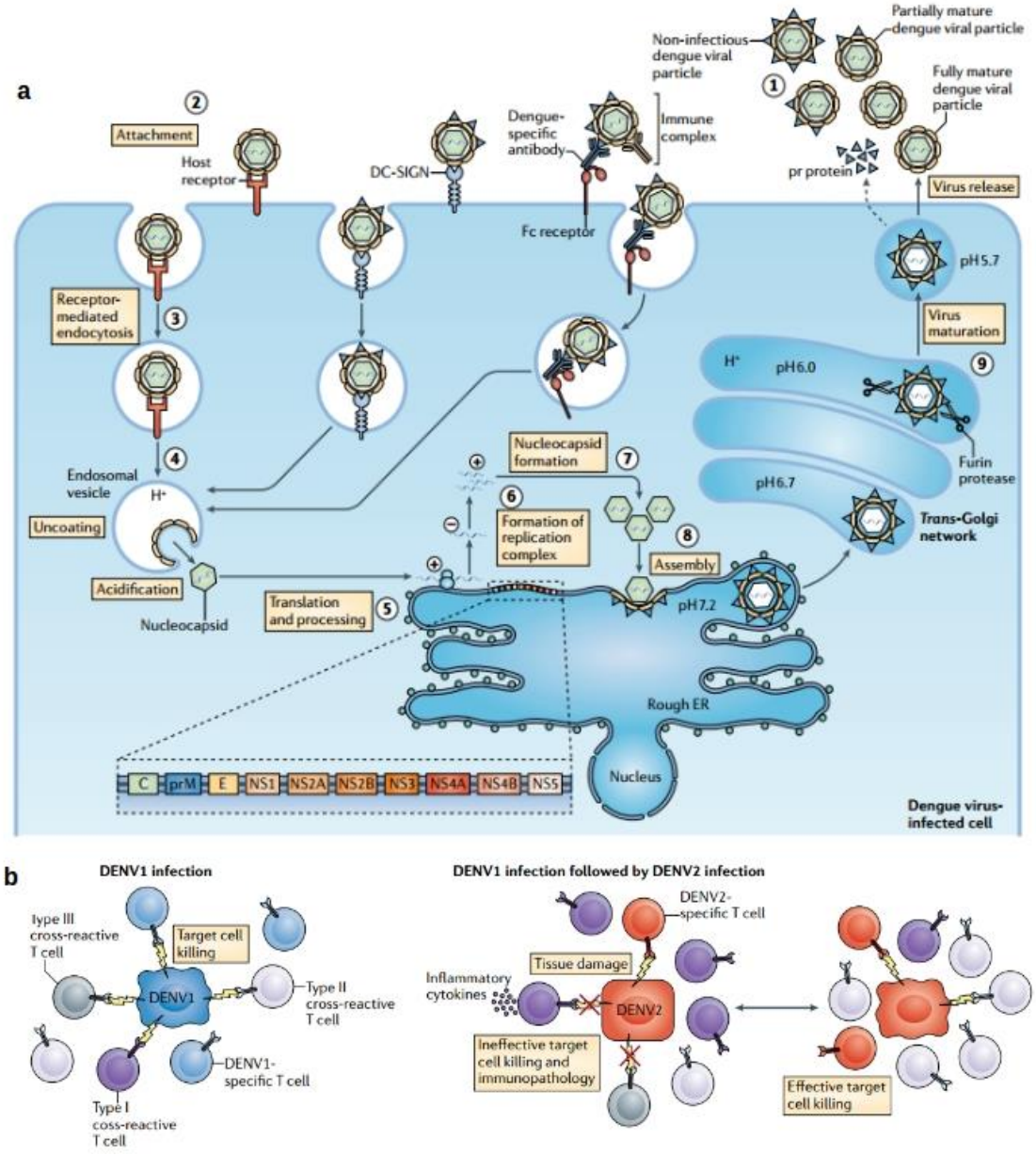


Figure 4. Life cycle and adaptive immune response after infection with *Dengue virus*. a. 1) Virus release; 2) attachment to cell receptors; 3) receptor mediated endocytosis and formation of endosomal vesicle; 4)acidification of endosomal vesicle and release of nucleocapsid; 5) translation and processing in the rough endoplasmic reticulum; 6) formation of replication complex; 7) nucleocapsid formation; 8) assembly and 9) virus maturation. **b.** T cell is activated in a secondary infection with a different serotype of the previous infection and lead to the dominance of that response by cross-reactive T cells generated in the previous infection. Adapted from G. Screaton *et al.* [43]. Reproduced with permission from Nature Publishing Group, license number 4096581281699.

For individuals living in endemic regions after a first infection a secondary infection by a different DENV virus is likely to occur and has been described to increase the risk of a severe condition. This is attributed to the condition known named Antibody-Dependent Enhancement (ADE) in which the host antibodies generated to a previous infection seem to enhance the viral load and disease severity after infection by genetically different

DENV [11], [12], [15], [36], [41], [46]. It has been suggested that a variation greater than 30-35% in the E protein of the DENV from a secondary infection might be sufficient to induce ADE [43]

Another relevant trigger of ADE seems to be the activation of a memory T cell after a secondary infection with a genetically different DENV leading to the dominance of the response by cross-reactive T cells generated in previous infections (Figure.4b). ADE has been demonstrated in several animal and in vitro cell models expressing Fc receptors and in which the viral loads have been enhanced [49], [50]. This is possible related with the increase in the severity of the disease in breast feeding children within the first year during exposure to antibodies transmitted from the previously infected mother .and where the levels of these antibodies are inadequate for effective neutralization of the virus [51]–[53].

Genomics and Immunoinformatics applied for the study of Dengue

Many factors like climate change, global travel, and rapid increase population and urbanization have facilitated the increment of the incidence and geographic range of DENV in recent decades [31], [54]. Although, significant advances have been made in the knowledge of vector interaction, pathogenicity, immunogenetics and virus evolution. These approaches have been achieved thanks to the development of new technologies that are increasingly available to study in depth the complete genome of various pathogens, to understand their transmission mechanisms and in this way to create new strategies for disease control. Genomics and immunoinformatics are two branches of computational biology that contribute with methods and tools and widely used for the study of infectious diseases, cancer genetics and the design of new therapies and vaccines [55]–[59].

The study of the evolution, genetic diversity and other aspects related to the biology of the host/pathogen interaction through genomic analyses is essential to understand the epidemiology and the emergence of DENV [54], [60]. The phylogenetic analysis of the gene encoding the E protein of the DENV genome is currently used to determine the viral genotype. This consensus has been reached since this gene has enough length and diversity (1485bp) to allow differentiating DENV from other viruses and between genotypes [32]. Although phylogeny based on the sequence encoding E protein has been widely used in the last decade allowing important advances in the identification and emergence of novel outbreak-related genotypes [34], [61], [62] phylogenetic analysis using complete DENV genome sequences has an even greater potential. A good example of the use of these approaches for a better understanding of DENV genetic diversity is the recent study by *Pyke et* showing an unique DENV strain likely to belong to a "new genotype" [34]. Moreover, in Japan a platform to visualize all the available genome sequences of DENV allowing the follow up of new DENV outbreaks has been created with great potential to reinforce the dengue molecular epidemiology surveillance [63].

Presently, there are several public databases from enabling meta-analysis of the exponentially increasing amount of next-generation sequencing data being generated [31]. These include Virus Variation Database (NCBI) and Virus Pathogen Database Resource (ViPR) [64]. Another area in clear development is the application of bioinformatics to the study of immunology, commonly referred as Immunoinformatics. This area of research includes the study of T- and B-cells epitopes using various computational algorithms with the objective of

decreasing the time and cost of laboratory testing [65]. It is increasingly used in the design of vaccines, diagnostics and host directed immunotherapy strategies [57] (Figure 5). In terms of public databases, the Immune Epitope Database (IEDB) stands out as the biggest database collecting information on epitopes incorporating more than 280,000 experimentally validated epitopes [66].

The study of conserved genomic regions between diverse RNA viruses has been proposed by many researchers for the development of new immunotherapies and vaccines [39]. Conserved regions often play a significant role in the structure of the virus and could be good targets for the neutralization of antigens such as the ones from HIV and Influenza [39]. In the case of DENV it is well accepted that $CD4^+$ and $CD8^+$ memory T cells are generated by a primary infection might be cross-reactive to other DENV serotype or another *Flavivirus* induced protective or adverse responses. The concentration and degree of epitope conservation and the specificity of these lymphocytes present might be to be in balance for a protective response to be formed [67]. A relevant step in the level and type of response induced by an epitope is its binding kinetics to MHC molecules (also known as human leukocyte antigen (HLA) in humans) since the presentation of epitopes to $CD8^+$ or $CD4^+$ T lymphocytes is always performed in the context of MHC class I or class II molecules[68], [69].

The identification of epitopes that might induce protective responses for the four serotypes of the DENV could be of great relevance for immunotherapy [70]. The most conserved region described so far among *Flaviviruses* (including the 4 serotypes of DENV) is the NS5 region, which has a significant impact on $CD8^+$ T-cells recognition making it a good potential candidate for vaccine development [22], [71],[72].

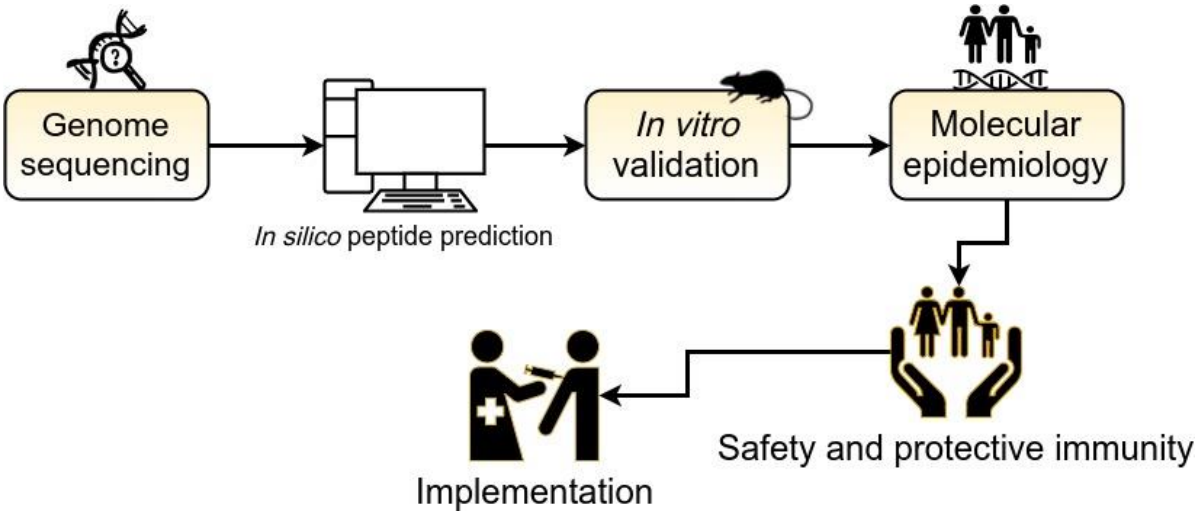


Figure 5. Schematic pipeline summarizing the pathways of vaccine development starting from reverse vaccinology. Genomes are screened for antigenic regions (peptide predictions) using bioinformatic tools. Epitope candidates are tested *in vitro* using animal models and strains of different species. Candidate vaccines are tested for safety and protective immunity and after clinical trials and efficacy studies approved for implementation. Adapted from A. Sette *et al.* [73]. Reproduced by permission of Elsevier publisher, license 4096600828981.

Aims

On the last decades, there has been a considerable expansion of DENV, there are several factors that drive this pandemic like globalization, the spread of the *Aedes* mosquito vector, inadequately planned urbanization and the absence of effective vaccines or specific anti-dengue therapies [9], [15].

Some reports indicate that the circulation pattern of DENV serotypes may have changed lately in different regions raising the importance in molecular surveillance of this pathogen [74]. Moreover, the complex interaction of different DENV serotypes with the human host often inducing immunopathology in secondary infections is the main challenge for dengue therapy development [75].

Projecting that dengue is a multifactorial problem involving the interaction of a genetically diverse population of the virus in different human populations with varying human leukocyte antigen (HLA) molecules this study proposes to address the following objectives:

1. Analyze the genetic diversity of DENV using all available whole genomes sequences (WGS).

2. Identify the conservancy level of experimentally validated T-cells epitopes in different DENV serotypes and genotypes.

3. Predict the human population coverage of the experimentally validated T-cell epitopes.

Accomplishing these objectives will allow a better understanding the genetic diversity of DENV and how its impacts the conservancy levels of validated T cell epitopes. This will be of relevance for the development of innovative and potentially more effective immunotherapies for dengue.

Results

The analysis of available DENV genomes by geographic site of sampling suggest widespread distribution of 4 serotypes

Until December 2016, a total of 2,179 complete genomes were sequenced from clinical isolates in Africa, Europe, Oceania, Asia, North, Central and South America, and made available on public databases. The analysis of these sequences by serotype and geographic site of sampling, showed all 4 DENV serotypes were sequenced in all major world region affected by dengue. The only region where genomes of the four serotypes have not been sequenced was Western Europe possibly due to the low incidence and very low number of sequences available (Figure 6).

Seventy four percent (1664/2179) of the available genome sequences came from Asian. From the 58 countries represented, Vietnam was the country will a larger number of sequenced isolates representing 42% (920/2179) of the sequences (Figure 6 and Supplementary Data Figure 12). Importantly, the most sequenced serotype in the regions of Africa, Oceania, South, North and Central America was DENV2. On the contrary, in Asia, DENV1 was found to be the most frequent serotype (Figure 6). Overall, the serotype distribution showed that DENV1 with 51.9% (1131/2179) and DENV2 with 28.41% (619/2179) were the serotypes with more available genome sequences. On the other hand, DENV3 with 14.09% (307/2179) and DENV4 with 5.60% (122/2179) were the less sequenced (Supplementary Data Figure 13).

The publicly available DENV whole-genomes did not result from an organized sampling strategy allowing it to be representative of the population of virus in the different regions. Nonetheless, it can be used to support the conclusion that all four DENV serotype are worldwide spread.

At the genotype level, so far least 20 genotypes have been sequenced around the world. In this work we were genotyped 2,179 DENV genomes according to have an estimate of their presence in the endemic regions. The results show that the genotypes DENV2III have been the most frequent in South, North and Central America. In Africa the DENV2VI and Oceania DENV2I were the most sequenced genotypes respectively. However, in Asia the most frequent genotype was DENV1I. The reduced number of cases isolated in the European Union were n DENV1I or DENV2VI genotypes (Figure 7). However, several other DENV genotypes have been sequenced in other regions, indicating a great diversity of DENV genotypes spread in different regions. (Figure 7 and Supplementary Data Figure 14)

Most interestingly, a known genotype could not be attributed to five out of the 2,179 sequences genotyped using the *in silico* tool "Dengue, Zika, Chikungunya and Yellow Fever Virus" [32]. One of these sequences belonged to serotype DENV1, two sequences belonged to DENV3 and the two sequences belonging to DENV4. The sequences of unknown genotype were isolated in Asia (6/7) or Africa (1/7). This was then further analyzed by phylogenetic analysis.

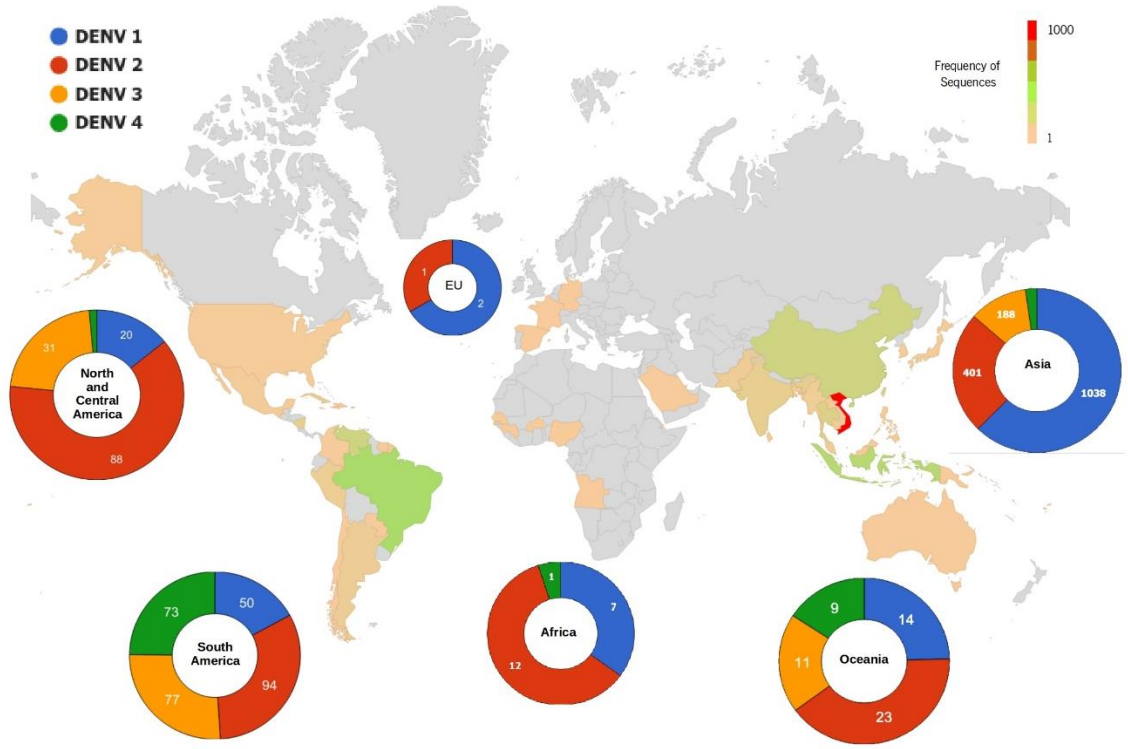


Figure 6. Distribution of *Dengue virus* (DENV) genomes according to serotypes. DENV1 (blue), DENV2 (red), DENV3 (orange) and DENV4 (green) WGS in different regions. Background map shows the countries where the sequences came from and gradient colored by their frequency.

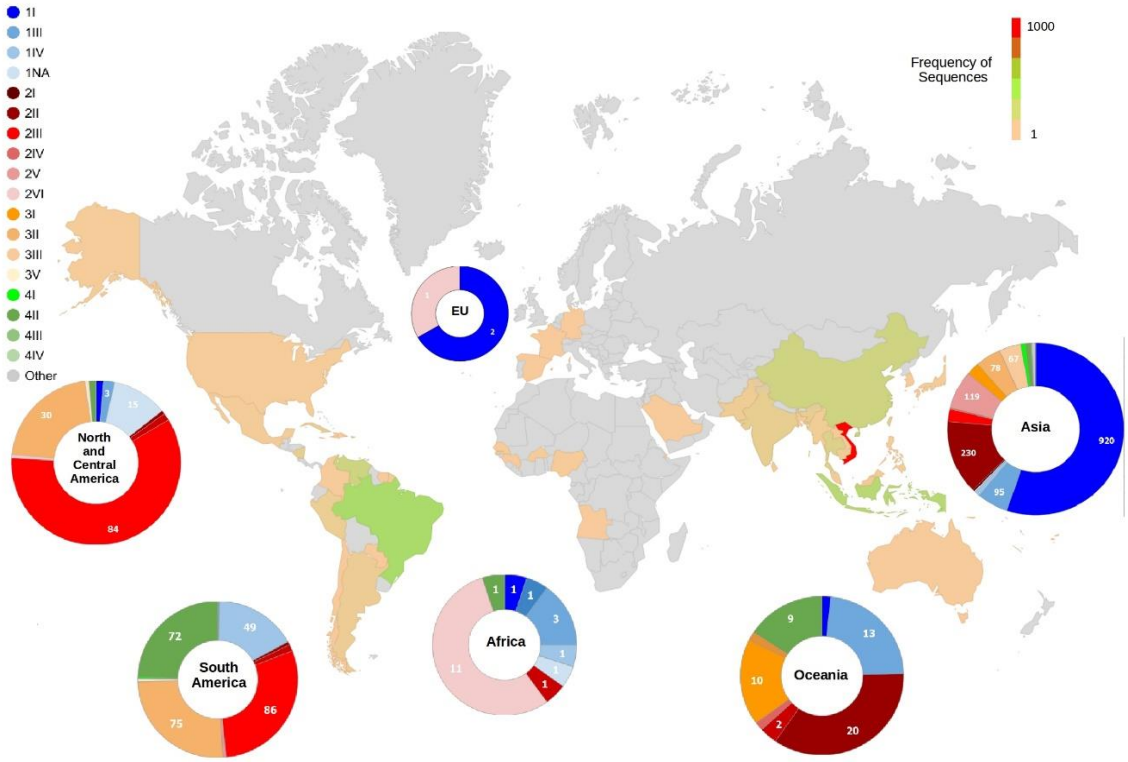


Figure 7. Distribution of *Dengue virus* genomes according to genotypes. DENV1 genotypes (gradient blue), DENV2 genotype (gradient red), DENV3 genotype (gradient orange) and DENV4 genotype (gradient green) in different regions. Background map shows the countries where the sequences come from and their frequency.

Phylogenetic analysis of *Dengue virus*

A phylogenetic unrooted tree of 2179 complete DENV genomes was constructed using a Maximum Likelihood approach as implemented in PHYML v.3.0 [76].Two sequences had to be excluded from the phylogenetic analysis due to the large amount of gaps present. The tree using 2175 complete DENV genome sequences can be observed in Figure 8.

In the rectangular cladogram four clusters are clearly identified and correspond to viral sequences from different serotypes. According to these phylogenetic analysis serotype 4 was the first to diverge from the other subtypes, while serotype 1, 2 and 3 share a more recent common ancestor. The most phylogenetic related serotypes were 2 and 3 (Figure 8).

As for the genotypes it was observed that the genotypes III and V share a recent common ancestor as well as genotypes I and II. (Figure 8)

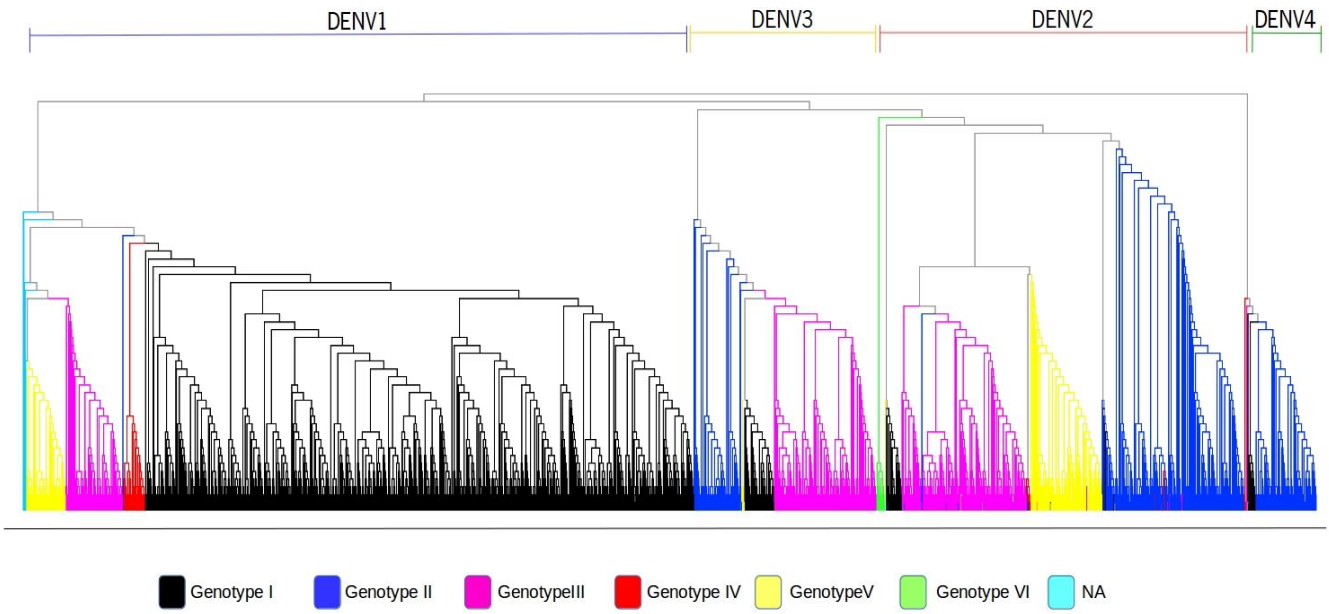


Figure 8. Phylogenetic tree constructed with 2177 complete genomes of *Dengue virus*. A cluster of serotypes framed by the top lines and the genotypes of each lineage defined by colors: Genotype I (Black), Genotype II (blue), Genotype III (purple) Genotype IV (red), Genotype V (yellow), Genotype VI (green), Genotype NA (light blue).

More interestingly, the phylogenetic analysis reinforces that the 5 sequences that could not be genotyped by the SNP-based tool [32] are part of novel genotypes. These sequences appear separately from the other sequences in the phylogenetic analysis (Figure 9.). These 5 unknown sequences were isolated from Brunei (KR919820) [34], Angola (KF184975) [77], Malaysia (EF457905) and India (KF289073, JQ922546) did not fit into any group. One of these sequences was described in a recent publication suggesting the emergence of a new genotype [34].

Level of sequence conservation of the T cells epitopes across the diversity of DENV

Identifying highly conserved epitopes across genetic variants of DENV can be of great interest for designing vaccines with broader efficacy or diagnostic tools. To address this question we extracted from IEDB [66] 1613 experimentally validated T cells epitopes from DENV. We then mapped all of these epitopes on all unique proteomes (n=1739) encoded by the 2175 DENV genomes under study.

The frequency of each epitope in the proteomes of the different serotypes and genotypes was represented in a heatmap (Figure 9). The figure clearly shows clusters of epitopes that are serotype-specific and epitopes shared by 2 or 3 serotypes. Interestingly, there was a small cluster of epitopes (17/1613 T cell epitopes) that were highly conserved in all or most of the DENV serotypes and genotypes. These hyperconserved epitopes were located in nonstructural proteins (NS1, NS3, NS4B, and NS5).

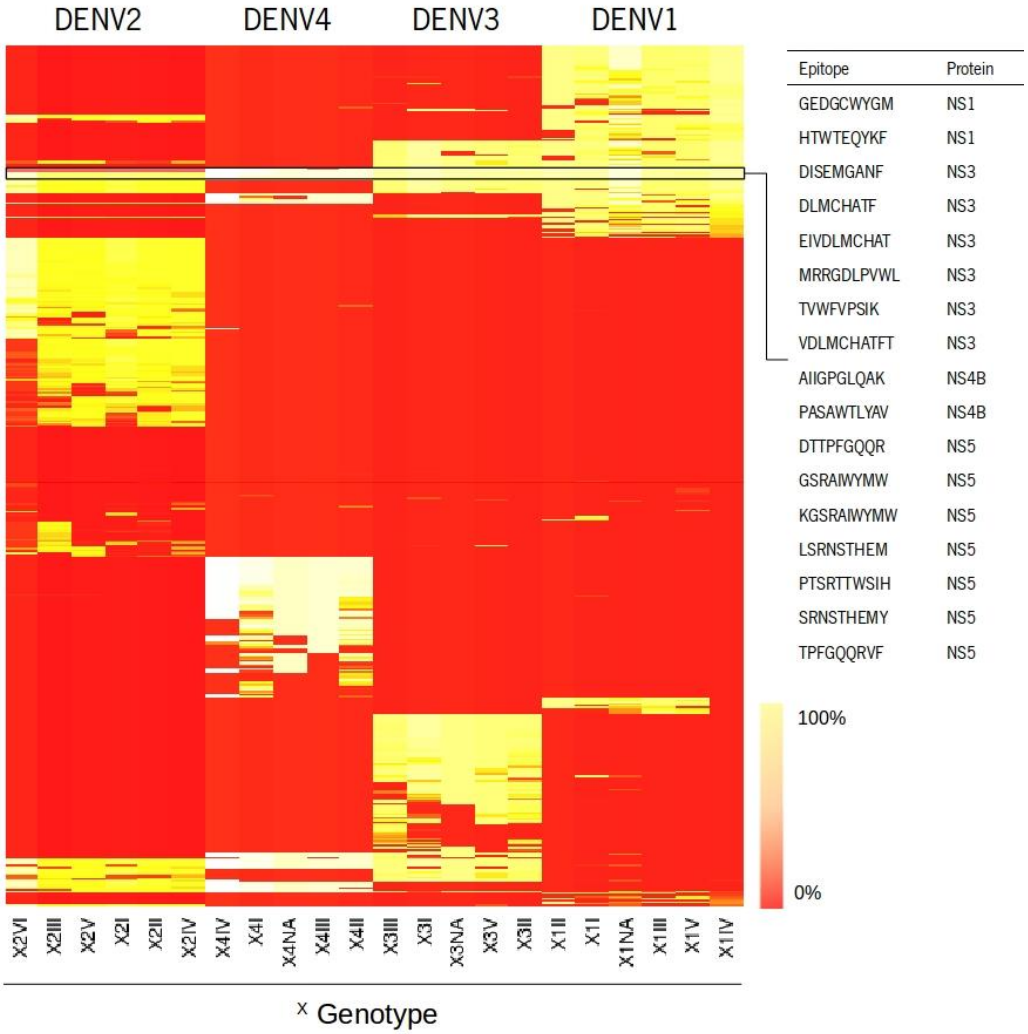


Figure 9. Heatmap representing the level of conservation of 1613 epitopes in 1739 complete proteomes of *Dengue virus*. The red zone represents a very low level of conservation of the epitopes, the white and yellow areas representing medium and highly conserved epitopes in the proteomes. Highly conserved epitopes among all serotypes and genotypes are highlighted in the heatmap with their corresponding location in the proteome.

Predicted population coverage of hyperconserved DENV epitopes

Epitopes that were widely conserved in DENV genomes were analyzed for their estimated coverage in the human population taking in consideration the global frequency of different class I Human Leukocyte Antigen (HLA) haplotypes and the predicted binding of the epitopes to these HLAs. The predicted binding of the hyperconserved epitopes was calculated for 2924 class I HLAs using NetMHCpan 3.0 (Table 2 and Supplementary data table 4). The population coverage for each epitope with the Immune Epitope Database (IEDB) population coverage tool using the list of HLAs predicted to have at least one strong binding epitope (affinity bellow 50 nM and rank bellow 0.5%). As a result we observed that 6/17 experimentally validated T cells epitopes located in NS3, NS4 and NS5 had an estimated a world population coverage between 31.61 and 42.87 percent and other 6/17 epitopes located in NS1, NS3 and NS5 obtained a coverage between 8.68 and 12.4. The remaining 5 epitopes had negligible estimated coverage levels.

Separately, these epitopes did not reach a coverage greater than 30 percent of the world population. However, by combining 14 of these epitopes we can reach more than 92.5% in the world population coverage (Figure 10 and Supplementary Data Figure 15).

Table 2. HLA Class I World Population Coverage of Hyperconserved T cells Epitopes. The percentage of the worldwide human population predicted to recognize hyperconserved $CD8^+$ T cells epitopes (HLA Class I World Population Coverage) was calculated using the IEDB analysis resource HLA class I alleles predicted to have at least one high binding affinity peptide (nHLAs SB) were used as input.

Epitope	HLA Class I World Population Coverage	nHLAs SB	Protein
HTWTEQYKF	12.4	83	NS1
GEDGCWYGM	9.32	80	NS1
DLMCHATF	0	0	NS3
EIVDLMCHAT	0	0	NS3
VDLMCHATFT	0	2	NS3
DISEMGANF	10.49	55	NS3
TVWFVPSIK	41.17	182	NS3
MRRGDLPVWL	42.87	179	NS3
PASAWTLYAV	0	0	NS4B
AIIGPGLQAK	33.29	120	NS4B
PTSRTTWSIH	0	0	NS5
GSRAIWYMW	10.91	67	NS5
KGSRAIWYMW	11.34	70	NS5
LSRNSTHEM	31.61	107	NS5
SRNSTHEMY	35.69	120	NS5
TPFGQQRVF	36.78	270	NS5
DTTPFGQQR	8.68	38	NS5

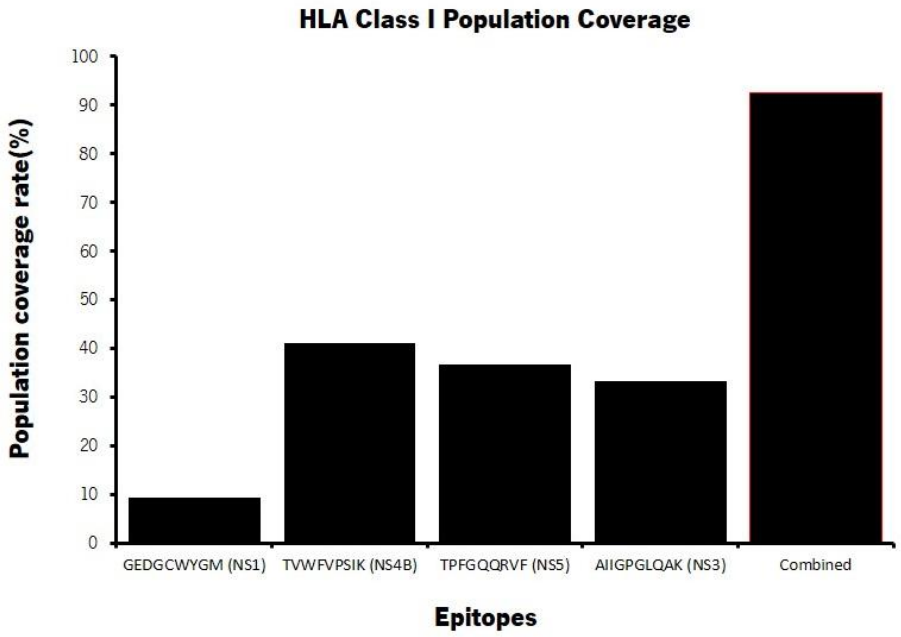


Figure 10. HLA class I world population coverage of four hyperconserved epitopes, separately and combined. The HLA class I world population coverage of four distinct antigenic peptides (NS1, NS4B, NS5, NS3) was calculated separately and in combination.

Discussion

This study corroborated that DENV has a great genetic diversity that is spread in different regions of the world [9]. This characteristic is likely to raise the challenge for disease control by rendering more complex the design of diagnostic and treatment tools. The previous studies [2], [10], [80] suggested that all 4 DENV serotypes are spread in all endemic countries, increasing the possibility of severe dengue in these populations. As described by several authors [16], [81] the possibility of contracting the virus is high and complications are largely related to the variety of circulating serotypes and genotypes.

More than 20 genotypes of DENV have been sequenced since the 1944 globally. Nowadays complete and partial genomes available in the different databases allowed us gain additional insights on the genotype diversity of DENV [40], [63], [64]. However, it was observed in our analysis (Figure 7) that in regions such as Africa and Central America the number of available sequences of clinical isolates are insufficient to know the real distribution of the genotypes in these areas.

Our results also highlight that new genotypes are likely emerging, possibly by the transmission of sylvatic lineages to humans caused by human demographic expansion [34], [82], [83]. This is a concern for the authorities as it might be the reason for the emergence of new outbreaks. In this work a total of 5 sequences coming from Angola (KF184975) [77], India (KF289073, JQ922546), [79], Malaysia (EF457905) [4], and Brunei (KR919820) [34] were identified since they could not be genotyped. One of these sequences has already been described by other authors as a possible new genotype [34], [63]. As highlighted above a better understanding of DENV diversity at the genotype and sub-genotype levels will be important to develop strategies for the control and elimination of the virus [84].

Although, there is a vaccine approved to immunize against Dengue and other vaccines that are in the final stages of clinical trials it is not clear these vaccines will be successful in protecting against all the genotypes of DENV. Thus, further studies are necessary. Approaches like the ones explored in this thesis based on immunoinformatics might be effective ways to for the design of novel vaccines for dengue, similarly to what was accomplished for Group B Meningococcal vaccine, [73], [87].

This study describes that most of the epitopes in IEDB that have been validated are serotype specific and that in many cases these epitopes are not highly conserved among the genotypes belonging to this group. Only a very small percentage of epitopes are highly conserved between serotypes and genotypes. These epitopes show a promising ability to have an antigenic effect against all or most of the DENV genotypes as they are highly conserved in their genomes [39], [88], [89].

Our analysis of the estimated HLA population coverage of the highly conserved epitopes showed by combining 14 of these they could reach a coverage of above 90 percent. Thus the strategy using these epitopes for designing a multiepitope vaccine could be considered [72], [92], and possibly represent and effective vaccine with broad efficacy.

Conclusion

Despite the existing vaccines, the DENV continues to be a threat to public health and efforts should be made to improve molecular epidemiology in areas of high vulnerability to understand the existent genetic variability and if it relates with vaccination success. The conclusions of this study are in line with previous works, supporting a global spread of the 4 existing DENV serotypes. Taking advantage of genetic typing and phylogenetic analysis, we also were able to analyze the distribution of the DENV genotypes. At the genotype level, it was possible to observe a more clear geographic profile in what regards to the genotypes that are more abundant in different world regions.

Although this study was performed under more than two thousand available sequences, analysis of genomes and the study of conserved epitopes between serotypes and genotypes will profit on a larger number of sequences in order to cover all genetic variants and avoid that non-common lines emerge as new threats.

Overall, this work supports that the diversity of DENV is very large and that new genotypes are likely to exist. Despite the high diversity, it was possible to find T cell epitopes with high levels of conservation. A chimeric vaccine with the epitopes that are more conserved could be a new alternative for diagnosis or new vaccines. We have found that by combining 14 hyperconserved epitopes we could get a predicted coverage of more than 92.5% of human population. Nevertheless further studies are necessary to test if these epitopes are able to induce protective responses.

Material and methods

Sequence retrieval

In order to have a representation of all the genetic variants of DENV, 2179 whole genomes sequence were extracted from the Virus Variation database (NCBI) (retrieval until Dec. 2016) by previously removing the duplicates and selecting only the sequences from environmental isolates. After classified in serotypes 1131 DENV1, 619 DENV2, 307 DENV3 and 122 DENV4 were obtained. Four Reference genomes for each serotype was obtained following the recommendations of the WHO International Standard Candidates for Nucleic Acid Testing [93]. The sequences were genotyped *in silico* using Dengue, Zika, Chikungunya and Yellow Fever virus typing tool [32].

Genome distribution and Phylogenetics analysis of DENV.

After classified the sequences were distributed by serotype/genotype and mapped using Google Spreadsheet. Multiple sequence alignment (MSA) of the 2179 was performed using MAFFT v.7 [94]. Two sequences were excluded by large amount of GAPS. A Maximum Likelihood tree was constructed using PhyML v.3.0 software [76] configured with the GTR (Generalized Time Reversible) nucleotide substitution model, model. Best for tree topology search NNI (Nearest Neighbor Interchanges) and SPR (subtype pruning regrafting) in Geneious software v.9.1 [95] Trees were visualized an edited using Dendroscope v 3.5 [96].

T cell epitopes conservancy and prediction

A total of 1613 experimentally and validated T-cells epitopes of DENV were obtained from IEDB database [66]. Using a Linux/BASH script each epitope was searched for conservancy in 1739 DENV proteomes. (The proteomes used were the WGS extracted from the NCBI that were translated into the amino acid code and again filtered the duplicates).The conservation level was represented in a heatmap using R studio v.0.99 and the D3heatmap package.v0.6.T-cell Epitopes highly conserved in genomes were used to predict their binding affinity using NETMHCpan 2.8 and NETMHCpan 3.0. Epitopes not conserved in some genotypes were identified and their mutations located using IEDB tool for the conservation of epitopes.

World population coverage of DENV epitopes

The percentage of the human population with the capacity of recognized the epitopes (population coverage) was calculated using an algorithmic described by Bui *et al* and available in IEDB. This tool is implemented to calculate the fraction of individuals predicted to respond to a given epitope set on the basis of HLA genotypic frequencies and on the basis of MHC binding and/or T cell restriction data [97]. A table with the population coverage percentage per geographic region was constructed with this information.

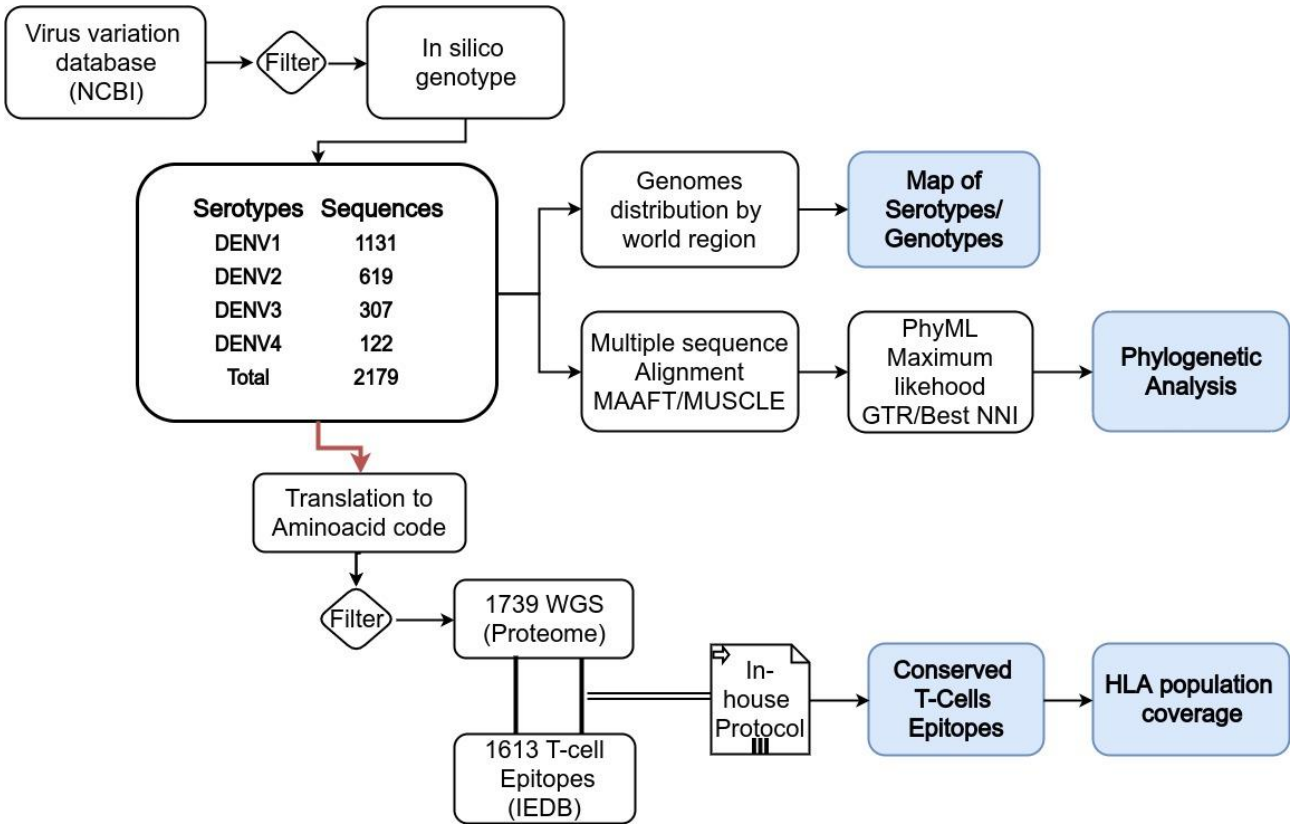


Figure 11. Bioinformatic pipeline used for the study of DENV genomes and analysis of T cell epitopes.

Reference

[1] WHO, "Global Strategy for Dengue Prevention and Control 2012–2020," World Heal. Organ., p. 43, 2012.

[2] S. Bhatt *et al.*, "The global distribution and burden of dengue," Nature, vol. 496, pp. 504–507, 2013.

[3] S. C. Weaver and A. D. T. Barrett, "Transmission cycles, host range, evolution and emergence of arboviral disease," Nat. Rev. Microbiol., vol. 2, pp. 789–801, 2004.

[4] N. Vasilakis *et al.*, "Evolutionary processes among sylvatic dengue type 2 viruses," J Virol, vol. 81, pp. 9591–9595, 2007.

[5] S. S. Twiddy, E. C. Holmes, and A. Rambaut, "Inferring the rate and time-scale of *Dengue virus* evolution," Mol. Biol. Evol., vol. 20, pp. 122–129, 2003.

[6] J. G. Rigau-Pérez, G. G. Clark, D. J. Gubler, P. Reiter, E. J. Sanders, and A. V. Vorndam, "Dengue and dengue haemorrhagic fever," Lancet, vol. 352, pp. 971–977, 1998.

[7] E. A. Henchal and J. R. Putnak, "The *Dengue virus*es," Clin. Microbiol. Rev., vol. 3, pp. 376–396, 1990.

[8] E. C. Holmes and S. S. Burch, "The causes and consequences of genetic variation in *Dengue virus*," Trends Microbiol., vol. 8, pp. 74–77, 2000.

[9] S. Bhatt *et al.*, "The global distribution and burden of dengue," Nature Publishing Group, 2013.

[10] M. G. Guzman *et al.*, "Dengue: a continuing global threat," Nat. Rev. Microbiol., vol. 8, pp. S7–S16, 2010.

[11] S. B. Halstead, "In vivo enhancement of *Dengue virus* infection in rhesus monkeys by passively transferred antibody," J. Infect. Dis., vol. 140, pp. 527–533, 1979.

[12] E. Z. Ong, S. L. Zhang, H. C. Tan, E. S. Gan, K. R. Chan, and E. E. Ooi, "*Dengue virus* compartmentalization during antibody-enhanced infection," Sci. Rep., vol. 7, p. 40923, 2017.

[13] T. Pang, M. J. Cardosa, and M. G. Guzman, "Of cascades and perfect storms: the immunopathogenesis of dengue haemorrhagic fever-dengue shock syndrome (DHF/DSS)," Immunol Cell Biol, vol. 85, pp. 43–45, 2007.

[14] D. S. Burke, A. Nisalak, D. E. Johnson, and R. Scott McN., "A prospective study of dengue infections in Bangkok," Am. J. Trop. Med. Hyg., vol. 38, pp. 172–180, 1988.

[15] Yacoub S, Mongkolsapaya J, Screaton G. Recent advances in understanding dengue. F1000Research. 2016.

[16] World Health Organization (WHO) Regional Office for South-East Asia, "Dengue: guidelines for diagnosis, treatment, prevention, and control," Spec. Program. Res. Train. Trop. Dis., p. 147, 2009.

[17] F. J. Carod-Artal, O. Wichmann, J. Farrar, and J. Gascón, "Neurological complications of *Dengue virus* infection," Lancet Neurol., vol. 12, pp. 906–919, 2013.

[18] C. Klungthong, R. Putnak, M. P. Mammen, T. Li, and C. Zhang, "Molecular genotyping of *Dengue virus*es by phylogenetic analysis of the sequences of individual genes," J. Virol. Methods, vol. 154, no. 1–2, pp. 175–181, 2008.

[19] World Health Organization, "Weekly Epidemiological Report," World Heal. Organiszation, vol. 30, pp. 349–364, 2016.

[20] B.-T. Teoh *et al.*, "The Use of NS1 Rapid Diagnostic Test and qRT-PCR to Complement IgM ELISA for Improved Dengue Diagnosis from Single Specimen.," Sci. Rep., vol. 6, p. 27663, 2016.

[21] S. S. Whitehead, "Development of TV003/TV005, a single dose, highly immunogenic live attenuated dengue vaccine; what makes this vaccine different from the Sanofi-Pasteur CYD™ vaccine?," Expert Rev. Vaccines, vol. 9, pp. 509–17, 2015.

[22] N. Khetarpal and I. Khanna, "Dengue Fever: Causes, Complications, and Vaccine Strategies," J. Immunol. Res., vol. 2016, 3, 2016.

[23] J. Flipse and J. M. Smit, "The Complexity of a dengue vaccine: A review of the human antibody response," PLoS Negl. Trop. Dis., vol. 9, pp. 1–18, 2015.

[24] J. Torresi, G. Ebert, and M. Pellegrini, "Vaccines licensed and in clinical trials for the prevention of dengue," Hum. Vaccin. Immunother., vol. 5515, pp. 1–14, 2017.

[25] D. J. Gubler, "The economic burden of dengue," Am. J. Trop. Med. Hyg., vol. 86, no. 5, pp. 743–744, 2012.

[26] B. P. Drumond, A. Mondini, D. J. Schmidt, R. V. M. de Bronzoni, I. Bosch, and M. L. Nogueira, "Circulation of Different Lineages of *Dengue virus* 2, Genotype American/Asian in Brazil: Dynamics and Molecular and Phylogenetic Characterization," PLoS One, vol. 8, no. 3, 2013.

[27] R. T. Sasmono *et al.*, "Genomic analysis and growth characteristic of *Dengue virus*es from Makassar, Indonesia," Infect. Genet. Evol., vol. 32, pp. 165–177, 2015.

[28] World Health Organization, "Dengue control: Epidemiology," 2016. [Online]. Available: http://www.who.int/denguecontrol/epidemiology/en/.

[29] J. Patterson, M. Sammon, and M. Garg, "Dengue, Zika and Chikungunya: Emerging Arboviruses in the New World," West. J. Emerg. Med., vol. 17, pp. 671–679, 2016.

[30] A. Amarilla *et al.*, "Genetic diversity of the E protein of dengue type 3 virus," Virol. J., vol. 6, p. 113, 2009.

[31] S. Sim and M. L. Hibberd, "Genomic approaches for understanding dengue: insights from the virus, vector, and host," Genome Biol., vol. 17, p. 38, 2016.

[32] "Dengue, Zika and Chikungunya viruses typing too version 1l: A phylogenetic algorithm for accurate identification of Dengue, Zika and Chikungunya viruses species and genotypes." 2015. [Online] available at: www.bioafrica.net/software.php

[33] F. J. Diaz, W. C. Black IV, J. A. Farfan-Ale, M. A. Lorono-Pino, K. E. Olson, and B. J. Beaty, "*Dengue virus* Circulation and Evolution in Mexico: A Phylogenetic Perspective," Arch. Med. Res., vol. 37, pp. 760–773, 2006.

[34] A. T. Pyke *et al.*, "Highly divergent *Dengue virus* type 1 genotype sets a new distance record," Sci. Rep., vol. 6, no. October 2015, p. 22356, 2016.

[35] H. A. Tissera *et al.*, "New *Dengue virus* type 1 genotype in Colombo, Sri Lanka," Emerg. Infect. Dis., vol. 17, pp. 2053–2055, 2011.

[36] E. C. Holmes and S. S. Twiddy, "The origin, emergence and evolutionary genetics of *Dengue virus*," Infect. Genet. Evol., vol. 3, pp. 19–28, 2003.

[37] J. Whitehorn and C. P. Simmons, "The pathogenesis of dengue," Vaccine, vol. 29, pp. 7221–7228, 2011.

[38] I. J. Sánchez and B. H. Ruiz, "A single nucleotide change in the E protein gene of *Dengue virus* 2 Mexican strain affects neurovirulence in mice," J. Gen. Virol., vol. 77, pp. 2541–2545, 1996.

[39] R. C. Fleith *et al.*, "Genome-wide analyses reveal a highly conserved *Dengue virus* envelope peptide which is critical for virus viability and antigenic in humans," Sci. Rep., vol. 6, p. 36339, 2016.

[40] B. E. Pickett *et al.*, "Virus pathogen Database and Analysis Resource (ViPR): A comprehensive bioinformatics Database and Analysis Resource for the Coronavirus research community," Viruses, vol. 4, pp. 3209–3226, 2012.

[41] A. T. Bäck and A. Lundkvist, "*Dengue virus*es - an overview," Infect. Ecol. Epidemiol., vol. 3, pp. 1–21, 2013.

[42] S. Suwanmanee and N. Luplertlop, "Dengue and Zika viruses: lessons learned from the similarities between these Aedes mosquito-vectored arboviruses," J. Microbiol., vol. 55, pp. 81–89, 2017.

[43] G. Screaton, J. Mongkolsapaya, S. Yacoub, and C. Roberts, "New insights into the immunopathology and control of *Dengue virus* infection," Nat. Rev. Immunol., vol. 15, pp. 745–759, 2015.

[44] M. A. Schmid, M. S. Diamond, and E. Harris, "Dendritic cells in *Dengue virus* infection: Targets of virus replication and mediators of immunity," Front. Immunol., vol. 5, pp. 1–10, 2014.

[45] L. Rivino *et al.*, "Differential Targeting of Viral Components by CD4+ versus CD8+ T Lymphocytes in *Dengue virus* Infection," J. Virol., vol. 87, pp. 2693–2706, 2013.

[46] C. M. Midgley *et al.*, "An in-depth analysis of original antigenic sin in *Dengue virus* infection," J. Virol., vol. 85, pp. 410–21, 2011.

[47] E. Z. Ong, S. L. Zhang, H. C. Tan, E. S. Gan, K. R. Chan, and E. E. Ooi, "*Dengue virus* compartmentalization during antibody-enhanced infection," Sci. Rep., vol. 7, p. 40923, 2017.

[48] I. A. Rodenhuis-Zybert, J. Wilschut, and J. M. Smit, "*Dengue virus* life cycle: Viral and host factors modulating infectivity," Cell. Mol. Life Sci., vol. 67, pp. 2773–2786, 2010.

[49] S. Zompi and E. Harris, "Animal models of *Dengue virus* infection," Viruses, vol. 4, pp. 62–82, 2012.

[50] K. W. K. Chan, S. Watanabe, R. Kavishna, S. Alonso, and S. G. Vasudevan, "Animal models for studying dengue pathogenesis and therapy," Antiviral Research, vol. 123. pp. 5–14, 2015.

[51] I. Adam, A. M. Jumaa, H. M. Elbashir, and M. S. Karsany, "Maternal and perinatal outcomes of dengue in PortSudan, Eastern Sudan.," Virol. J., vol. 7, p. 153, 2010.

[52] S. H. Pouliot *et al.*, "Maternal dengue and pregnancy outcomes: a systematic review," Obstet. Gynecol. Surv., vol. 65, pp. 107–118, 2010.

[53] K. Pengsaa *et al.*, "*Dengue virus* infections in the first 2 years of life and the kinetics of transplacentally transferred dengue neutralizing antibodies in thai children," J. Infect. Dis., vol. 194, pp. 1570–1576, 2006.

[54] H. Salje *et al.*, "Dengue diversity across spatial and temporal scales: Local structure and the effect of host population size," Science (80-.), vol. 355, pp. 1302–1306, 2017.

[55] Y. Li *et al.*, "VIP: an integrated pipeline for metagenomics of virus identification and discovery," Sci. Rep., vol. 6, no. September 2015, p. 23774, 2016.

[56] L. Backert and O. Kohlbacher, "Immunoinformatics and epitope prediction in the age of genomic medicine," Genome Med., vol. 7, p. 119, 2015.

[57] L. Moise and A. S. De Groot, "Putting immunoinformatics to the test," Nat. Biotechnol., vol. 24, pp. 791–792, 2006.

[58] D. R. Tobergte and S. Curtis, Bioinformatics and functional genomics, vol. 53, 2013.

[59] Y. Xu, J. Cui, and D. Puett, Cancer bioinformatics. Springer-Verlag, p. 1-368, 2014.

[60] L. Rubio, J. Guerri, and P. Moreno, "Genetic variability and evolutionary dynamics of viruses of the family Closteroviridae," Front. Microbiol., vol. 4, pp. 1–15, 2013.

[61] H. C. Hapuarachchi *et al.*, "Intra-epidemic evolutionary dynamics of a *Dengue virus* type 1 population reveal mutant spectra that correlate with disease transmission," Sci. Rep., vol. 6, p. 22592, 2016.

[62] P. A.C.O *et al.*, "Asian genotypes of *Dengue virus* 4 in Brazil," Epidemiol. Infect., vol. 143, pp. 3114–3117, 2015.

[63] A. Yamashita, T. Sakamoto, T. Sekizuka, K. Kato, T. Takasaki, and M. Kuroda, "DGV: Dengue genographic viewer," Front. Microbiol., vol. 7, pp. 1–8, 2016.

[64] J. R. Brister *et al.*, "Virus Variation Resource - Recent updates and future directions," Nucleic Acids Res., vol. 42, no. D1, pp. 1–6, 2014.

[65] N. Tomar and R. K. De, "Immunoinformatics: An integrated scenario," Immunology, vol. 131, pp. 153–168, 2010.

[66] N. Salimi, W. Fleri, B. Peters, and A. Sette, "The immune epitope database: A historical retrospective of the first decade," Immunology, vol. 137, pp. 117–123, 2012.

[67] Y. Okamoto, I. Kurane, A. M. Leporati, and F. A. Ennis, "Definition of the region on NS3 which contains multiple epitopes recognized by *Dengue virus* serotype-cross-reactive and *Flavivirus*-cross-reactive, HLA-DPw2-restricted CD4+ T cell clones," J. Gen. Virol., vol. 79, pp. 697–704, 1998.

[68] S. Paul, D. Weiskopf, M. A. Angelo, J. Sidney, B. Peters, and A. Sette, "HLA Class I Alleles Are Associated with Peptide-Binding Repertoires of Different Size, Affinity, and Immunogenicity," J. Immunol., vol. 191, no. 12, pp. 5831–5839, 2013.

[69] J. W. Yewdell and J. R. Bennink, "Immunodominance in Major Histocompatibility Complex Class I–Restricted T Lymphocyte Responses 1," Annu. Rev. Immunol., vol. 17, pp. 51–88, 1999.

[70] R. Appanna, L. H. Tan, L. L. C. See, L. T. Phoay, J. Vadivelu, and S. Devi, "Cross-reactive T-cell responses to the nonstructural regions of *Dengue virus*es among dengue fever and dengue hemorrhagic fever patients in Malaysia," Clin. Vaccine Immunol., vol. 14, pp. 969–977, 2007.

[71] J. Shi *et al.*, "Inferring protective CD8+ T-cell epitopes for NS5 protein of four serotypes of *Dengue virus* Chinese isolates based on HLA-A,-B and-C allelic distribution: Implications for epitope-based universal vaccine design," PLoS One, vol. 10, pp. 1–15, 2015.

[72] X. Y. Chen, D. Z. Li, X. Z. Zhong, B. Chen, Z. L. Duan, and J. S. Wen, "Induction of multiple cytotoxic T lymphocyte responses in mice by a multiepitope DNA vaccine against *Dengue virus* serotype 1," Microbiol. Immunol., vol. 60, pp. 835–845, 2016.

[73] A. Sette, "Reverse Vaccinology: Developing Vaccines in the Era of Genomics," Immunity, vol. 33, pp. 530–541, 2010.

[74] N. Afreen, I. H. Naqvi, S. Broor, A. Ahmed, and S. Parveen, "Phylogenetic and molecular clock analysis of dengue serotype 1 and 3 from New Delhi, India," PLoS One, vol. 10, pp. 1–13, 2015.

[75] R. E. Soria-Guerra, R. Nieto-Gomez, D. O. Govea-Alonso, and S. Rosales-Mendoza, "An overview of bioinformatics tools for epitope prediction: Implications on vaccine development," J. Biomed. Inform., vol. 53, pp. 405–414, 2015.

[76] S. Guindon, F. Lethiec, P. Duroux, and O. Gascuel, "PHYML Online - A web server for fast maximum likelihood-based phylogenetic inference," Nucleic Acids Res., vol. 33, pp. 557–559, 2005.

[77] O. M. Sessions *et al.*, "Exploring the origin and potential for spread of the 2013 dengue outbreak in luanda, Angola," Glob. Health Action, vol. 6, 2013.

[78] N. Vasilakis *et al.*, "Genetic and phenotypic characterization of sylvatic *Dengue virus* type 2 strains," Virology, vol. 377, pp. 296–307, 2008.

[79] M. A. Islam *et al.*, "Molecular characterization and clinical evaluation of dengue outbreak in 2002 in Bangladesh," Jpn. J. Infect. Dis., vol. 59, pp. 85–91, 2006.

[80] J. P. Messina *et al.*, "Global spread of *Dengue virus* types: Mapping the 70 year history," Trends Microbiol., vol. 22, pp. 138–146, 2014.

[81] H. Y. Lei, T. M. Yeh, H. S. Liu, Y. S. Lin, S. H. Chen, and C. C. Liu, "Immunopathogenesis of *Dengue virus* infection," J Biomed Sci, vol. 8, pp. 377–388, 2001.

[82] B. E. Martina *et al.*, "Human to human transmission of arthropod-borne pathogens," Curr. Opin. Virol., vol. 22, pp. 13–21, 2017.

[83] N. Vasilakis *et al.*, "Potential of ancestral sylvatic dengue-2 viruses to re-emerge," Virology, vol. 358, no. 2, pp. 402–412, 2007.

[84] V. P. Waman, P. Kolekar, M. R. Ramtirthkar, M. M. Kale, and U. Kulkarni-Kale, "Analysis of genotype diversity and evolution of *Dengue virus* serotype 2 using complete genomes," PeerJ, vol. 4, p. e2326, 2016.

[85] B. Guy and N. Jackson, "Dengue vaccine: hypotheses to understand CYD-TDV-induced protection," Nat. Rev. Microbiol., vol. 14, pp. 45–54, 2016.

[86] E. P. Abrão, D. L. A. Espósito, F. Lauretti, and B. A. L. da Fonseca, "Dengue vaccines: what we know, what has been done, but what does the future hold?" Rev. Saude Publica, vol. 49, pp. 1–6, 2015.

[87] Y. He, R. Rappuoli, A. S. De Groot, and R. T. Chen, "Emerging vaccine informatics," J. Biomed. Biotechnol., vol. 2010, 2010.

[88] D. Weiskopf *et al.*, "The Human CD8+ T Cell Responses Induced by a Live Attenuated Tetravalent Dengue Vaccine Are Directed against Highly Conserved Epitopes," J. Virol., vol. 89, pp. 120–8, 2015.

[89] J. S. Testa *et al.*, "Conserved MHC class I-presented *Dengue virus* epitopes identified by immunoproteomics analysis are targets for cross-serotype reactive T-cell response," J. Infect. Dis., vol. 205, pp. 647–655, 2012.

[90] H. Loke *et al.*, "Strong HLA class I–restricted T cell responses in dengue hemorrhagic fever: a double-edged sword?" J. Infect. Dis., vol. 184, pp. 1369–73, 2001.

[91] H. Harapan, "Genetic Polymorphisms of HLA and HLA-related Proteins: Implications on *Dengue virus* Infection," J. Mosq. Res., vol. 3, pp. 1–10, 2013.

[92] S. Jafarpour, H. Ayat, and A. M. Ahadi, "Design and Antigenic Epitopes Prediction of a New Trial Recombinant Multiepitopic Rotaviral Vaccine," vol. 28, pp. 1–6, 2015.

[93] G. Añez, D. A. Heisey, E. Volkova, and M. Rios, "Complete Genome Sequences of *Dengue virus* Type 1 to 4 Strains Used for the Development of CBER / FDA RNA Reference Reagents," Am. Soc. Microbiol., vol. 4, pp. 4–5, 2016.

[94] K. Katoh and D. M. Standley, "MAFFT multiple sequence alignment software version 7: Improvements in performance and usability," Mol. Biol. Evol., vol. 30, pp. 772–780, 2013.

[95] M. Kearse *et al.*, "Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data," Bioinformatics, vol. 28, pp. 1647–1649, 2012.

[96] D. H. Huson, D. C. Richter, C. Rausch, T. Dezulian, M. Franz, and R. Rupp, "Dendroscope: An interactive viewer for large phylogenetic trees," BMC Bioinformatics, vol. 8, p. 460, 2007.

[97] H. H. Bui, J. Sidney, K. Dinh, S. Southwood, M. J. Newman, and A. Sette, "Predicting population coverage of T-cell epitope-based diagnostics and vaccines," BMC Bioinformatics, vol. 7, p. 153, 2006.

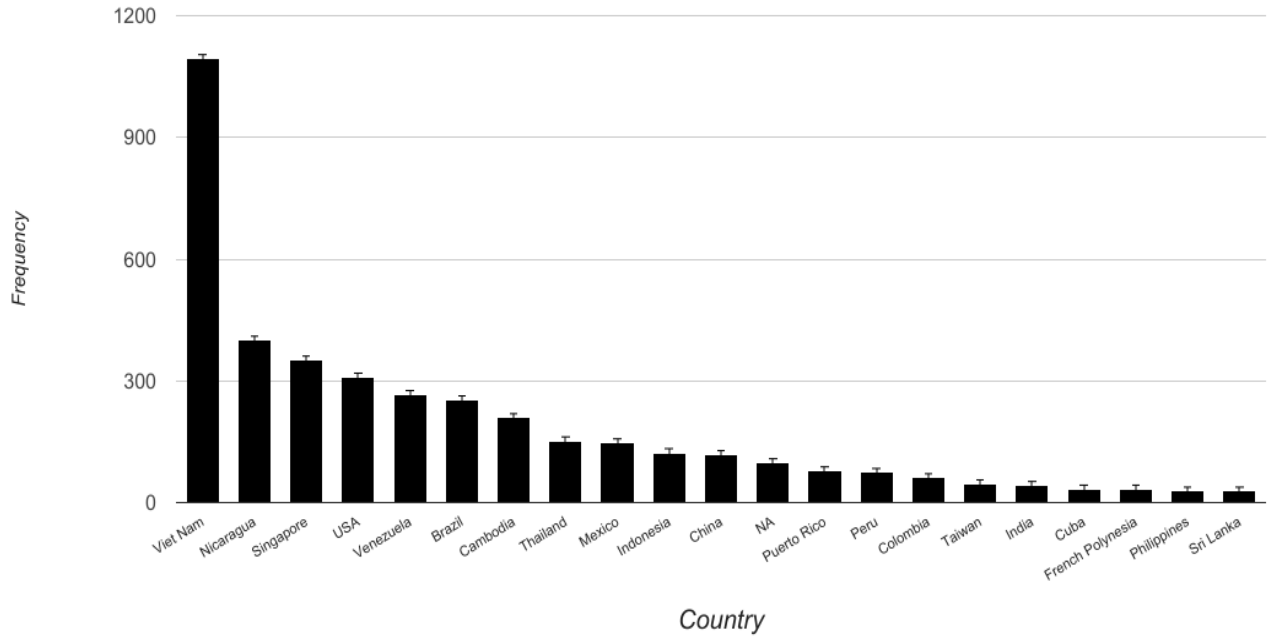


Figure 12. Distribution of DENV genomes according to isolation countries.

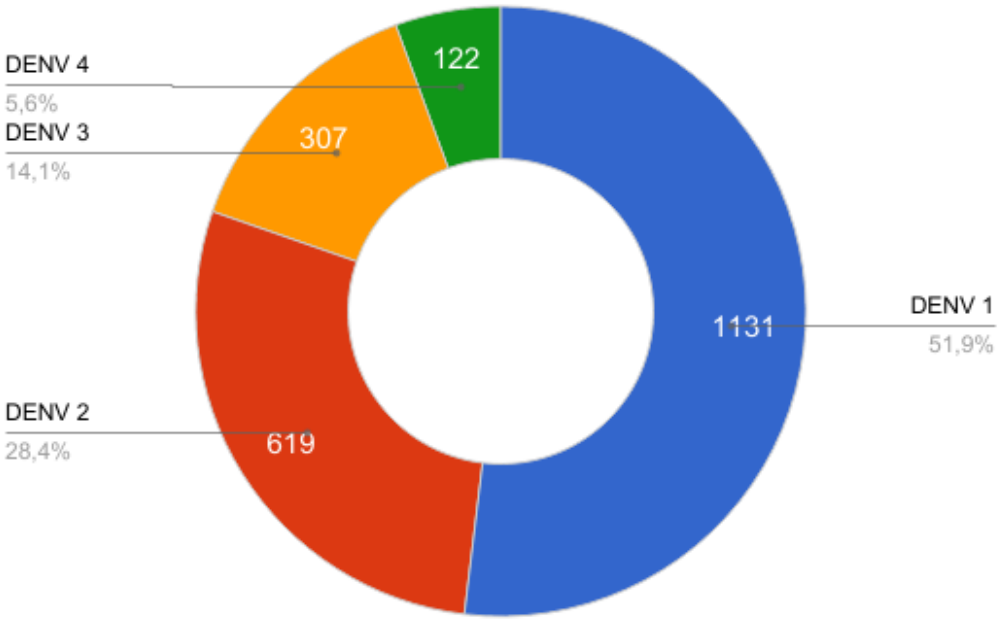


Figure 13. Distribution of DENV genomes according to serotypes.

Table 3. Genomes of DENV serotypes distributed according to the regions.

	Africa	Asia	Europe	North/ Central America	South America	Oceania	Total
DENV1	7	1038	2	20	50	14	1131
DENV2	12	401	1	88	94	23	619
DENV3	0	188	0	31	77	11	307
DENV4	1	37	0	2	73	9	122
Total	20 (0.9%)	1664 (76.4%)	3 (0.1%)	141(6.5%)	294(13.5%)	57(2.6%)	2179(100%)

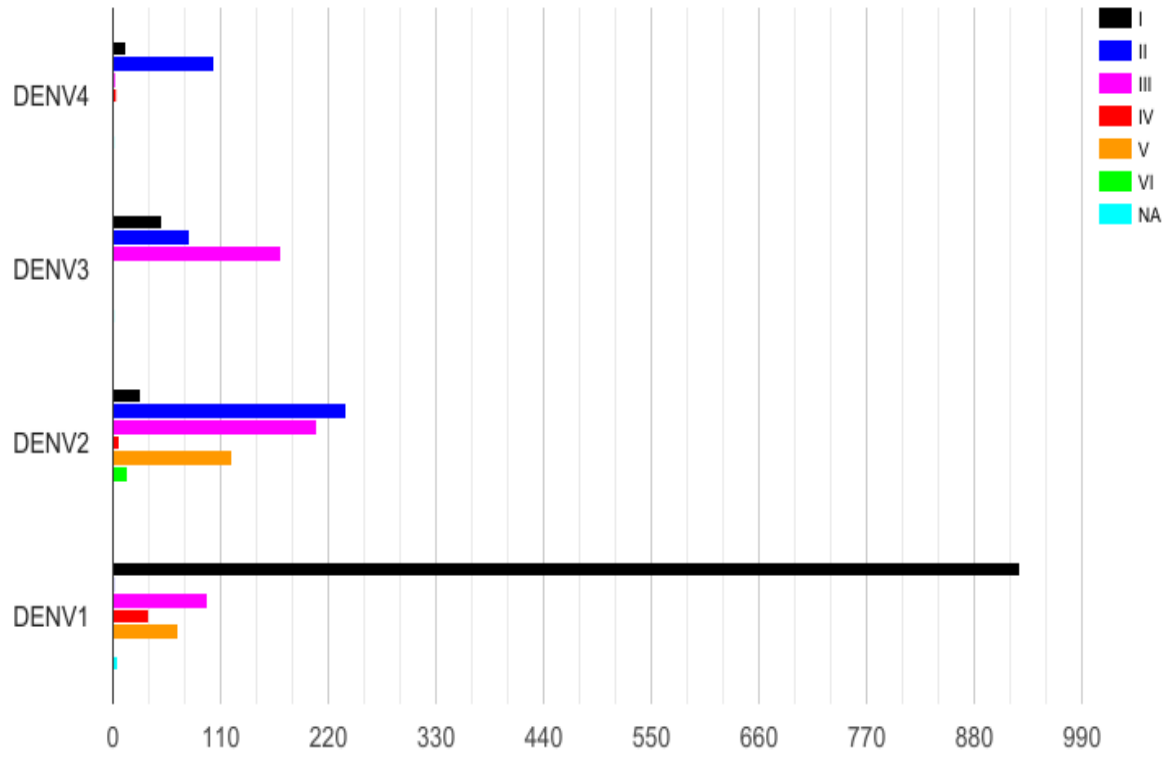


Figure 14. Distribution of DENV genomes according to genotypes.

Table 4. HLA class I population coverage of hyperconserved DENV epitopes estimated for the regions of Asia and Europe.

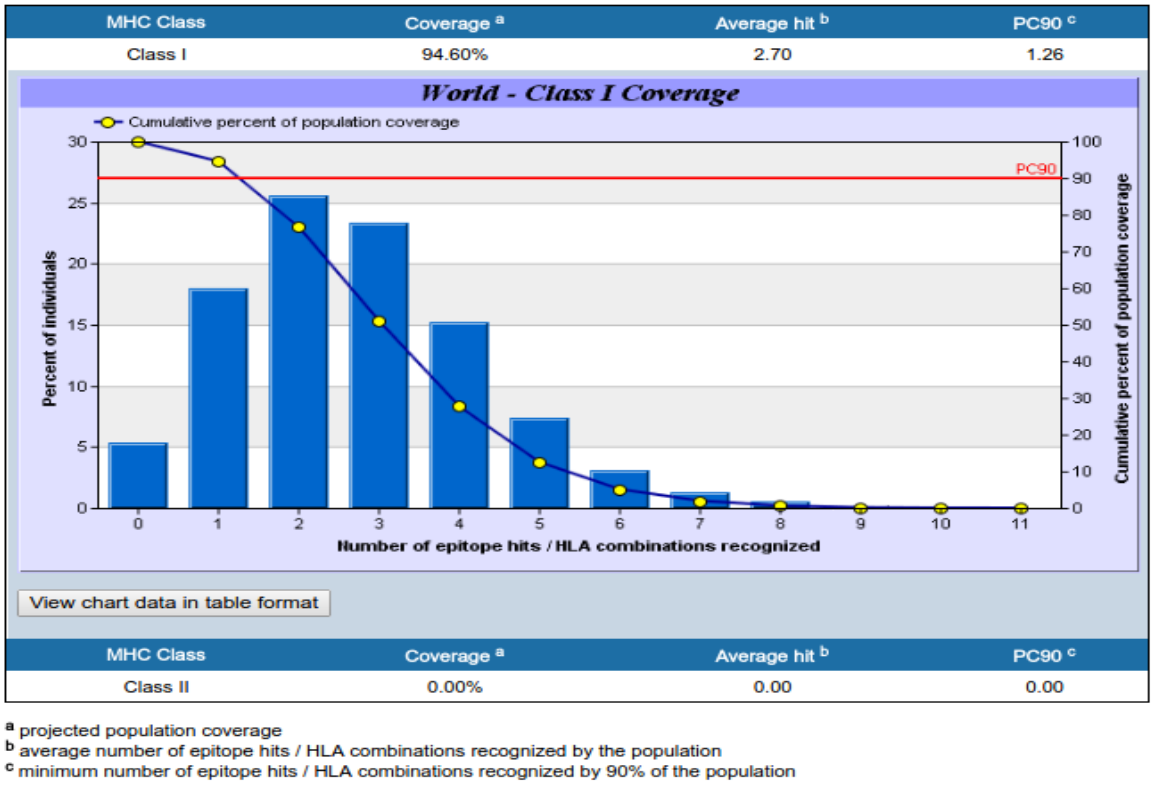
Epitope	East Asia	Northeast Asia	South Asia	Southeast Asia	Southwest Asia	Europe
AIIGPGLQAK	18.24	54.68	33.83	43.42	22.71	34.82
DISEMGANF	16	6.53	10.66	5.81	11.96	15.88
DLMCHATF	0	0	0	0	0	0
DTTPFGQQR	3.79	4.39	14.04	0.6	9.66	7.81
EIVDLMCHAT	0	0	0	0	0	0
GEDGCWYGM	7.58	16.29	05.02	24.93	2.59	7.68
GSRAIWYMW	4.12	9.66	13.75	14.45	9.52	8.21
HTWTEQYKF	7.41	10.48	16.58	14.9	17	09.08
KGSRAIWYMW	7.41	9.8	14.14	14.45	12.13	8.42
LSRNSTHEM	29.77	37.81	43.03	33.03	27.07	39.31
MRRGDLPVWL	18.2	16.97	37.82	23.81	35.6	50.78
PASAWTLYAV	0	0	0	0	0	0
PTSRTTWSIH	0	0	0	0	0	0
SRNSTHEMY	17.49	22.07	30.68	13.15	31.12	41.65
TPFGQQRVF	24.99	14.52	39.12	15.29	41.62	40.14
TVWFVPSIK	19.88	56.97	44.6	44.2	34.27	41.25
VDLMCHATFT	0	0	0	0	0	0

Table 5. HLA class I population coverage of hyperconserved DENV epitopes estimated for the regions of Africa.

Epitope	East Africa	West Africa	Central Africa	North Africa	South Africa
AIIGPGLQAK	14.52	17.21	18.46	17.13	26.8
DISEMGANF	04.08	6.29	5.61	4.83	2.28
DLMCHATF	0	0	0	0	0
DTTPFGQQR	4.3	14.51	15.75	11.85	7.86
EIVDLMCHAT	0	0	0	0	0
GEDGCWYGM	1.6	11.96	2.88	2.53	0
GSRAIWYMW	31.55	17.63	25.25	12.28	30.11
HTWTEQYKF	34.89	19.2	32.85	19.55	33.92
KGSRAIWYMW	32.48	17.63	30.49	16.77	31.89
LSRNSTHEM	34.16	25.92	29.96	26.75	26.9
MRRGDLPVWL	58.25	35.77	47.41	51.98	57.07
PASAWTLYAV	0	0	0	0	0
PTSRTTWSIH	0	0	0	0	0
SRNSTHEMY	51.99	33.4	44.52	48.78	53.72
TPFGQQRVF	38.2	42.65	44.07	29.29	25.48
TVWFVPSIK	27.33	32.26	32.4	27.92	33.98
VDLMCHATFT	0	0	0	0	0

Table 6. HLA class I population coverage of hyperconserved DENV epitopes estimated for the regions of West Indies, America and Oceania.

Epitope	West Indies	North America	Central America	South America	Oceania
AIIGPGLQAK	26	25.18	0	14.61	37.91
DISEMGANF	7.81	8.44	0.8	4.3	3.17
DLMCHATF	0	0	0	0	0
DTTPFGQQR	14.39	10.61	0	15.82	4.11
EIVDLMCHAT	0	0	0	0	0
GEDGCWYGM	9.5	10.09	0	10.87	17.53
GSRAIWYMW	13.9	13.1	1.4	4.46	3.18
HTWTEQYKF	15.54	15.52	1.4	6.73	3.68
KGSRAIWYMW	13.9	13.47	1.4	4.99	3.45
LSRNSTHEM	0.88	27.53	0	16.42	11.16
MRRGDLPVWL	3.51	34.93	0	25.34	19.41
PASAWTLYAV	0	0	0	0	0
PTSRTTWSIH	0	0	0	0	0
SRNSTHEMY	0	29.6	0	19.37	9.6
TPFGQQRVF	41.77	40.57	2.98	45.46	19.91
TVWFVPSIK	31.17	34.33	0	30.79	39.8
VDLMCHATFT	0	0	0	0	0



a projected population coverage
b average number of epitope hits / HLA combinations recognized by the population
c minimum number of epitope hits / HLA combinations recognized by 90% of the population

Figure 15. World – Class I Coverage of hyperconserved DENV epitopes estimated in IEDB platform.

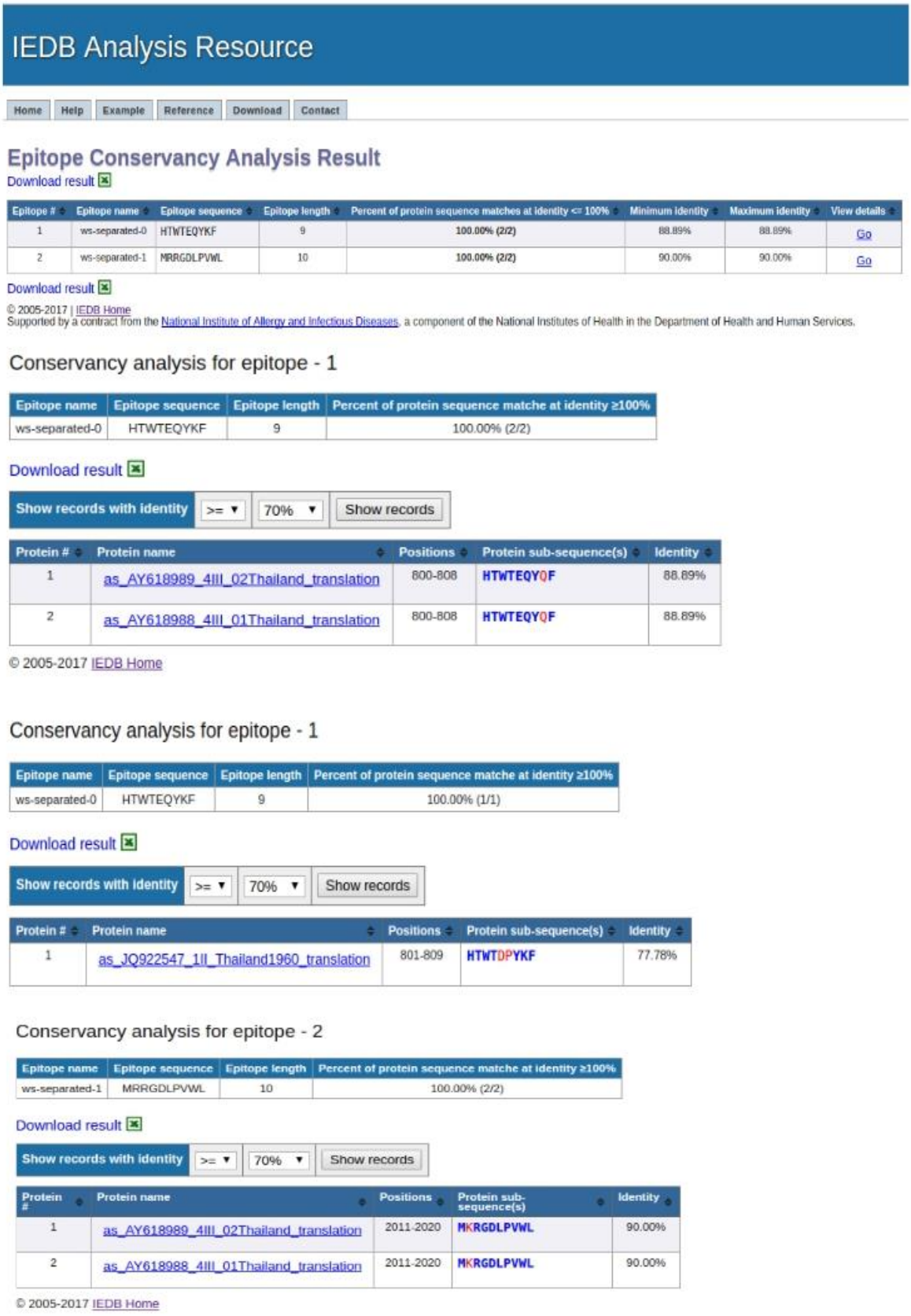


Figure 16. Print screen of conservancy analysis of two hyperconserved DENV epitopes in genomes of DENV1II and DENV4III genotypes.