Danielle Elise Gill

An Epidemiological Investigation of Circulating Arboviruses in the Brazilian state of Amapá during the Outbreaks of 2013-2016

Dissertation presented at the Institute of Tropical Medicine of Sao Paulo of the University of Sao Paulo to obtain the title of Master of Science.

Area of concentration: Tropical Diseases and International Health
Graduate advisor: Prof. Dr. Ester Cerdeira Sabino

Sao Paulo
2019

Catalog card elaborated by the Library of the Instituto de Medicina Tropical de São Paulo
of the Universidade de São Paulo – Librarian Carlos José Quinteiro, CRB-8 5538

Gill, Danielle Elise
An epidemiological investigation of circulating arboviruses in the Brazilian state of Amapá during the outbreaks of 2013-2016 / Danielle Elise Gill. – São Paulo, 2019.
Dissertation (Master of Science) – Instituto de Medicina Tropical de São Paulo of the Universidade de São Paulo
Area of concentration: Tropical Diseases and International Health
Graduate advisor: Ester Cerdeira Sabino

Descriptors: 1. EPIDEMIOLOGY. 2. MOLECULAR VIROLOGY. 3. ARBOVIRUS. 4. GENOMICS.

USP/IMTSP/BIB-07/2019.

Dedication

I dedicate this work to my family and academic mentors that have supported and guided me through the years.

Epigraph

“Name one genius that ain’t crazy.”
— Kanye West

“Shoot for the moon. Even if you miss, you'll land among the stars.”
— Norman Vincent Peale

“You miss one hundred percent of the shots you don't take.”
— Wayne Gretzky

“IN THE END… We only regret the chances we didn’t take, the relationships we were afraid to have, and the decisions we waited too long to make.”
— Lewis Carroll

“I want to try the impossible to show that it can be done.”
— Terry Fox

"All progress takes place outside the comfort zone."
— Michael John Bobak

"Success is not final; failure is not fatal: It is the courage to continue that counts."
— Winston S. Churchill

"The real test is not whether you avoid this failure, because you won't. It's whether you let it harden or shame you into inaction, or whether you learn from it; whether you choose to persevere."
— Barack Obama

Work hard in silence; let your success be your noise.
— Frank Ocean

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

To Dr. Ester Sabino for being my graduate advisor and for her kind patience, guidance, and support throughout this process.

To Dr. Antonio Charlys da Costa for being a mentor and partner in this project and for being a dear friend through the process of completing a degree in a completely new geographical, cultural, and academic environment.

To Dr. Camila Malta Romano for being a mentor and helping me gain the knowledge and skills I needed to be successful with this project.

To my collaborators and friends, the talented researchers of the Instituto Adolfo Lutz for lending time, knowledge, and recourses to help me be successful with this project: Dr. Renato Pereira de Souza, Dr. Fabiana Cristina Pereira dos Santos, Dr. Adriana Yurika Maeda, Dr. Adriana Luchs, and Dr. Mariana Sequetin Cunha.

To the friends that I made here in Sao Paulo that gave me a family away from home: Fabiana Cristina Pereira dos Santos and her kind family, Antonio Charlys da Costa, Karolina Morales Barrio-Nuevo and Victor Lima, Beatriz Pereira Gonçalves, Carol Saldanha Martins, Dr. Lucia Maria Almeida Braz, and all of the characters that I met while living in the FMUSP moradia.

To all of the people at home who give love unconditionally and who remind me that I am strong enough to pick myself up and become stronger for having fallen: Mom and Daddy, my sister Bree, Grandmamma and Granddaddy, Granny and Papa, and my dear friend Natalia Gonzalez Varela.

Resumo

Gill DE. Uma investigação epidemiológica de arbovirais circulantes no estado brasileiro Amapá durante os surtos de 2013-2016 (dissertação). São Paulo: Instituto de Medicina Tropical de São Paulo da Universidade de São Paulo; 2019.

Introdução: As arboviroses causam graves problemas de saúde pública no Brasil e em muitos dos países da América Latina. A epidemiologia molecular é um instrumento valioso na compreensão da dispersão, persistência e diversidade desses patógenos virais. **Objetivos**: Neste projeto, buscamos investigar a dinâmica epidemiológica molecular dos arbovíroses (com especial enfoque aos vírus da dengue-DENV, chikungunya - CHIKV e zika -ZIKV) que circularam no estado do Amapá entre os anos de 2013 e 2016. **Métodos**: 824 amostras de plasma humano foram coletadas pelos laboratórios de Saúde Publica (LACEN) no estado do Amapá entre os anos de 2013 e 2016; essas amostras foram obtidas de pacientes que apresentavam sintomas consistentes com uma das arboviroses. O material genético viral presente nestas amostras foi extraído e os ensaios de qPCR foram realizados. Todas as amostras foram submetidas inicialmente a um ensaio triplex (ZIKV/DENV/CHIKV), as amostras negativas foram posteriormente submetidas a um ensaio de pan-flavivírus. As amostras positivas para um dos ensaios foram submetidas a NGS (sequenciamento de nova geração). **Resultados**: Das 824 amostras testadas, 36 foram positivas para DENV, ZIKV ou CHIKV; desses 36 positivos, 24 foram para DENV, 11 para CHIKV e 1 para ZIKV. Foram obtidos 27 genomas completos: 16 de DENV (15 DENV1, genótipo V e 1 DENV2, genótipo III) e 11 de CHIKV (genótipo asiático / caribenho). Das 788 amostras testadas com o ensaio de pan-flavivírus, 22 amostras foram positivas; porem apenas uma amostra produziu genoma completo pela técnica de NGS. Este genoma foi relacionado com um flavivírus com semelhante em 76,81% com o vírus Long Pine Key – LPKV, que anteriormente só tinha sido descrito em mosquitos. Árvores de *Maximum likelihood* e *Maximum clade credibility* foram construídas utilizando os genomas do DENV1 obtidos neste estudo. Essas árvores exibiam duas linhagens distintas de DENV1, genótipo V presentes na América Latina. Uma destas linhagens tem um padrão de circulação que inclui países do Caribe, América Central e América do Sul (incluindo Brasil); a outra linhagem distinta circula dentro das fronteiras do Brasil. As árvores também indicam que o DENV1 presente no estado do Amapá é da linhagem que tem o padrão de circulação que inclui o Caribe e as Américas Central e do Sul e que essa linhagem surgiu no Amapá entre 2005 e 2010. **Conclusão**: Este estudo fornece dados importantes sobre as arboviroses no Amapá e os dados genômicos mais recentes disponíveis para a região, bem como o contexto brasileiro e latino-americano para esses dados. Dados dessa natureza são inestimáveis nos esforços das autoridades de saúde pública para a prevenção e controle de epidemias por estes agentes.

Descritores: Epidemiologia. Virologia molecular. Arbovirus. Genômica.

Abstract

Gill DE. An Epidemiological Investigation of Circulating Arboviruses in the Brazilian state of Amapá during the Outbreaks of 2013-2016 (dissertation). São Paulo: Instituto de Medicina Tropical de São Paulo da Universidade de São Paulo; 2019.

Introduction: Arboviral febrile illnesses plague the nation of Brazil and many of its surrounding Latin America countries. Molecular epidemiology is a growing and increasingly invaluable tool in the field of public health for understanding the dispersal, persistence, and diversity of these impactful viral pathogens. **Objectives:** In this project, the identities and molecular epidemiological dynamics of arboviruses circulating in the Brazilian state of Amapá between the years 2013 and 2016, with special focus on DENV, CHIKV and ZIKV, were investigated and given Brazilian and Latin American geographical and temporal context via molecular epidemiological analyses. **Methods:** 824 human blood plasma samples were collected from LACEN laboratories in the state of Amapá between the years 2013 and 2016; these samples originated from patients showing symptoms consistent with any of the common arboviral febrile illnesses. The viral genetic material present in these samples was extracted and qPCR diagnostics assays were performed; all samples first underwent a triplex assay (ZIKV/DENV/CHIKV - ZDC), then the samples yielding negative results for the triplex assay underwent a pan-flavivirus assay. The samples yielding positive results for either assay were submitted for NGS and all whole viral genomes subsequently obtained underwent phylogenetic molecular epidemiological analyses. **Results:** Of the 824 samples tested, 36 tested positive for the ZDC assay; of those positives, 24 tested positive for DENV, 11 for CHIKV, and 1 for ZIKV. 27 full genomes were obtained from these ZDC positives: 16 of DENV (15 DENV1, genotype V and 1 DENV2, genotype III) and 11 of CHIKV (Asian and Caribbean genotype). Of the 788 samples tested with the pan-flavivirus assay, 22 samples yielded positive results, from only one of which a genome was obtainable. This genome was found to be closely related to a flavivirus previously only found in mosquitoes (76.8% identity with Long Pine Key Virus - LPKV). Maximum likelihood and maximum clade credibility trees were constructed using the DENV1 genomes obtained from this study. These trees displayed two distinct lineages of DENV1, genotype V present in Latin America, one of which has a circulation pattern spanning widely across the Caribbean and Central and South America (including Brazil), while the other circulates within Brazilian borders. The trees also indicate that the DENV1 present in the state of Amapá is of the lineage having the wider circulation pattern and that this lineage emerged in Amapá between 2005 and 2010. **Conclusion:** This study provides important data concerning the range of the arboviral landscape in Amapá and the most recent genomic data available for the region as well as Brazilian and Latin American context to that data. Data of this nature are invaluable in the efforts of public health officials for the prevention and control of epidemics of these impactful arboviral pathogens.

Descriptors: Epidemiology. Molecular virology. Arbovirus. Genomics.

List of Figures

List of Tables

List of Abbreviations

DENV – any one of the four dengue RNA viruses (DENV1, DENV2, DENV3, DENV4)
RNA – Ribonucleic Acid
ZIKV – Zika virus
CHIKV – Chikungunya virus
ECSA – East-Central-South African genotype of the Chikungunya virus
NGS – Next generation sequencing
qPCR – Quantitative polymerase chain reaction
LACEN – Laboratório Central de Saúde Pública (Central Laboratory of Public Health)
PBS – Phosphate-buffered saline
cDNA – Complementary deoxyribonucleic acid
LPKV – Long Pine Key Virus
ISF – Insect-specific virus
cISFs – Classical insect-specific flaviviruses
dISFs – Dual host affiliate insect-specific flaviviruses

Table of Contents

INTRODUCTION

1 Introduction

1.1 Background and significance

Dengue and Dengue Hemorrhagic Fever are febrile diseases caused by any one of the four dengue (DENV) RNA viruses, DENV1, DENV2, DENV3, DENV4, and are transmitted primarily by the *Aedes aegypti* mosquito. In addition to possessing considerable genetic diversity, the DENV enzyme for replication is prone to introducing RNA errors; therefore, the virus accumulates genetic changes rapidly over time[1]. Dengue is one of the most impactful tropical-climate diseases in the world, infecting more than 390 million people annually in more than 100 countries and endemic areas[2]. Since its reintroduction to Brazil in 1982, DENV has grown to be a serious public health concern; Brazil now has the highest number of notified annual infections of DENV in the Americas[3-5]. The highest numbers of reported cases were observed between 2010 and 2015; cumulatively, during these years there were more than 5 million reported cases. Furthermore, the transmission of all four serotypes of DENV has been documented in Brazil[5-6].

Like DENV, Zika *(*ZIKV) is an RNA virus of the family *Flaviviridae* that is transmitted by *Aedes* sp*.* mosquitoes (*Ae. aegypti* and *Ae. albopictus)*[7-9]. The Zika virus does display some distinctive characteristics from the rest of its relatives in the *Flaviviridae* family, however. For instance, although the replication of *flaviviruses* is known to occur in the cell cytoplasm, it has been reported that ZIKV antigens have been found in the nucleus of infected cells[10]. Another important distinction of ZIKV is that it has the potential to be transmitted sexually, from mother to unborn child, and via blood products[11-13]. The virus gained notoriety after outbreaks were reported in Yap Island in 2007, French Polynesia in

2013, and Cook Island and New Caledonia in 2014[7,13-15]. In 2015 in Brazil, cases were reported in all regions of the country; due to an increase in cases of microcephaly in infants occurring after this surge of infections, concern arose of a connection between ZIKV infection and the birth defect[16]. Despite evidence that the incidence of ZIKV in Brazil is increasing, its introduction and dissemination in the country continue to be poorly understood.

Chikungunya (CHIKV) is a member of the genus *Alphavirus*, within the family *Togaviridae*. *Alphaviruses* that are known to cause disease in humans can be divided into two primary, phylogenetically distinct groups: one group that causes arthralgia, including the Chikungunya, Semliki Forest, O'nyong nyong, Rio Ross, and Barmah Forest viruses, and another group that causes encephalitis, including the Equine Encephalitis virus and the Western and Eastern Encephalitis Equine Encephalomyelitis viruses[17,18]. After the inoculum of the mosquito, CHIKV induces an acute febrile illness, typically accompanied by arthralgia, which can last from weeks to months. The virus was discovered in Africa in the 1950's, and has since been reported to have caused numerous outbreaks around the world. Three genotypes of CHIKV have been identified: the Asian genotype, the East-Central-South African (ECSA) genotype, and the West African genotype. In 2005, a new strain of the ECSA genotype, called the *Indian Ocean Lineage*, took hold in the Indian Ocean islands and has reached into India and Southeast Asia, causing millions of cases, as well as extending into Southern Europe through hundreds of autochthonous cases[17,18]. Near the end of 2013, the Asian genotype emerged in the Caribbean and has since spread into North, South and Central America[17,18]. In Brazil, only in the year 2014 did more than 3 thousand notified cases appear across 5 federal states. In 2015, these numbers grew exponentially; more than 17,000 cases

were reported in Brazil, spanning across more than 10 states, while more than 1 million cases were reported across the Americas, with still more cases appearing due to autochthonous transmission in more than 50 countries[19]. In 2004, the ECSA genotype was introduced in the Comoros Islands from Kenya where, within months, 63% of the people were infected; this genotype was introduced to Brazil in 2014 in Feira de Santana, Bahia from Angola and it is estimated that 94% of the Brazilian population is at risk of infection[20,21].

In recent years, molecular epidemiology has become an essential tool for investigating outbreaks of new and established pathogens. Molecular epidemiology incorporates models of ecological and evolutionary processes, enabling researchers to investigate factors such as epidemic origins, transmission rates and dissemination patterns. In a study conducted by Nunes *et al*.[20] in 2015, a combination of genetic, epidemiological and ecological data was applied to describe the introduction of CHIKV to Brazil and to predict its dissemination (Figure 1); the team predicted that the dissemination patterns of the Asian and the East-Central-South African (ESCA) genotypes are likely to overlap and that up to 99% of the Brazilian population could be at risk of infection[20]. Molecular epidemiological investigation of the circulation of DENV in the Americas and in Brazil suggests repeated, isolated entrances of the virus into Brazil from other nations in the Americas as well as from other parts of the world[22]. It has also been concluded that a consistent gene flow exists among nations in the Americas and that it maintains DENV genetic diversity between outbreaks; this gene flow commenced at a steadier, slower rate after an initial explosion of DENV genetic diversity that occurred concurrently with the first introductions of the virus to the Americas[4,5].

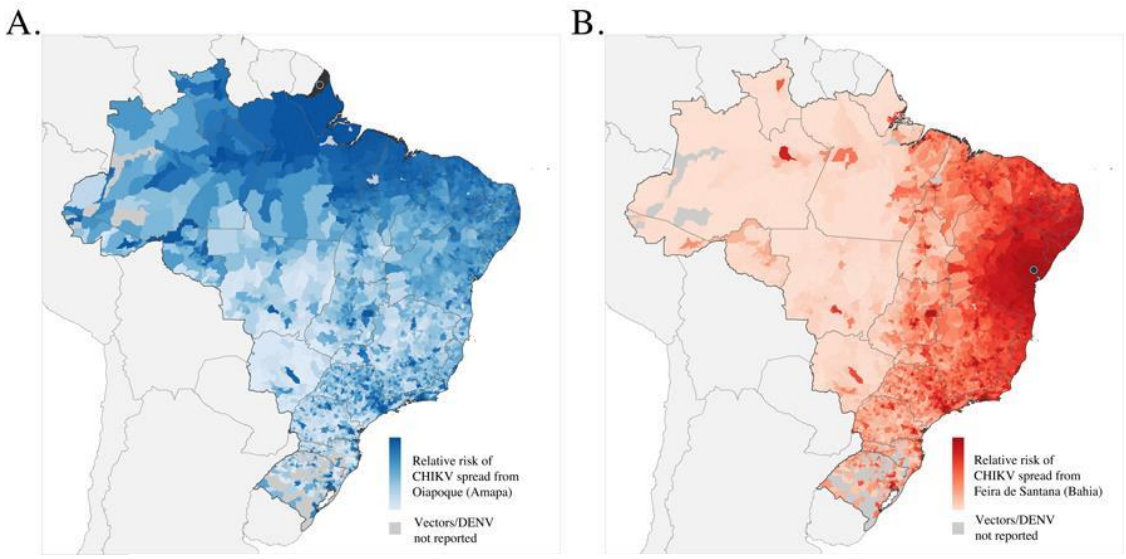

Figure 1 - Risk maps of the dissemination of the Asian CHIKV genotype from Oiapoque, in the State of Amapá (Panel A) and that of the dissemination of the ECSA CHIKV genotype from Feira de Santana, Bahia (Panel B). The Asian genotype is known to have been circulating in Caribbean since 2013, while the emergence of the ECSA genotype in Feira de Santana represents the first introduction of this genotype to the Americas[20].

In Brazil, the infrastructure of public health, housing, tourism and the economy are growing steadily; nevertheless, outbreaks of common arboviruses, such as CHIKV, DENV, and ZIKV, are highly capable of causing significant problems in the country[23]. The rapid diversification and dissemination of arboviruses such as DENV, CHIKV and ZIKV make their genetic analysis a crucial component not only of understanding epidemics but also of the development of effective preventative and control measures. The development of such strategies requires a combination of genetic, environmental, clinical and surveillance data in order to identify factors that drive the transmission of a virus and influence its spatial dissemination[24]. Investigative methods, such as the complete sequencing of viral genetic material extracted from samples collected throughout Brazil using next generation sequencing (NGS), may aid in the elucidation and prediction of transmission and spatial dynamics of these viruses in Brazil. Due to its high incidence of arboviral diseases and to

its continental dimensions, Brazil is an ideal country to collect samples for use in epidemiological studies.

1.2 Study objectives

In this project, the identities and molecular epidemiological dynamics of arboviruses circulating in the Brazilian state of Amapá between the years 2013 and 2016, with special focus on DENV, CHIKV and ZIKV, were investigated. The primary goal of this project was to provide a picture of the arboviral landscape in the state of Amapá and to give Brazilian and Latin American geographical and temporal context to that landscape using molecular epidemiological techniques.

The data collected and presented through this work represents the most recent information available concerning arboviral circulation in the state of Amapá and the most recent genomic data of DENV1 available in this region of Brazil. This information is currently lacking in published literature, and another major goal of this study is to contribute to the filling this gap of updated information available to scientists and public health officials who work towards the control and prevention of infections by these impactful arboviral pathogens.

1.3 Study design

The primary phases of this study were as follows: RNA extraction, qPCR diagnostic assays, NGS, and molecular epidemiological analyses.

Over eight hundred raw human blood plasma samples were collected from LACEN laboratories in the state of Amapá between the years 2013 and 2016; these samples originated from patients showing symptoms consistent with any of the common arboviral

febrile illnesses. The viral genetic material present in these samples was extracted and qPCR diagnostics assays were performed, followed by NGS.

The analyses of this study focused on the identification of the arboviral species and strains circulating in Amapá, and evaluation of how they epidemiologically and genetically compared with strains circulating in other regions of Brazil and Latin America through the application of molecular epidemiological techniques.

Extraction and qPCR lab work was conducted in laboratories of the Adolfo Lutz Institute of São Paulo, the reference public health laboratory of the state. Sequencing was performed at the Institute of Tropical Medicine of São Paulo. Genome assembly was also performed at the Institute of Tropical Medicine of São Paulo, with the assistance of the Blood Systems Research Institute, San Francisco, CA, USA[25]. The phylogenetic and epidemiological analyses were conducted at the Institute of Tropical Medicine of São Paulo, with aid from the Adolfo Lutz Institute.

MATERIALS AND METHODS

2 Materials and methods

This study is part of the project approved by the CEP FMUSP under the number CAAE: 53153916.7.0000.0065.

2.1 Sample processing and preliminary data acquisition

Through the years 2013-2016, blood plasma samples were obtained from the LACEN in Amapá; 800 samples were taken in 2016, 800 in 2015, 800 in 2014, and 800 in 2013. Of those original samples, 96, 240, 283, and 205 samples were used for this study from the years 2013, 2014, 2015, and 2016, respectively. The numbers of samples from each year were chosen arbitrarily and based on accessibility.

Viral RNA was extracted from the samples using a MagNa Pure 2.0 Roche automatic nucleic acid extraction machine (MagNA Pure LC instrument, Roche Applied Science, Indianapolis, Ind.). The reagent kits used for extraction were from the MagNa Pure LC Total Nucleic Acid Isolation Kit- High Performance, Version 8, by Roche (Roche Applied Science, Indianapolis, Ind.); the protocol used was that specified by the kit instructions. 200uL of blood plasma was used from each sample for extraction; if the sample did not have a full 200uL of volume, PBS was added to the sample up to a total volume of 200uL and the contents of the sample tube were agitated gently with a pipette. The final elution volume for each sample was 60uL. After extraction, the samples were stored in a -80ºC degree freezer.

The samples were then submitted to a series of qPCR assays, as is described in Figure 2. First, the ZDC (Zika, Dengue, Chikungunya) Multiplex qPCR Assay, by BIORAD (Bio-Rad Laboratories, Inc.; Hercules, California), was applied to all samples. The assay

was performed according to the manufacturer's protocol, which is specified in the kit; 5uL of extracted RNA was used for the assay. The samples that showed positive results for the ZDC assay were submitted to NGS.

The samples that showed negative results for the ZDC assay were then submitted to a pan-Flavivirus multiplex qPCR assay, using the primers and protocol described in the article, *Development of one-step quantitative reverse transcription PCR for the rapid detection of flaviviruses*, by Patel *et al*.[26]. Again, 5uL from each sample of extracted RNA was used for the assay.

The samples that showed negative results for the pan-Flavivirus assay were sent to a partner lab for the application of a viral metagenomics protocol for the complete nucleic acid characterization of these samples. Those which yielded positive results for the pan-Flavivirus assay were sent for NGS.

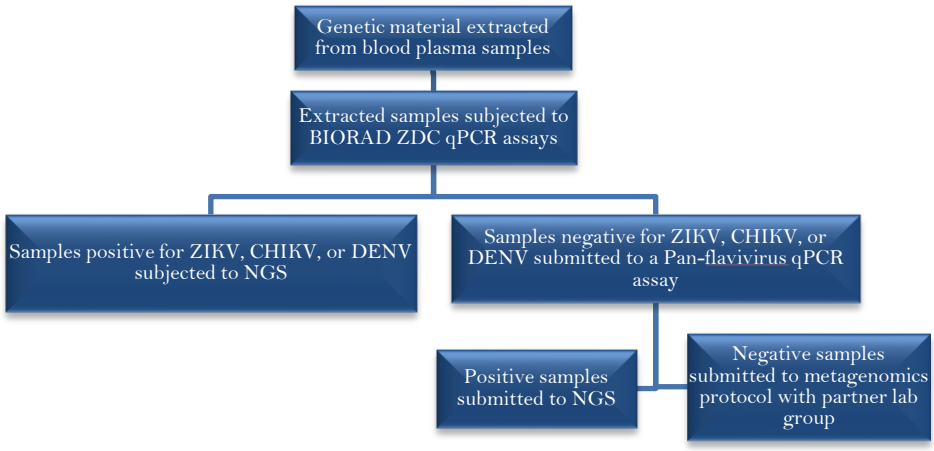

Figure 2 - Scheme of sample processing procedures and lab work

The afore mentioned viral metagenomics protocol, which has been performed on the samples that showed negative results for both the ZDC and the pan-Flavivirus assays, has been conducted by the lab of Dr. Eric Delwart at the Blood Systems Research Institute in San Francisco, California. According to this protocol, 0.3mL of the sample was

centrifuged at 12,000 x g for 5 minutes at 8 ºC and the supernatant was filtered through a filter of 0.45M (Millipore, Billerica, MA, USA) to remove host and bacterial cellular debris. The filtrate was then treated with a mixture of nucleases at 37 ºC for 1.5 hours to reduce nucleic acids, keeping only the infectious viral nucleic acids, which are protected from digestion by their capsids. Total nucleic acid extraction was then performed using the ZR & ZR-96 DNA/RNA Kit (Zymo Research, Irvine, CA, USA). According to the manufacturer's instructions, the elution volume was 0.05mL of nuclease free water. cDNA synthesis was performed with the SuperScript III kit (Life Technologies, Grand Island, NY, USA) and the second strand of cDNA with the Klenow FRAGMENT kit (New England Biolabs, Ipswich, MA, USA). The resulting product was then submitted directly to Nextera XT (Illumina, San Diego, CA, USA) for library preparation. The paired-end, 300pb sequences generated by MiSeq were demultiplexed using Illumina software. The data was then submitted to the "virus discovery" pipeline in the supercomputers of the Blood Systems Research Institute[25]. The sequences were filtered to remove remaining human, bacterial, and fungal sequences using bowtie2. Later, assemblers were used for the reconstruction of viral genomes, including SOAPdenovo2, Abyss, meta-Velvet and CAP3, Mira and SPADES programs. The contigs were then submitted to TBLASTX. The full or partial genomes were then evaluated in terms of cover using *Geneious R8* (Biomatters, San Francisco, CA, USA). In samples for which it may not be possible to assemble the complete genome, specific primers are used for amplification and subsequent sequencing.

2.2 Sequence analysis and model development

The newly obtained DENV1 genomes were submitted for phylogenetic analysis with previously published genomes collected throughout Latin America (including genomes

from the Caribbean and South America) in order to explore their molecular epidemiological context on the continental scale. This first tree contains genomes from all 5 DENV1 genotypes. For this tree, complete genomic sequences were aligned using the SeaView[27] software package component, Muscle[28], and subsequently visually inspected and manually adjusted. Phylogenetic reconstructions were then produced using Maximum Likelihood methods in PhyML (NNI algorithm), also found within the SeaView[27] program. The substitution model used was GTR+I+G, as suggested by jModelTest[29].

In order to more closely explore the Brazilian national context of the newly obtained DENV1 genomes, phylogenetic analysis was performed using these new genomes with a set of previously published DENV1 genomes collected throughout Brazil. For this tree, the alignment containing all sequences from the Latin American tree was edited in Seaview to exclude all non-Brazilian sequences. Phylogenetic reconstructions were then performed using Bayesian settings in Beast package software[30]. Settings in Beauti were as follows: substitution model [GTR], base frequencies [Estimated], site heterogeneity model [Gamma + Invariant Sites], number of gamma categories [4], clock type [uncorrelated relaxed clock], relaxed distribution [lognormal], tree prior [coalescent: Bayesian Skyline], number of groups [10], skyline model [piecewise-constant], tree model [random starting tree], length of chain [100000], echo state to screen every [1000], log parameters every [1000]. A maximum clade credibility tree (MCC tree) was then generated using Tree Annotator in the Beast package.

FigTree software was used for visualization and editing of both Maximum likelihood and Bayesian trees. All previously published reference sequences used for both trees were obtained from GenBank. Genotyping of the newly obtained genomes was performed using the online Genome Detective Virus Tool[31].

RESULTS

3 Results

3.1 qPCR assays

Figure 3 below shows the collection year and virus distribution of the positive results of the ZDC assay. Of the 824 samples tested, 788 were negative and 36 samples (4.3%) were positive (0 from 2013, 8 from 2014, 23 from 2015, and 5 from 2016). Of the 36 positive samples, 24 were DENV (5, from 2014; 16, from 2015; 3, from 2016), 11 were CHIKV (3, from 2014; 7, from 2015; 1, from 2016), and 1 was ZIKV (from 2016).

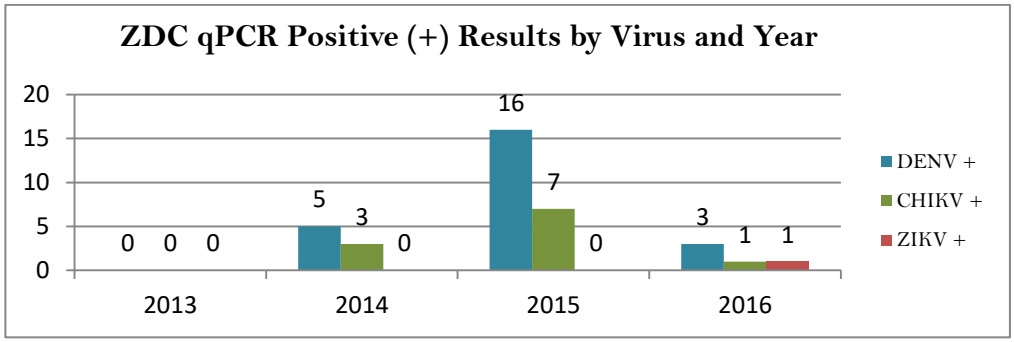

Figure 3 – Collection year and virus distribution of the positive results of the ZDC assay

Figure 4 below shows the distribution of the Pan-flavivirus qPCR assay positives, with respect to collection year. Of the 788 samples that underwent this assay, 22 showed positive results (0 from 2013, 5 from 2014, 14 from 2015, and 3 from 2016).

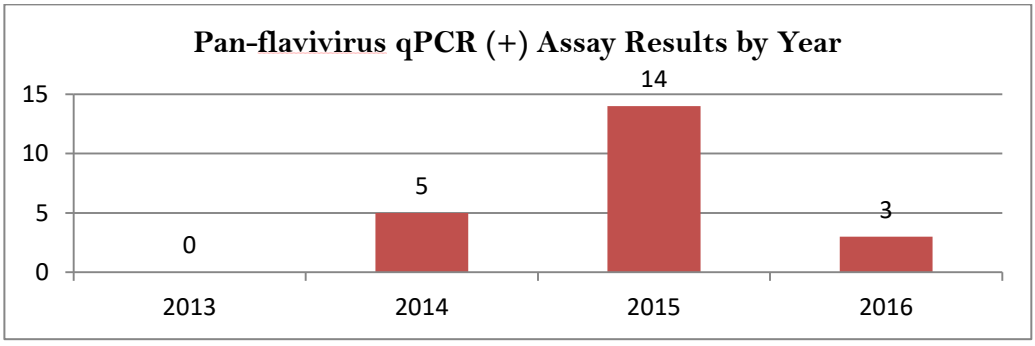

Figure 4 – Distribution of the Pan-flavivirus qPCR assay positives, with respect to collection year

Figure 5 below describes the Ct values for all of the positive samples of the ZDC assay (series in blue) and the pan-Flavivirus assay (series in green). The red diamonds represent samples that yielded full genomes after NGS, 28 in total (not including any that may be obtained from the viral metagenomic investigation of the negative samples, as those results are still pending). Of the 36 ZDC assay positives, 27 resulted in full genomes, 11 of which were CHIKV and 16 were DENV (15 DENV1 full genomes and 1 DENV2 full genome). The genotypes of these genomes are as follows: all 15 DENV1 genomes were determined to be of genotype V, the single DENV2 genome was determined to be of genotype III (Southern Asian/American), and all 11 CHIKV genomes were determined to be of the Asian and Caribbean genotype. The single full genome produced from the 22 Pan-flavivirus assay positives is that of a Flavivirus that is highly related to a recently discovered insect-specific Flavivirus called Long Pine Key Virus (LPKV), which was collected from mosquitoes captured in the Florida keys[32].

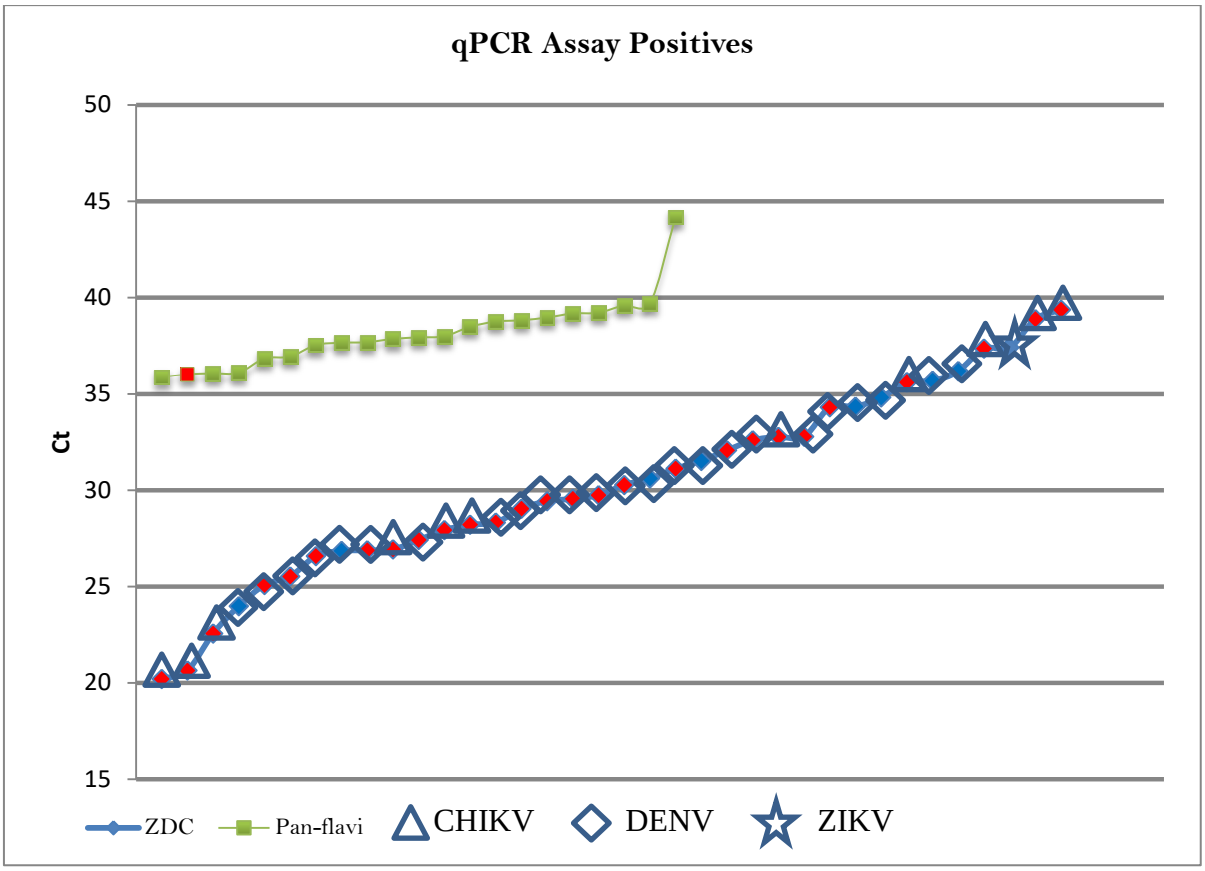

Figure 5 – Ct's of all positives of the ZDC and Pan-flavi assays; shown in red are the samples that yielded full genomes after NGS; for the ZDC assay, the samples' virus identities specified by the assay results are distinguished by the designated shapes

Figure 6 below displays and compares the distributions of the Ct's of the positives of both the ZDC assay and the pan-Flavivirus assay. The box plot shown in red represents the distribution summary of the Ct values of the ZDC assay positives; the average Ct value was 30.35, the range was 12.49, and the minimum, quartile 1, median, quartile 3, and maximum values were 26.88, 26.85, 30.01, 34.34, 39.37, respectively. The box plot shown in blue represents the distribution summary of the Ct values of the pan-Flavivirus assay positives; the average Ct value was 38.16, the range was 8.31, and the

minimum, quartile 1, median, quartile 3, and maximum vales were 35.87, 36.88, 37.93, 39.07, 44.18, respectively.

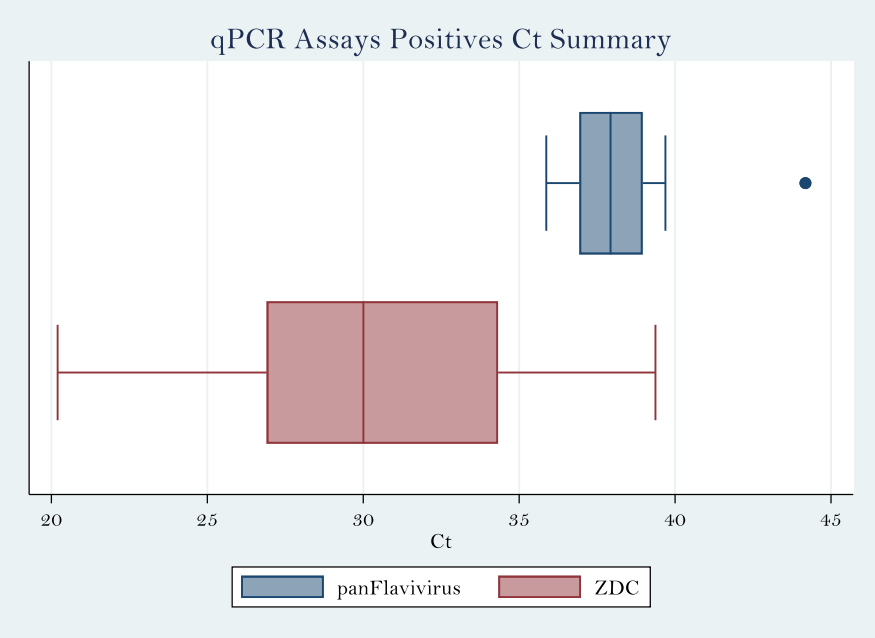

Figure 6 – Summary of positive Ct values of the ZDC and Pan-flavivirus assays

3.2 Phylogenetic reconstructions containing newly obtained DENV1 genomes

Figure 7 below displays a maximum likelihood tree containing the newly obtained DENV1 genomes collected in Amapá with other DENV1 genomes collected throughout Brazil and Latin America. Brazilian genomes are indicated in red. This tree contains reference sequences for all 5 genotypes of DENV1. As can be seen in Figure 7, genotypes 1-4 of DENV1 (highlighted in yellow in the tree) grouped outside of the two primary visible clusters, indicating that the two clusters are two distinct lineages within genotype 5 of DENV1 (labeled as L1 and L2 in the tree). Lineage one, L1, shows clustering of Brazilian sequences with those from several other countries in Latin America including Venezuela, Colombia, Argentina, Puerto Rico, Nicaragua, Mexico, and Haiti. L1

also contains all of the newly obtained genomes collected in Amapá; they are all clustered at the most terminal positions of the tree. Lineage 2, L2, contains mostly genomes collected within Brazil, with the exception of a couple of genomes collected from small Caribbean islands, including the British Virgin Islands. It should be noted that both lineages, L1 and L2, circulate in similar regions of Brazil, namely Sao Paulo and Pernambuco states, during the same time period.

Figure 8 below displays a tree made of the newly obtained DENV1 genomes collected in Amapá state with other DENV1 genomes collected throughout Brazil. There are two major lineages displayed in this tree, represented by two distinct branches, the nodes of which are indicated in Figure 8 by L1 and L2. The tree shows that between the years 1990 and 2015, there were two major, distinct lineages of DENV1 genotype 5 circulating in Brazil.

The most terminal branch, labeled L1, contains all of the Amapá samples from this study as well as genomes originating in Pernambuco and Sao Paulo states. The Amapá genomes rest in the most terminal positions of the branch. This branch of the tree shows that strains of DENV1 circulating in Amapá during the years 2014 and 2015 are closely related to those circulating in Sao Jose de Rio Preto, in Sao Paulo state, and Recife, in Pernambuco state in 2010 and 2011. These locations represent three very distant corners of the country, indicating that this particular strain is widely spread across the country.

The other major branch visible in this tree, labeled L2, represents another distinct lineage of DENV1 and contains genomes originating in Pernambuco and Sao Paulo states as well as other unspecified regions throughout Brazil. This branch shows that strains of DENV1 circulating in Pernambuco in 2010 are closely related to those circulating in Sao

Paulo state in the years 2011-2014. Again, these locations represent distant corners of the country, indicating that this strain must too be widespread throughout the country.

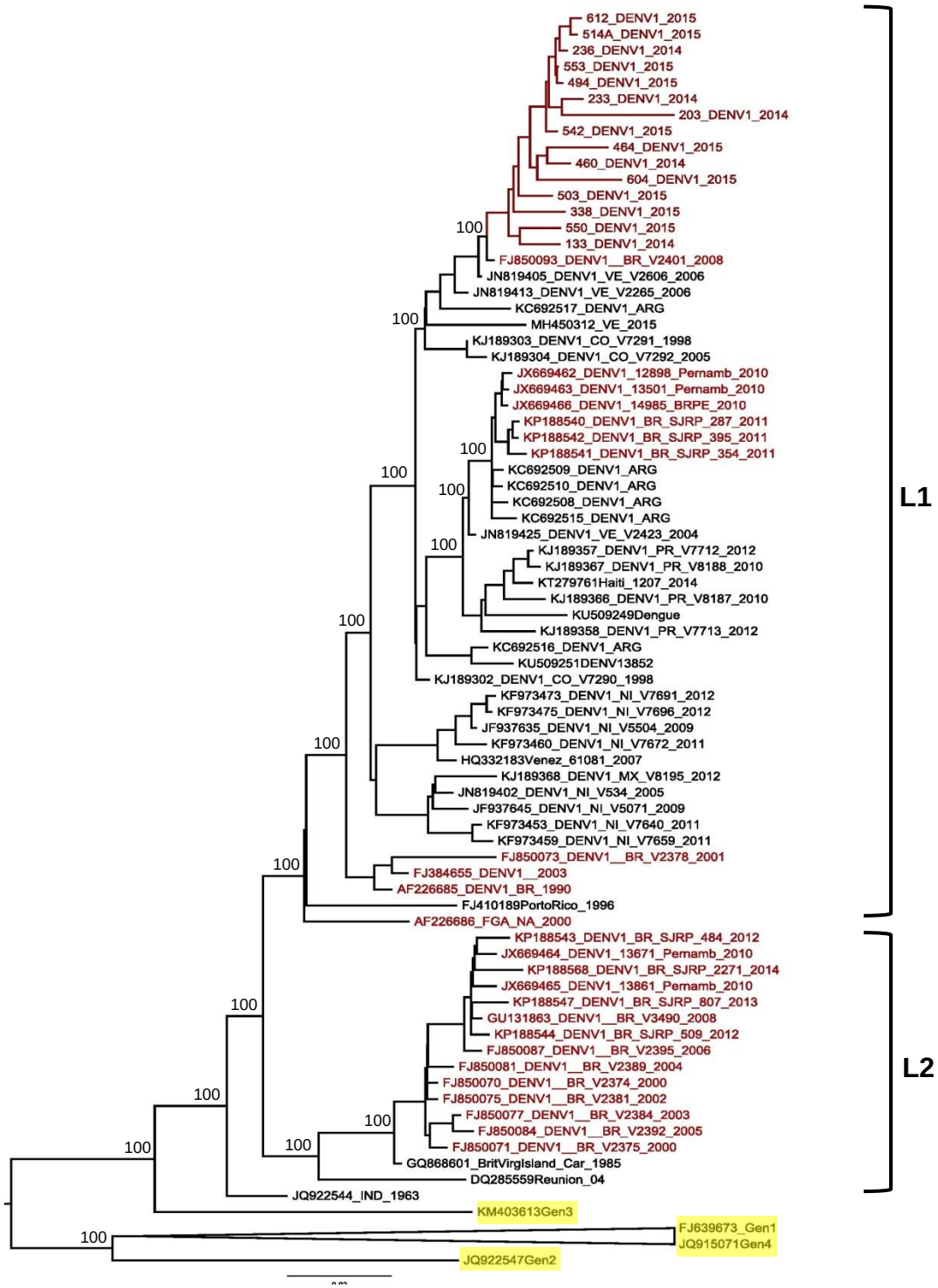

Figure 7 – Phylogenetic reconstruction of newly obtained DENV1 genomes from Amapá with other DENV1 genomes obtained throughout Brazil and Latin America; Brazilian genomes indicated in red; two distinct lineages of DENV1 genotype 5 indicated by L1 and L2; main nodes having bootstrap values of 70 or above are labeled

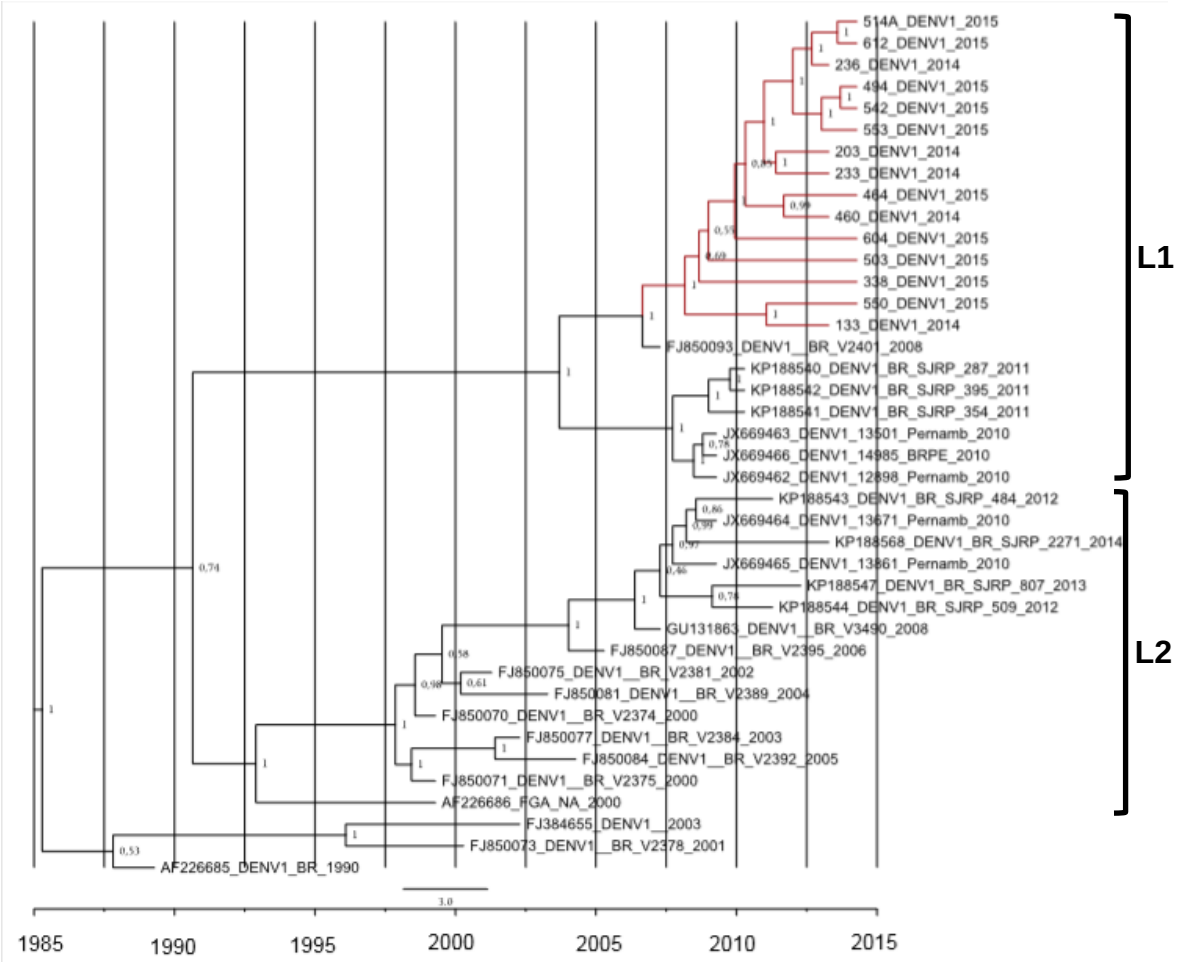

Figure 8 – MCC Phylogenetic reconstruction of newly obtained DENV1 genomes from Amapá with other DENV1 genomes obtained throughout Brazil over time; two distinct lineages of DENV1 genotype 5 indicated by L1 and L2; posterior probability values are displayed at the nodes

3.3 BLAST investigation of newly obtained LPKV-related Flavivirus

Figure 9 below displays a partial view of the alignment between the newly obtained LPKV-related flavivirus and the first hit resulting from a BLAST search of the new genome. This first hit has a query cover of 97%, an e value of 0.0 and a percent identity of 76.81%.

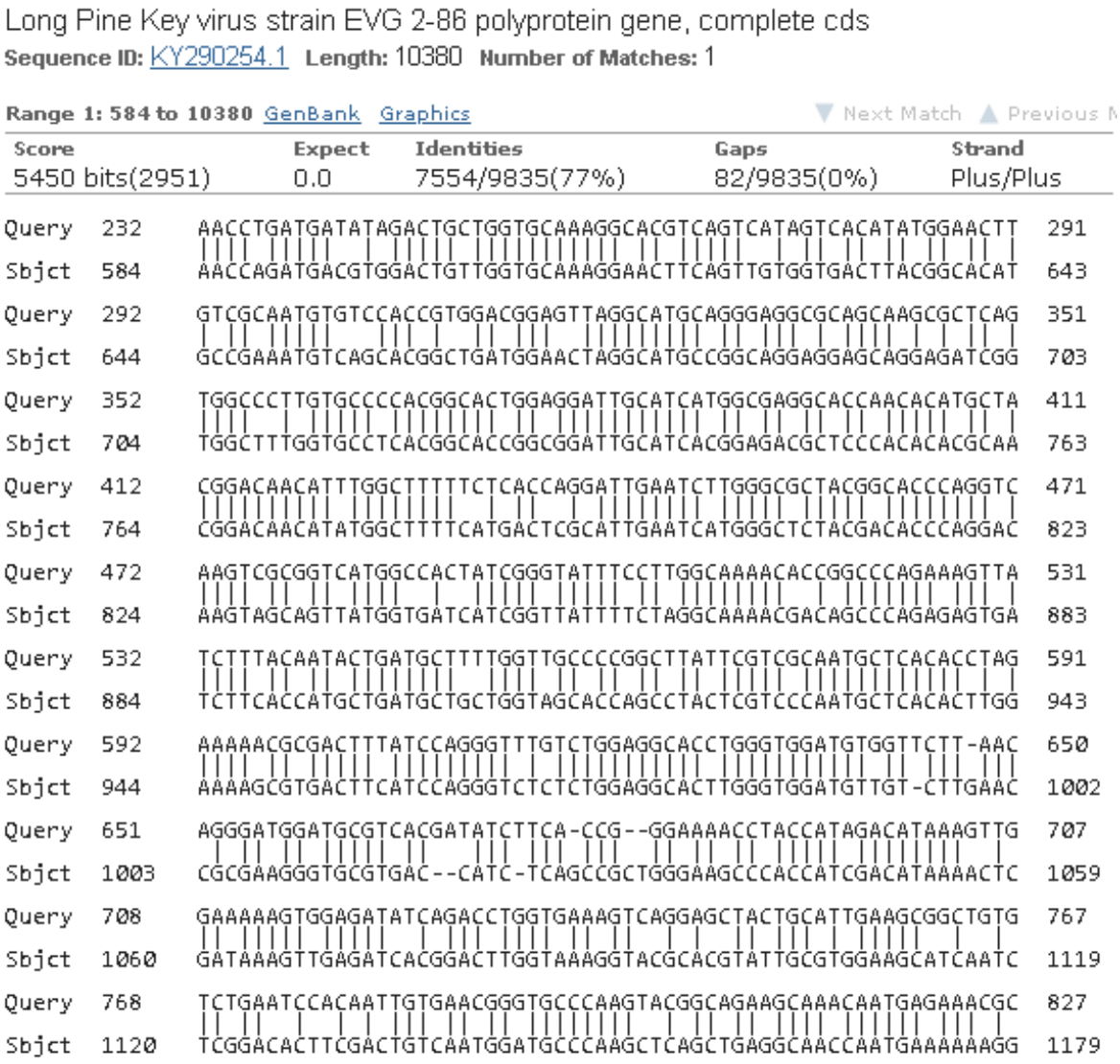

Figure 9 – Partial view of alignment between newly obtained LPKV-like flavivirus with first hit resulting from a BLAST search

Table 1 below describes the first five results of a BLAST search using the new LKPV-like flavivirus genome. The query covers of the first five hits were all 97% and the e values were all 0.0. The percent identities were 76.81%, 76.79%, 76.79%, 76.78% and 76.75%, respectively.

Table 1 – Description of the first five results from a BLAST search of newly obtained LPKV-like flavivirus

Description	Query Cover	E value	Percent Identity	Accession
Long Pine Key virus strain EVG 2-86 polyprotein gene, complete cds	97%	0.0	76.81%	KY290254.1
Long Pine Key virus strain EVG 5-61 polyprotein gene, complete cds	97%	0.0	76.79%	KY290255.1
Long Pine Key virus strain EVG 1-33 polyprotein gene, complete cds	97%	0.0	76.79%	KY290249.1
Long Pine Key virus strain EVG 5-72 polyprotein gene, complete cds	97%	0.0	76.78%	KY290256.1
Long Pine Key virus strain EVG 2-81 polyprotein gene, complete cds	97%	0.0	76.75%	KY290253.1

DISCUSSION

4 Discussion

4.1 Implications of diagnostics assays

Among the arboviruses to be positively identified as circulating in the state of Amapá during the years 2013-2015 are: the Asian and Caribbean genotype of CHIKV, genotype V of DENV1, genotype III of DENV2, and one unspecified strain of ZIKV. Another 22 strains of unspecified Flavivirus were detected through this study, although only one produced a full genome, allowing us to further characterize it as a previously undescribed virus similar to LPKV.

Not only is it important to understand what viruses are circulating in an area in order to construct effective plans of action for their control, but viral circulation in Amapá state, specifically, requires special attention. Amapá has been shown to be an entry point for the Asian genotype of CHIKV[20], which subsequently spread throughout the country, indicating that it may have the ability to do the same for other viral pathogens. Additionally, the climate and road coverage of Amapá state may lend it specific ability for the perpetuation and dissemination of these viruses. Amapá has the highest concentrations of roads compared with other regions in northern Brazil, many of which are accessible by boat or plane, and connects the Northern region to other regions of Brazil. For these reasons, it is especially necessary to understand the viral landscape of Amapá state, information which is already lacking in published literature, adding to the significance of this study.

4.2 Long Pine Key Virus-related Flavivirus

Long Pine Key Virus (LPKV) has been reported upon in one paper only, having been revealed as a newly discovered insect-specific virus (ISF) of the family

Flaviviridae, genus *Flavivirus*[32]. ISFs refer generally to viruses that "naturally infect hematophagous Diptera and that replicate in mosquito cells in vitro, but do not replicate in vertebrate cells or infect humans or other vertebrates"[32]. The paper describes two classifications of ISV's: classical insect-specific flaviviruses (cISFs) and arbovirus-related or dual host affiliate insect-specific flaviviruses (dISFs). LPKV is described as a dISF, which are closely related, phylogenetically and antigenically, to flavivirus vertebrate pathogens.

As reported by the study which first published on LPKV[32], the samples that yielded LPKV genomes (8 total) were collected, as whole mosquitoes, between June and July of 2013 in the Everglades National Park of southern Florida, specifically from Long Pine Key. The reported host mosquito species included *Aedes sp.*, *Culex sp*., and *Anopheles sp*. mosquitoes[32]. Although the researchers who discovered the virus emphasize the close relatedness of the dISFs with vertebrate pathogenic flaviviruses, they did run tests to determine if LPKV replicates in vertebrate cells, and these tests seemed to indicate that it does not.

However, the authors speculate concerning the potential for dISFs to develop the ability to infect vertebrates. They go on to reference evidence from recent metagenomics studies that suggest that flavivirus pathogens may have evolved from earlier arthropod viruses. Evidence seems to support the possibility that dISFs have the potential to evolve the ability to act as vertebrate pathogens[33,34]. The discovery of LPKV in vertebrate cells could provide evidence that they may indeed develop this ability. Additionally, there is some evidence that dISFs can reduce vector competence for various pathogenic flaviviruses due to heterologous interference; for this reason, it has been proposed that these viruses could be used as a means of arboviral disease control. Consequently, is important that our team investigate the relatedness of this

virus with the LPKV, as the resulting findings could potentially have impacts on our understanding of how arboviral diseases evolve and develop pathogenicity, as well as the potential utility of dISFs as means for arboviral control.

4.3 Implications of phylogenetic analyses of newly obtained DENV1 genomes

The tree presented in Figure 7 of this work demonstrates 2 distinct lineages of DENV1 genotype 5, one which circulates widely among countries in the Caribbean, Central America and South America, and another that circulates primarily within Brazil. Upon further study of factors such as level of viremia in hosts and infectivity in vectors, it may be possible to determine why one lineage was able to spread so widely across countries, while the other circulated in a much tighter geographical circumference. As mentioned, both lineages circulate within similar regions in Brazil simultaneously. This finding is of interest, as the two lineages must compete for hosts within these regions, as immune responses would not allow for subsequent infections by the same genotype. This tree may represent a snapshot of clade replacement in these regions, in which a clade having better fitness would replace a less fit clade over time in a pool of hosts[35]. This is a recognized phenomenon suggested to allow for the cycles of DENV outbreaks in many endemic countries[35].

The tree presented in Figure 8 of this study characterizing DENV1 circulation in Brazil between the years 1990 and 2015 highlights the presence of two major lineages of DENV1 circulating in Brazil, also shown in Figure 7, both of which are spread widely across the country. This study also shows that the most recently diverged lineage of DENV1 has been circulating in Amapá up to the year 2015, which is the most recent DENV1 circulation data in the region and in the entire country of Brazil. Additionally, according to this tree, the emergence of this lineage into Amapá occurred sometime

between 2005 and 2010. However, besides the genomic data resented in this study, there is no other DENV1 genomic data from Amapá that would allow for better exploration of the evolution and dispersion of this serotype in Amapá through time.

It is important to understand which serotypes of DENV are circulating in which regions; such knowledge not only helps with general DENV prevention and control efforts, but also identifies regions at risk for the severest forms of illness caused by dengue viruses. Infection by dengue viruses causes a wide range of clinical presentations, from asymptomatic infection to lethal hemorrhagic fever. Specifically, the most aggressive forms of illness caused by dengue viruses are the result of a phenomenon called Antibody-dependent enhancement[36], in which subsequent infections by differing serotypes of dengue cause more severe forms of illness than the first infection. Therefore, identifying which regions harbor which serotype/s is highly important for knowing areas at risk for severe infections. This study contributes to the understanding of DENV1 circulation in Brazil, knowledge that aids in the development of more effective prevention and control efforts.

4.4 Study limitations

Limitations of this study include the less-than-optimal conditions of the samples at the initiation of the study. The samples were 2-5 years old at the start of the study, with no guarantee of proper storage in the meantime. Having all come from the LACEN laboratory, it can be assumed that most or all of the samples were collected out of the viremia period, and were sent to LACEN for the performance of serological tests. It is also possible that the samples were stored in conditions that were not conductive of long-term preservation of the viral RNA contained within them, namely storage in -20ºC

freezers; therefore, it can be expected that the samples may not yield a high percentage of positives from the qPCR assays that were performed.

Additionally, the lack of DENV1 genomic data that is more recent than that of the present study stunts the potential for solid conclusions about the role of Amapá in the distribution of arboviruses in Brazil to be made from the present study. Because the Amapá sequences are the most recent of the tree, it is expected that they would be the most terminal; therefore, any patterns indicating direction of movement of DENV1 to or from Amapá that may be suggested by the tree are entirely preliminary and unverifiable until there is more recent DENV1 genomic data available from Latin America and Brazil. However, the genomic and phylogenetic data presented in this study nevertheless carry heavy significance, as they represent the most recent DENV1 data in Brazil and Latin America and demonstrate the need for more recent data.

4.5 Final thoughts and future work

There are a few important tangents to this study that will later contribute to the overall picture of the data presented. Respective patient information, which is currently partially available, may add significantly to the utility of the genomic data. Additionally, the metagenomic data of the samples that showed negative results for both of the ZDC and Pan-flavivirus assays is still pending, but should prove to be highly valuable to the understanding of the causes of febrile diseases in the area. This data, in addition to further investigation of the LPKV-like virus obtained during the course of this study, will provide important information on little-known and previously unidentified arboviruses. Also pending are the phylogenetic analyses of the CHIKV and DENV2 genomes obtained in this study. Collectively, the pending data and analyses tangent to this study

will add to the overall understanding of the circulation and characteristics of arboviruses in Brazil and Latin America.

CONCLUSION

5 Conclusion

This study provides the most recent DENV1 genomic data available in Brazil as well as molecular epidemiological examination of these Amapá strains with those circulating in other areas of Brazil. This work also provides characterizations of the arboviral landscape of the state of Amapá, including arboviral pathogens previously identified and unidentified in academic literature.

The most recent epidemics of arboviral pathogens, such as DENV, CHIKV and ZIKV, imposed significant and direct impacts on the infrastructure of public health as well as indirect effects on tourism and the economy. This study provides a better understanding of the dissemination and persistence of increasing arboviral threats to public health in Brazil, which is crucial for successful epidemic management and control.

References

1. Jenkins GM, Rambaut A, Pybus OG, Holmes EC. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J Mol Evol. 2002;54:156-65.
2. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, *et al.* The global distribution and burden of dengue. Nature. 2013;496:504-7.
3. Laughlin CA, Morens DM, Cassetti MC, Costero-Saint Denis A, San Martin JL, Whitehead SS, *et al.* Dengue research opportunities in the Americas. J Infect Dis. 2012;206:1121-27.
4. 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;29:1533-43.
5. Carrington CV, 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;79:14680-7.
6. Nunes MRT, Palacios G, Rodrigues Faria N, Sousa EC, 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;8:2769.
7. Lanciotti RS, Kosoy OL, Laven JJ, Velez JO, Lambert AJ, Johnson AJ, *et al*. Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. Emerg Infect Dis. 2008;14:1232-9.
8. McCrae AW, Kirya BG. Yellow fever and Zika virus epizootics and enzootics in Uganda. Trans R Soc Trop Med Hyg. 1982;76:552-62.

9. Kuno G, Chang GJ, Tsuchiya KR, Karabatsos N, Cropp CB. Phylogeny of the genus Flavivirus. J Virol. 1998;72:73-83.

10. Buckley A, Gould EA. Detection of virus-specific antigen in the nuclei or nucleoli of cells infected with Zika or Langat virus. J Gen Virol. 1988;69:1913-20.

11. Musso D, Roche C, Robin E, Nhan T, Teissier A, Cao-Lormeau VM. Potential sexual transmission of Zika virus. Emerg Infect Dis. 2015;21:359-61.

12. Besnard M. Lastere S, Teissier A, Cao-Lormeau V, Musso D. Evidence of perinatal transmission of Zika virus, French Polynesia, December 2013 and February 2014. Euro Surveill. 2014;19:20751.

13. Musso D, Nhan T, Robin E, Roche C, Bierlaire D, Zisou K, *et al*. Potential for Zika virus transmission through blood transfusion demonstrated during an outbreak in French Polynesia, November 2013 to February 2014. Euro Surveill. 2014;19:20761.

14. Cao-Lormeau VM, Roche C, Teissier A, Robin E, Berry AL, Mallet HP, *et al*. Zika virus, French Polynesia, South Pacific, 2013. Emerg Infect Dis. 2014;20:1085-6.

15. Duffy MR, Chen T, Hancock WT, Powers AM, Kool JL, Lanciotti RS, *et al*. Zika virus outbreak on Yap Island, Federated States of Micronesia. N Engl J Med. 2009;360:2536-43.

16. Brady OJ, Osgood-Zimmerman A, Kassebaum NJ, Ray SE, de Araujo VEM, da Nobrega AA, *et al*. The association between Zika virus infection and microcephaly in Brazil 2015-2017: An observational analysis of over 4 million births. PLoS Med. 2019;16:1002755.

17. Horwood PF, Buchy P. Chikungunya. Rev Sci Tech. 2015;34:479-89.

18. Weaver SC, Forrester NL. Chikungunya: Evolutionary history and recent epidemic spread. Antiviral Res. 2015;120:32-9.

19. The Pan American Health Organization (PAHO). Data, Maps and Statistics. Washington, D.C.: 2015.

20. Nunes MR, Faria NR, de Vasconcelos JM, Golding N, Kraemer MU, de Oliveira LF, *et al*. Emergence and potential for spread of Chikungunya virus in Brazil. BMC Med. 2015;13:102.

21. Sergon K, Yahaya AA, Brown J, Bedja SA, Mlindasse M, Agata N, *et al*. Seroprevalence of Chikungunya virus infection on Grande Comore Island, union of the Comoros, 2005. Am J Trop Med Hyg. 2007;76:1189-93.

22. Nunes MRT, Faria NR, Vasconcelos HB, Medeiros DBA, Lima CPS, Carvalho VL, *et al*. Phylogeography of dengue virus serotype 4, Brazil, 2010-2011. Emerg Infect Dis. 2012;18:1858-64.

23. Teixeira MG, Siqueira JB, Ferreira GLC, Bricks L, Joint G. Epidemiological trends of dengue disease in Brazil (2000-2010): a systematic literature search and analysis. PLoS Negl Trop Dis. 2013;7:2520.

24. Heesterbeek H, Anderson RM, Andreasen V, Bansal S, De Angelis D, Dye C, *et al*. Modeling infectious disease dynamics in the complex landscape of global health. Science. 2015;347:4339.

25. Deng X, Naccache SN, Ng T, Federman S, Li L, Chiu CY, *et al*. An ensemble strategy that significantly improves of new assembly of microbial genomes from metagenomic next-generation sequencing data. Nucleic Acids Res. 2015;43:46.

26. Patel P, Landt O, Kaiser M, Faye O, Koppe T, Lass U, *et al*. Development of one-step quantitative reverse transcription PCR for the rapid detection of flaviviruses. Virol J. 2013;10:58.

27. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010;27:221-4.

28. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792-7.

29. Posada D. jModelTest: phylogenetic model averaging. Mol Biol Evol. 2008;25:1253-6.

30. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012;29:1969-73.

31. Emweb. Genome Detective Virus Tool [internet]. 2017. [30/4/2019] Disponível em: https://www.genomedetective.com/

32. Guzman H, Contreras-Gutierrez MA, Travassos da Rosa APA, Nunes MRT, Cardoso JF, Popov VL, *et al*. Characterization of three new insect-specific flaviviruses: Their relationship to the mosquito-borne flavivirus pathogens. Am J Trop Med Hyg. 2018;98:410–419.

33. Shi M, Lin X, Vasilakis N, Tian J, Li C, Chen L, *et al*. Divergent viruses discovered in arthropods and vertebrates revise the evolutionary history of the Flaviviridae and related viruses. J Virol. 2016;90:659–69.

34. Li CX, Shi M, Tian JH, Lin XD, Kang YJ, Chen LJ, *et al*. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. Elife. 2015;4.

35. Teoh BT, Sam SS, Tan KK, Johari J, Shu MH, Danlami MB, *et al*. Dengue virus type 1 clade replacement in recurring homotypic outbreaks. BMC Evol Biol. 2013;13:213.

36. Katzelnick LC, Gresh L, Halloran ME, Mercado JC, Kuan G, Gordan A, *et al*. Antibody-dependent enhancement of severe dengue disease in humans. Science, 2017;358:929-32.