SHAHAB ZAKI POUR

Desenvolvimento e validação experimental de uma metodologia *in house* para amplificação e sequenciamento do genoma completo do Zika vírus

Dissertação apresentada ao Departamento de Microbiologia do Instituto de Ciências Biomédicas da Universidadede São Paulo, para obtenção do Título de Mestre em Ciências.

São Paulo
2018

SHAHAB ZAKI POUR

Development and validation of an in house method for whole genome amplification and sequencing of Zika virus

Dissertation presented to the Department of Microbiology of the Institute of Biomedical Sciences of the University of São Paulo, to obtain the degree of Master in Science.

São Paulo
2018

SHAHAB ZAKI POUR

Development and validation of an in house method for whole genome amplification and sequencing of Zika virus

Dissertation presented to the Department of Microbiology of the Institute of Biomedical Sciences of the University of São Paulo, to obtain the degree of Master in Science.

Concentration area: Microbiology

Advisor: Prof. Dr. Paolo Marinho de Andrade Zanotto

Corrected Version

São Paulo
2018

SHAHAB ZAKI POUR

Desenvolvimento e validação experimental de uma metodologia *in house* para amplificação e sequenciamento do genoma completo do Zika vírus

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

Área de concentração: Microbiologia

Orientador: Prof. Dr. Paolo Marinho de Andrade Zanotto

Versão Corrigida

São Paulo
2018

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

_____

Candidato: Shahab Zaki Pour

Título da Dissertação: **Desenvolvimento e validação experimental de uma metodologia *in-house* para amplificação e sequenciamento do genoma completo do Zika vírus**

Orientador: Prof. Dr. Paolo Marinho de Andrade Zanotto

 A Comissão Julgadora dos trabalhos de Defesa da Dissertação de Mestrado, em sessão pública realizada a ..../................./................, considerou

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

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

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

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

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

·ICB USP

São Paulo, 21 de setembro de 2017.

PARECER 1354/CEPSH
CAAE nº 75975517.8.0000.5467
Protocolo CEPSH-ICB nº 1398/17

O Comitê de Ética em Pesquisas em Seres Humanos do ICB, nesta data, **APROVOU** o projeto intitulado: "*Desenvolvimento e validação experimental de uma metodologia in-house para amplificação e sequenciamento do genoma completo do Zika vírus*", do pesquisador Prof. Dr. **Paolo Marinho de Andrade Zanotto** e do aluno Shahab Zaki Pour.

Caberá aos pesquisadores elaborar e apresentar a este Comitê, relatórios anuais (parciais e final) de acordo com a Resolução nº 466/12, item II, II.19 e II.20, do Conselho Nacional de Saúde, conforme modelo constante no site: ww2.icb.usp.br/icb/cepsh, como também finalizar o processo junto à Plataforma Brasil quando do encerramento deste projeto.

O primeiro relatório deverá ser encaminhado à Secretaria do CEP em **21/09/2018**, bem como anexado uma cópia à Plataforma Brasil.

Atenciosamente,

Profa. Dra. **CAMILA SQUARZONI DALE**
Coordenadora do Comitê de Ética em
Pesquisas com Seres Humanos - ICB/USP

PERSONAL ACKNOWLEDGEMENTS

FINANCIAL ACKNOWLEDGEMENTS

Resumo

ZAKI POUR, SH. Desenvolvimento e validação experimental de uma metodologia in-house para amplificação e sequenciamento do genoma completo do Zika vírus. [Dissertação (Mestrado em Microbiologia)]. São Paulo : Instituto de Ciências Biomédicas, Universidade de São Paulo, 2018.

O zika é um arbovírus emergente. Há evidências para a relação entre o zika e a microcefalia congênita e também com a síndrome de Guillain-Barre. Várias características do vírus são importantes, como a persistência do vírus no sêmen por vários meses, transmissão sexual e evidência de transmissão pré-natal. As mães grávidas infectadas com zika podem dar à luz crianças aparentemente saudáveis que podem apresentar manifestações e complicações tardias. Existe uma clara necessidade de diagnosticar e sequenciar amostras clínicas do ZIKV que circulam na América do Sul, especificamente no Brasil. No entanto, as baixas cargas virais observadas que são observadas comumente em amostras humanas constituem um fator complicador para detecção, amplificação e sequenciamento. Neste projeto, propor projetar um fluxo de trabalho otimizado para o sequenciamento completo do genoma com base no pré-enriquecimento por PCR (reação em cadeia da polimerase) e pools de amplicons.

Palavras-chave: Zika. Reação em cadeia da polimerase. Enriquecimento prévio. Sanger. Sequenciamento de nova geração. Amplificação de genoma viral completo.

Abstract

ZAKI POUR, SH. Development and validation of an in-house method for whole genome amplification and sequencing of Zika virus. [Master Dissertation (Microbiology)]. São Paulo : Instituto de Ciências Biomédicas, Universidade de São Paulo, 2018.

Zika is an emerging arbovirus. There is enough evidence for the relation between Zika and congenital microcephaly and also with the Guillain-Barre syndrome. Several characteristics of the virus are important, such as persistence of the virus in semen for several months, sexual transmission and evidence of prenatal transmission. Zika infected pregnant mothers may give birth to apparently healthy children that may show late manifestations and complications. There is a clear necessity of diagnosing and sequencing clinical samples of ZIKV circulating in South America, specifically in Brazil. Nevertheless, the observed low viral loads that are commonly in human samples constitute a complicating factor for detection, amplification and sequencing. In this project, we aim to design an optimized workflow for full genome sequencing based on pre-enrichment by PCR (polymerase chain reaction) and amplicon pools.

Keywords: Zika. Polymerase chain reaction. Enrichment. Sanger. New generation sequencing. Whole genome amplification.

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF ABBREVIATIONS & SYMBOLS

B19V - Parvovirus B19
BLAST - Basic Local Alignment Search Tool (GenBank)
bp - base pairs
CHIKV - Chikungunya virus
CMC - Carboxymethyl Cellulose
CC - Celular Culture
Ct value - Cycle Threshold
DC-Sign-Dentritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non -
integrin
DENV - Dengue Virus
DMEM - Dulbeco Modified Eagle Medium
DNA - Deoxyribonucleic Acid
ELISA - Enzyme – Linked Immunosorbent Assay
ER - Endoplasmic Reticulum
FBS - Fetal Bovine Serum
IFA - Immunofluorescence Assay
IEC - Zika Isolated by the Evandro Chagas Institute
NCBI - National Center for Biotechnology Information
NGS - Next generation sequencing
NS1 - Nonstructural Protein 1
nt - Nucleotides
NT - Neutralization Test
ORF - Open Reading Frame
PBS - Phosphate Buffered Saline Solution
PCR - Polymerase Chain Reaction
PFU - Plaque-Forming Unit
prM - Premembrane
PRNT - Plaque Reduction Neutralization test
qRT-PCR - Quantitative Reverse-Transcription Polymerase Chain Reaction
PS - Porcine Stable Kidney Cells
RdRp - RNA-dependent RNA polymerase
RNA - Ribonucleic acid
RPM - Round Per Minutes
Sanger - Is a Method of DNA sequencing based on dye-labaled terminator
sequencing
SfRNA - Non coding Subgenomic Flaviviruses RNA derived from 3' UTR
SNV - Single Nucleotide Variation
SSRNA - Single Stranded RNA
TAM - Tyro-3, Axl, and Mer constitute the TAM family of receptor Tyrosine kinases
Tm - Melting Temperature
UTR - Untranslated region
VP1 - Viral Protein 1
WHO - World Health Organization
WNV - West Nile virus
XRN1 - Exoribonuclease 1
YFV - Yellow Fever virus
ZIKV - Zika Virus

TABLE OF CONTENTS

1. Introduction

Zika virus (ZIKV) was first identified in a rhesus monkey in the forest of Ziika in Uganda in 1947 (1). Then the virus was recovered from the mosquito *Aedes africanus*, caught in the same forest in 1948 (2). The first human cases of ZIKV were detected in Uganda and the United Republic of Tanzania in 1952 (3). In 1964 a researcher from Uganda was infected while working with the virus, confirming the Zika virus disease in human (4). Human cases were confirmed, although no hospitalization was reported between the 1960's and 1980's. The disease then moved from Uganda to western Africa and Asia in the first half of the 20th century (5, 6). ZIKV was detected in mosquitoes found in equatorial Asia, including India, Indonesia, Malaysia, and Pakistan from 1969 to 1983 (7). The first ZIKV large outbreak in humans reported in the Pacific Island of Yap in the Federated States of Micronesia with an estimated 73% of residents infected in 2007. Prior to this, only 14 cases of human ZIKV were documented around the world (8). In 2008, a US scientist conducting field-work in Senegal fell ill with ZIKV infection. On his return home to Colorado, he infected his wife and that was the first documented case of sexual transmission of ZIKV (9). In 2013 and 2014, outbreaks occurred in the Pacific: French Polynesia, Easter Island, the Cook Island, and New Caledonia. After infection has been linked to microcephaly in Brazil, thousands of previous suspected infections in French Polynesia were re-investigated in order to establish and confirm a possible association between the ZIKV virus and congenital malformation and to severe neurological and autoimmune complications (10, 11). On March 2014, during the outbreak in French Polynesia, two mothers and their newborns were found infected. The infants possibly acquired the infection by transplacental transmission or during delivery (12). At the same time, during the outbreak in French Polynesia, 1,505 asymptomatic blood donors reported being (polymerase chain reaction) PCR positive for ZIKV alerting the authorities that the virus can be passed on through blood transfusion (13). Brazil notified WHO of an illness with the symptom of rash in north-eastern states on the 29 of March 2015. From February 2015 to 29 April 2015,

nearly 7000 mild cases are reported, with no associated deaths. Of 425 blood samples tested, 13% were dengue positive. Tests for Chikungunya, Measles, Rubella, Parvoviruses B19, and Enteroviruses were negative. At that time, ZIKV was not suspected and was no tests were carried out for its detection. On May 2015, Brazil National reference laboratory, The Oswaldo Cruz Institute confirmed ZIKV circulating in Brazil and this was the first report of locally acquired ZIKV in the Americas (14). On November 2015, Brazil declared a national public health emergency, as cases of suspected ZIKV associated with microcephaly continued to increase. At the same time, in Brazil it was detected ZIKV genome in blood and tissue samples of a baby with microcephaly (15). Afterward, during January of 2016, it was reported 3,893 suspected of microcephaly and 1,708 cases of Guillain-Barre syndrome in Brazil. On September 2016 WHO concluded that Zika virus infection during pregnancy is the cause of congenital brain abnormalities (14). Finally, on the18 of November 2016, WHO declares the end of the public health emergency of international concern regarding microcephaly, although concerns still prevail in Brazil and more research will be necessary to understand ZIKA evolution, host adaptation, viral virulence and molecular epidemiological investigations (16). Recently Zika was included in the top ten emerging pathogens with the potential of causing a public health emergency and the absence of effective drugs or vaccines by a panel of scientists and public health experts convened by WHO in the January of 2017, also in the Second annual review in the February 2018 (17).

1.1 The Zika virus

Arboviruses (Arthropod-borne viruses) comprise more than 500 viruses transmitted either by insect vectors or spread as a zoonotic agent. Arboviruses are classified according to antigenic and phylogenetic relationships, morphology, and replicative mechanisms. Arboviruses are included in different taxonomic families, including Flaviviridae (genus Flavivirus), Bunyaviridae (genus Nairovirus, Orthobunyavirus, Phlebovirus, and Tospovirus), Togaviridae (genus Alphavirus), Rhabdoviridae (genus Vesiculovirus), Orthomyxoviridae (genus Thogotovirus), and Reoviridae (genus Orbivirus and Coltivirus) according to International committee on taxonomy of viruses (18).

Mosquitoes are the main vectors for most arboviruses, although other biting flies, midges, and ticks may also transmit these diseases. Humans are understood as important hosts because of the huge outbreaks observed in the Americas. Humans transmit disease to female mosquitos during blood feeding. The virus replicates in midgut cells and moves to the hemocoel (the primary body cavity of most invertebrates, containing circulatory fluid) infecting subsequently the salivary glands, from where is transmitted back to humans.

ZIKV belongs to the genus Flavivirus in the family Flaviviridae. The genus Flavivirus includes 53 other viral species, as well as the Dengue, Yellow Fever, Saint Louis encephalitis and West Nile viruses (19). ZIKV belongs to the Spondweni serogroup which shares serological cross-reactivity and similar clinical presentations. It has a small virion of approximately (50 - 60) nm in size (20), single-stranded RNA of positive sense, around 11 kb in length (21), with a single ORF (open reading frame) flanked by 5' (106nt) and 3' (428nt) UTRs (*untranslated region*) at both ends. It has 5' cap structure at 5' end and not polyadenylated at the 3' end, but makes a secondary loop structure that leads to the formation of a subgenomic flavivirus RNA (sfRNA) that is abundant non coding subgenomic RNA in infected cells through genomic RNA degradation by the host XRN1 exonuclease. The sfRNA is an extension of the 3' UTRs and it is essential for pathogenicity (22, 23). The single ORF then encodes a polyprotein precursor, which is cleaved by both host and virus enzymes, resulting the tree structural viral proteins C (capsid), preM (pre-membrane), E (envelope) and seven non-structural proteins NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5 (24). NS1 is essential for virus replication and inhibition of complement-mediated immune response (25) and makes multimers with different functions during the infection cycle, including dimers involved in the replication complex in vesicles and hexamers, complexed with lipids that are secreted to the extra cellular environment. NS3 combines helicase/NTPase, serine protease, and RNA triphosphatase activity (26 - 28). NS2B is a cofactor for the protease activity of NS3. NS5 contains a methyltransferase and RNA-dependent RNA polymerase (RdRp) domains and is necessary for genome replication also capping of nascent RNA (29).

Figure 1 – ZIKV genome organization based on the MR-766 isolate
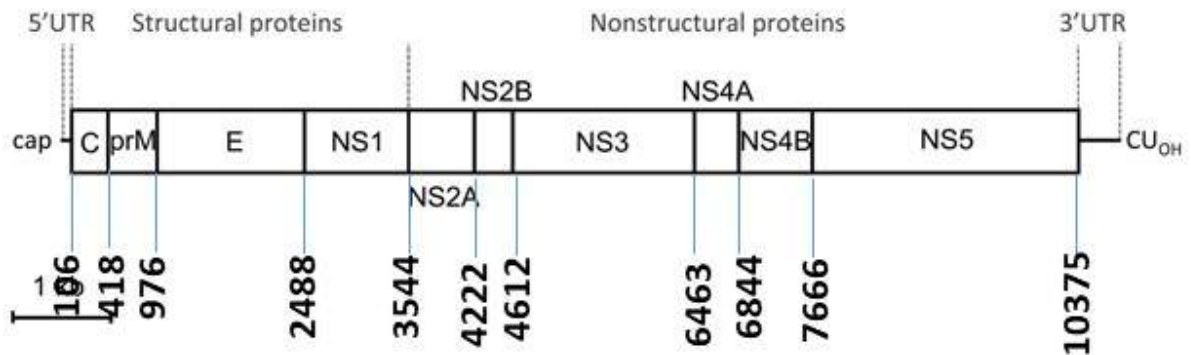
Untranslated regions play a fundamental role by RNA cyclization during the Flaviviruses replications (Figure 2) (30). This complementary requirement between 5' and 3' untranslated regions not only showed by RNA secondary structure prediction, but also by using infectious clones and replicon systems of DENV and WNV (31, 32). These RNA elements within the UTRs include 5' stem loops A and B (5' SLA and 5'SLB respectively), 5' and 3' upstream AUG, 3' cyclization sequence, 3' short hairpin structure (sHP), the highly-conserved 3'SL and the 5' cyclization sequence, and the capsid-coding region hairpin element (cHP) that lies within the ORF The translation initiator AUG at the 5' UTR found to be complementary to the region present at 3' SL (stem loops) which is called cyclization sequence 5'-3' UAR (Figure 2) (33).

Figure 2 – 5' and 3' untranslated regions structures (33)
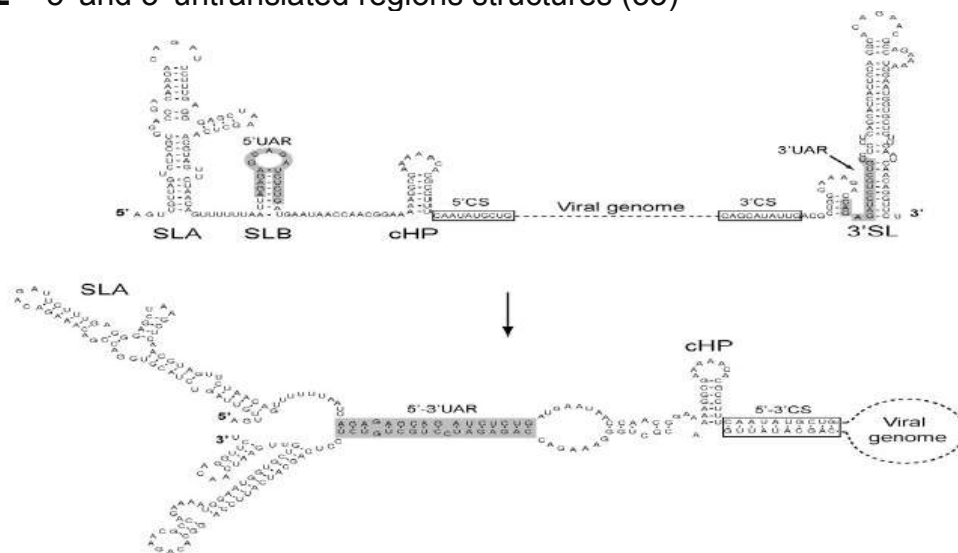
The transmission to humans occurs through the bite of infected mosquitoes. After the entrance of the virus, it interacts with host cell receptors like DC-SIGN (Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-Integrin) known as CD209, and also TAM family of the receptor, tyrosine kinase (Tyro-3, Axl, and Mer). With the attachment of viral envelope protein (E) and mediation of host cell receptor internalization occurs by endocytosis and the virus fuses with the endosome. Low pH causes the release of genomic RNA into the host cell cytoplasm. The positive ssRNA in replication vesicles is translated into a large polyprotein that is subsequently cleaved into mature structural and non-structural proteins. Negative ssRNA then is synthesized from the positive ssRNA serving as a template strand for viral genome replication by the viral-encoded RdRp and finally assembling. Polyprotein will be made. Then the polyprotein is cleaved into separate, mature proteins. Replication takes place at the surface of Endoplasmic Reticulum (ER). Virus assembly occurs in the endoplasmic reticulum and the virion buds at the ER and is translocate to the Golgi apparatus. The prM is cleaved into the Golgi and then the mature virion is released by exocytosis (34 - 40).

Based on the genome sequencing and phylogenetic trees, three distinct genotypes were identified, West African (Nigerian cluster), East African (MR766 prototype cluster), and Asian (41- 43), (Figure 3).

Figure 3 – Maximum likelihood phylogenetic tree of ZIKV
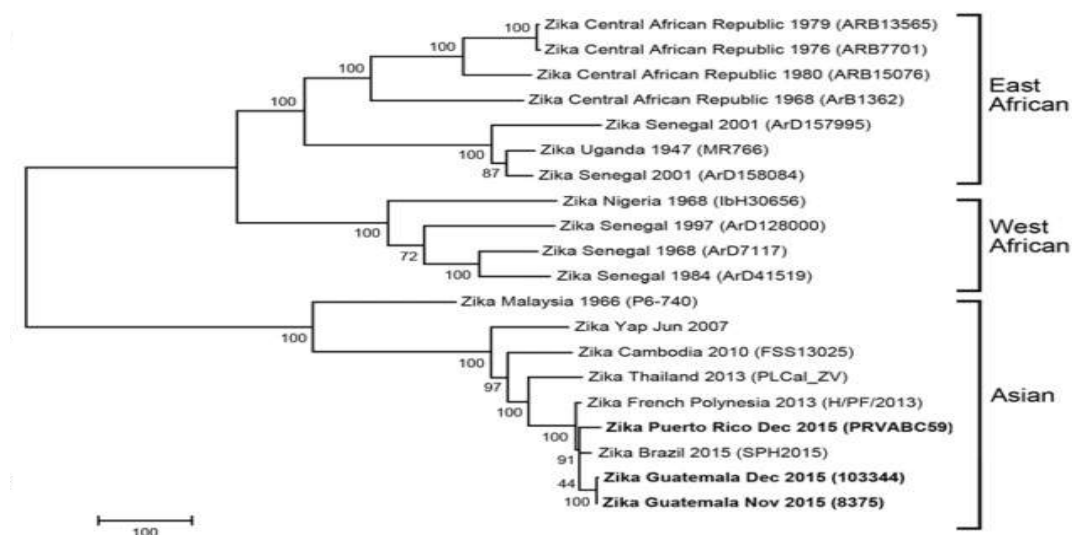
2. **Justification**

Although there is some proof of ZIKV persistence and shedding in vaginal secretion, urine, tears, saliva, semen for weeks, even months. Generally, the viremic period is short, allowing direct virus detection from blood and saliva, usually during the first 3–5 days after the onset of symptoms (sometimes up to 7–8 days) (44 - 49). Enough genomic material should be provided to sequencing with all different sequencing platforms, including Sanger and Next Generation Sequencing.

Figure 4 – ZIKV detection by different methods
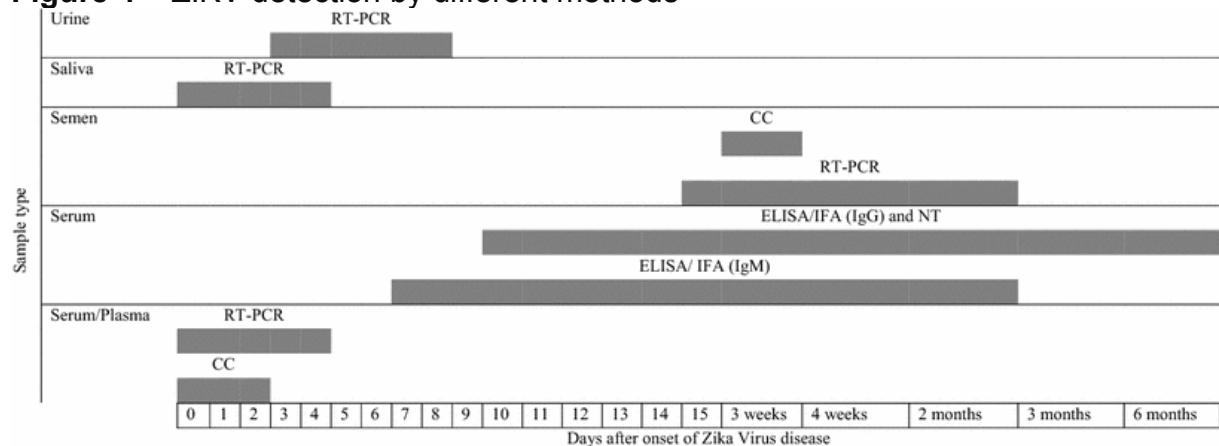
An overview of the laboratory diagnosis tests shows the optimal periods for ZIKV detection by different techniques. The genomic material is detectable in urine and saliva for a longer period in comparison with serum and plasma by RT-PCR. Cell culture (CC) can be useful to isolate virus until the second day of viremia, but up until 3 weeks in the case of semen. ELISA (Enzyme-linked Immunosorbent Assay), IFA (Immunofluorescent Assay), also NT (Neutralization test) is applicable from 7 days to 2 months after infection and in the case of IgM and much longer by IFA and also by detection of IgG.

When an individual shows clinical signs and some disease manifestations, he may seek medical attention, although a large majority of cases are asymptomatic. Typically, most infected individuals never develop symptoms and when symptoms occur, they are rather mild and can also be mistaken by other infectious diseases.

During the acute phase, generally the immune system is able to control the virus in the blood and viral titers tend to get low and eventually vanish (Figure 5).

Figure 5 – The course of symptoms, antibodies and viremia
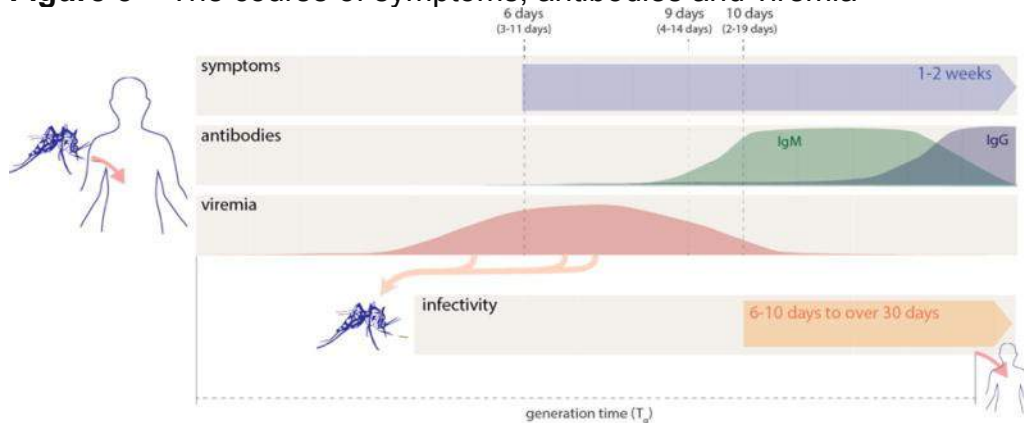
On average, the symptoms would appear in 6 to 11 days and during this period IgM (Immunoglobulin) and IgG is produced. Viremia is even shorter than 10 days (50).

Because of the low copy of ZIKV in the blood and also due to the short viral incubation time, the amplification and sequencing of Zika constitute a significant challenge for virologist and clinical laboratories. This is best shown by the fact that only around 68 complete genomes have been sequenced so far from Brazil. Although, better efficiency in the detection, due to higher viral load, has been reported from urine and saliva in comparison with other specimens is reported, in the beginning of 2016 the best choice was still serum. The results of RT-PCR from most of the blood samples tested in our laboratory shows that the CT (threshold cycle value) is more than 30, corresponding to less than 1000 genome copies of virus per milliliter. This could explain the difficulty found for amplification and sequencing. Sanger sequencing requires at least 3 - 75 ng depends on the product size and new generation sequencing needs 1ng of DNA input. It could be tedious, time-consuming and error-prone cloning whole viral genome into vectors with the objective of sequencing. Direct sequencing of clinical samples without previous passage in cell cultures is also a challenge. As mentioned above, a low number of Zika full genome sequences have been deposited in GenBank since the beginning of 2016. Moreover, low quality of the obtained sequences from some regions of the genome shows the difficulty of this job. In sum, the short viremic phase and low titers make the sequencing of ZIKA a great challenge.

There are a total of 68 complete genomes deposited at NCBI (National Center for Biotechnology Information), 13 sequences in 2016, 52 in 2017 and 3 sequences in 20018. From this 68 complete sequences, 27 were sequenced using the Illumina platform, 24 by MinION Oxford Nanopore, 5 by Sanger and three in not mentioned (Consulted at the 21 of March, 2018). Furthermore, the quality of the sequences available is not so satisfactory due to the presence of some gaps (unavailable sequences provided) (Figure 6).

Figure 6 – Schematic view of ZIKV sequences publicly available in NCBI
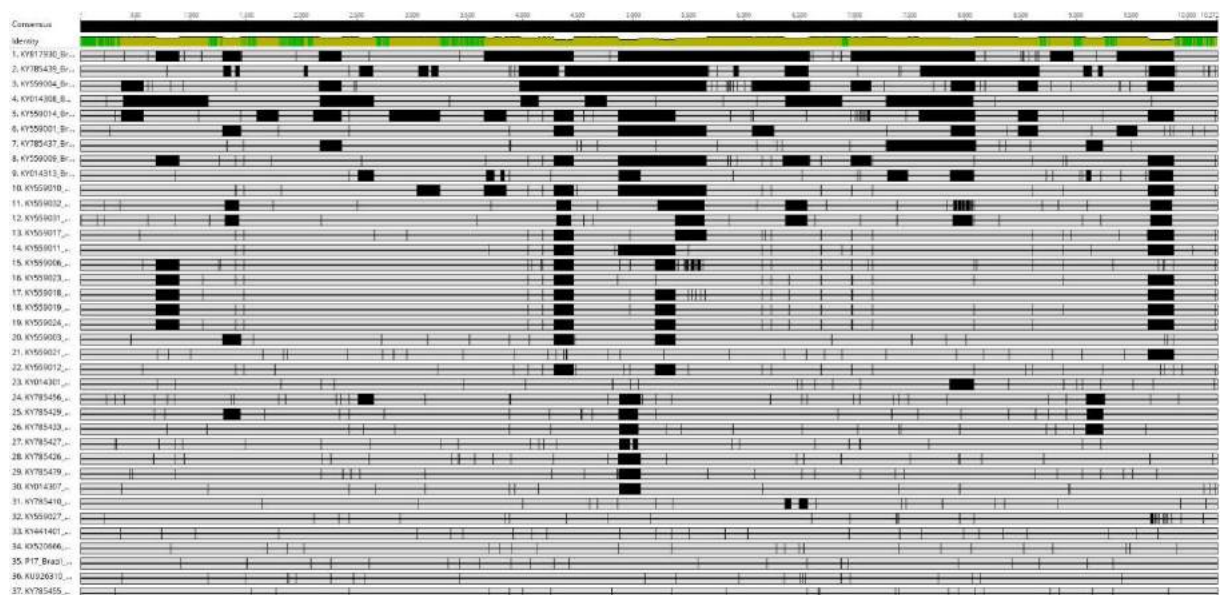
All 68 complete genomes publicly available from NCBI were aligned using MUSCLE V. 3.8.31. To have better alignment we manually cut the 5' and 3'UTRs. Then realign them again with the same program. The alignment opened with Geneious V 10.2.3 and reduced until all appear in a one page. The dark black lines appear due to the gaps and regions witch was not successfully sequenced by different platforms. Only the coding region was considered due to better alignment of this region.

3. **Objectives**

- Design new sets of overlapping primers for complete ZIKV genome Sequencing including 5´ and 3´ UTRs;
- Whole genome Sanger sequencing based on overlapping amplicons;
- Sequencing the amplicons generated by PCR with NGS;
- Evaluation of two amplification strategies for templates present at high and low copy number.

4. **Materials and methods**

4.1 Oligonucleotide primers

The fact that there is small diversity among ZIKV circulating in Brazil allowed us the design of a new set of primers to amplify the whole genome of ZIKV, including the 5' and 3' untranslated regions. We used all available complete ZIKV sequences retrieved from NCBI by March of 2016 (Accession numbers: KU365780, KU365777, KU365778, KU365779, KU729217, KU497555, KU321639). The sequences were aligned with ClustalX (V2.1). Subsequently the alignment was submitted as an input to the Primerdesign-M available in the Los Alamos HIV site (51, 52). Some options were considered to have the optimal results, according to the objective such as:

a) Gap stripping: This option will remove all columns of the submitted alignment having more than the selected percentage of gaps. A 50% value was chosen as a threshold to exclude regions that they have more than 50 % of the gaps.

b) Region of interest: The region of the alignment that needs to be amplified. In this case we choose 50 to 10600 avoiding regions of 5' and 3' UTRs due to high diversity. In our experience choosing such a diverse region would result in an error.

c) Multiple fragments: In this case multiple fragments were chosen for the suggestion of the set of primers and not just a single fragment.

d) Fragment overlap: For multiple primer pairs, these options dictate how much overlap will exist between each successive primer pair. There are some options available like Max, Min, Mid, Zero and Flex, which regard to the position of the next primer in the first 25%, near the middle, in the last 25%, zero or flexible to the position the next forward primer of the fragment.

e) Adaptor options: It was chosen no adaptor. It means there is no synthetic DNA attached to the construct. Because we chose "no adaptor", we needed to provide a read length that obviously depended on the system of primer design and on the different sequencing technologies due to different read length limitation. Therefore, it required from us a minimum and maximum read length of 600 to 900 bp, which is a desirable size for Sanger sequencing with acceptable quality.

f) Tag options: You can choose how to optimize the tags/barcodes. It was chosen no tag.

Primer options: Such a minimum and maximum primer size, detection limit (%) can be between zero to twenty for determining the level at which rare variants should be included in the design (the default was chosen which is 5), complexity limit, the maximum difference in Tm (Melting temperature), dimer window size, dimer maximum ratio, G/C clamp. We choose a minimum of 20 and maximum of 24 with the complexity limit of 1 for avoiding degenerate primers. Detection limit of 5 with is the default for avoiding degenerate primer design and 5 Cº of maximum Tm difference. The dimer window size of 10 by default was chosen. This was because it is the proper window size for primer-primer dimerization. Finally, G/C clamp, which allows to specify that the 3' end of the primer must be a G or C, which, helps promote specific binding due to the strong bonding of G and C bases. Nevertheless, more than 3 G's or C's should be avoided in the last 5 bases at the end of the 3' of primers.

Although the result was a suggestion list to cover all the regions of interest for amplifying whole genome (almost 400 primers), 39 desirable primers were chosen manually, using different criteria such as:

a) MegaBlast result for checking the specificity, genotype, query coverage, identity.

b) Tm of each primer again verified with different available tools in silica utilizing Oligo Calculator V 3.27 available at http://biotools.nubic.northwestern.edu/OligoCalc.html and also OligoAnalyzer 3.1 available at https://www.idtdna.com/calc/analyzer.

c) Overlap of at least 100 bp.

d) Amplicon size which can be between 700 to 900 bp.

e) Tm difference of not more than 5 Cº for the reverse and forward primers.

f) GC % between 40 and 50.

Initially, the properties of all primers were checked with the specific objective of determining the ideal annealing temperature *in silico* utilizing Oligo Calculator V 3.27 available at http://biotools.nubic.northwestern.edu/OligoCalc.html and also OligoAnalyzer 3.1 available at https://www.idtdna.com/calc/analyzer.

Table 1 – Oligo primers used in this study (Position based on KU940228)

Oligo Name	Sequence 5'-3'	Position	Size	Ta
ZIKbr-F1	GACAGTTCGAGTTTGAAGCGAA	28-49	22	55
ZIKbr-F2	TCAATGCTAGGAAGGAGAAGAAGA	389-412	24	56
ZIKbr-F3	AGCACAGTGGGATGATTGTTAATG	1415-1438	24	54.5
ZIKbr-F4	GTGTCATACTCCTTGTGTACTGC	1882-1904	23	53
ZIKbr-F5	GGGTCTGAACACAAAGAATGGATC	2400-2423	24	54
ZIKbr-F6	ATAACAGCTTTGTCGTGGATGGTG	2876-2899	24	56
ZIKbr-F7	GAAGGGTGATCGAGGAATGGT	3401-3421	21	53
ZIKbr-F8	CATTGAAAGAGCAGGTGACATCA	4377-4399	23	54.5
ZIKbr-F9	GGAGAGAGAGCGAGGAACATC	4918-4938	21	55.5
ZIKbr-F10	ACCTTCACTTCACGTCTACTACAG	5404-5427	24	55
ZIKbr-F11	ATACTTGATGGCGAGAGAGTCATT	5910-5934	24	56
ZIKbr-F12	AGAGTTTGTTCAGATCATGCGG	6403-6424	22	54.4
ZIKbr-F13	AGTAGGTCTTCTGGGCTTGATTAC	6885-6908	24	55
ZIKbr-F14	CATTGATCTTGGATGTGGCAGAG	7896-7918	23	55.5
ZIKbr-F15	AGTGAAATATGAGGAGGATGTGGA	8415-8438	24	54
ZIKbr-F16	AGGGGCAATATTTGAAGAGGAAAA	8895-8918	24	53.5
ZIKbr-F17	CCAAAACAAAGTGGTAAAGGTCCT	9387-9410	24	53
ZIKbr-F18	AGCACCAATCTTAATGTTGTCAGG	10377-10400	24	54.5
ZIKbr-R1	TCTATTGATGAGACCCAGTGATGG	286-309	24	55
ZIKbr-R2	CTAGTGGAATGGGAGGGGAGC	759-779	21	55
ZIKbr-R3	CCTCTGTCCACTAACGTTCTTTTG	1251-1274	24	55
ZIKbr-R4	GTGAACTGCTCCTTCTTGACTC	1752-1773	22	54
ZIKbr-R5	CTCCAACTGATCCAAAGTCCCAG	2259-2281	23	55.5
ZIKbr-R6	GGCAATCTCTGTGGACCTCTC	2781-2801	21	54
ZIKbr-R7	CCTTTCATTTGGGTCCTGTAGC	3263-3284	22	53.5
ZIKbr-R8	CTGACTTTGAATGCCGCTATCAG	3778-3800	23	57
ZIKbr-R9	ATCTCTATATCTGCCTTGGCGAA	4282-4304	23	56
ZIKbr-R10	ATCAAGTCTCCCTTCACCGCTT	4788-4809	22	56
ZIKbr-R11	GTTGGAGCTAAGATCACGGTG	5265-5285	21	53.5
ZIKbr-R12	CTCTGTCTCAAAAGTCTTTCTGCT	5770-5793	24	56
ZIKbr-R13	CAAAGCACCATCTTCTATCTGTGT	6278-6301	24	55
ZIKbr-R14	GAAACACAACAATGAGGACACATG	6779-6802	24	53
ZIKbr-R15	CTGGGATCAAGTACATGTAGTGC	7266-7288	23	54
ZIKbr-R16	CTCTTCTCTGCACACCTCGGT	7762-7782	21	56.5
ZIKbr-R17	CAGGAAGAGACCATGCTCATAAC	8785-8807	23	54.5
ZIKbr-R18	CAGTGTCATCTGCATACATCCTTC	9251-9274	24	55
ZIKbr-R19	ACTTCTTCCCAGTTGTCCCATC	9767-9787	22	53
ZIKbr-R20	CGCACCATGTTGACTGTGTTTT	10262-10283	22	54
ZIKbr-R21	CTCTAACCACTAGTCCCTCTTCTG	10649-10672	24	56

4.2 Study design

We included several samples to test our system: (*i*) the IEC isolated was used for optimization of whole workflow, from the extraction to sequencing. It was sequenced using the Sanger method using both short and long amplicons with Illumina; (*ii*) 6 temporal samples collected from the same patient during 2 months; (*iii*) a sample collected in late 2014, which was sequenced using short amplicons using the Illumina technology.

4.3 Optimizations assays and protocol validations

4.3.1 Cell culture infections

All positive control and viruses used for the optimization since extraction, primers optimizations and validations, and finally the sequencing were obtained from C6/36 and Vero infected cells. The virus used for this purpose was kindly supplied by the Evandro Chagas Institute (called ZIKV-br in this project). 200 µL of virus suspension were inoculated in 10E6 cells, seeded in the previous day that were growing at 28°C in Leibovitz's L- 15 and 100 g/ml, respectively. Green monkey kidney (Vero), cells were cultured as monolayers at 37°C at 5% CO_2 in Eagle's minimum essential medium with 5% calf serum and penicillin G, streptomycin, at 100 IU/ml and 100 g/ml, respectively. CPE (cytopathic effect) has been observed in a week.

4.3.2 Virus titration

Approximately 10E5 PS cells (porcine stable kidney cell) kindly provided by the Institute Pasteur in Dakar were seeded into a 24 well plate, each well containing 500 µL of complete DMEM (DMEM supplemented with 10% of FBS and 1x penicillin-streptomycin-glutamine solution) and incubated for 24 h at 37C° also 5% CO_2. On the next day, 10-fold serial dilution of viral stock consist of 100 µL of virus in 900 µL of DMEM supplemented with 10% of FBS was prepared and labeled from 10E-1 to 10E-10. 100 µL was discarded from the last tube. Afterwards, the medium was discarded and the cells were washed one time with FBS 37 C° to get rid of cell

derbies and dead cells. Respectively, each two wells from the plate seeded at the previous day got infected with 200 µL from each 10-fold serial dilution in duplicates. Two wells were left just with cells as a control. The plate was incubated at 37 ºC with 5% CO_2, shaked mildly each 20 minutes for two hours. The supernatant was then discarded, 250 µL of CMC 3.2% (Carboxymethyl Cellulose) and 250 µL of culture 2x medium with 10% FBS was added and mixed by gentle. The plate was incubated for 4 days, then the supernatant (CMC and culture medium) is discarded and washed with PBS 1x (Phosphate-buffered saline) three times, each time with 1 ml. 3% Paraformaldehyde was added and left for 20 minutes at room temperature. Supernatant was discarded and washed with 1 ml of PBS 1x three times. 300 µL of Triton X100 (0.5%) was added and incubated for 4 minutes. Again the supernatant was discarded and washed one time with PBS 1x. 300 µL of crystal violet was added and the plate stayed at room temperature for 30 minutes. Finally, the plate was washed with normal tap water and stayed at room temperature until dry.

The calculation of viral titer based on the plaque assay method is a quantitative measurement of the biological activity of the virus and is expressed as a PFU/ml. To estimate the titer, the number of each clear plaque that arises from a single virus infection is counted. The average of replicates of the same dilution is calculated and discounted wells less than 5 or greater than 100 plaques. The negative control should have a uniform monolayer form and should stay intact. Finally, PFU/ml is the Average of plaques divided by the dilution multiplied by the volume of diluted virus added to the plate (PFU/ml = Average of plaques/D x V). For instance, if around 30 plaques were counted and 10E7 is the dilution and the inoculum was 0.2 ml, then it would yield a titer of 30/10E-7 x 200 that would be 10E6 PFU/ml.

4.3.3 Sample used in this study

Samples were obtained from the central laboratory of Sergipe (LACEN) in Aracajú or were collected in the region by our research group or were provided by the staff of the LACEN, during our collaborative investigation of the 2014/15 ZIKA outbreak in Sergipe. Patient samples were obtained under a term of informed

consent, following the protocols approved by the Ethics Committee on Human Research (CEP-ICB) of the Biomedical Research Institute (ICB) at the University of São Paulo (USP). One of these samples S007 from late 2014 was chosen.

Samples from patient 17 (P17) were kindly provided by Prof. Dr. Edison Luiz Durigon. They were isolated by Érica A. Mendes, Ph.D, Nicholas Di Paola, Ph.D and also Danielle B.L Oliveira, Ph.D from the Department of Microbiology at the Institute of Biomedical Science at the University of São Paulo. Sperm samples were collected at different time points from the infected patient (P17) (Table 2).

Table 2 – All the samples utilized in this study

Name	Material	Collection date	Ct value
P17-A	Sperm	26/04/2016	19
P17-C	Sperm	10/05/2016	23
P17-H	Sperm	16/06/2016	24
P17-K	Sperm	05/07/2016	26
P17-S	Sperm	1/09/2016	33
P17-T	Sperm	06/09/2016	34
S007	Sera	2014	33
ZIKV-br	Cell culture	05/05/2016	15

4.3.4 RNA isolation

Viral RNAs were extracted from the supernatant culture medium and clinical samples using the QIAamp Viral RNA Mini Kit (Qiagen) according to manufacturer recommendation and finally eluted in 60 µM of elution buffer provided in the same kit. Samples were quantified using the nanodrop 1000 spectrophotometer (Thermofisher) and Qubit RNA Assay Kit with 3.0 Fluorometer instrument (Thermofisher).

4.3.5 qRT-PCR

The CT value of all our samples during the optimization process were estimated using a set of primers and probe designed by Lanciotti et al. (53). 25µL reactions containing 5 µL viral RNA, 12.5 µL of 2x RT-PCR buffer, 1 µL of the mixture of forward and reverse primer and probe with the concentration of 10 µM (for the final concentration of 400 nM of each one), 1 µL of25x RT-PCR Enzyme mix and 4 µL

nuclease-free water were incubated at 45°C for 10 min for cDNA synthesis, 95°C 10 min for reverse transcriptase inactivation, PCR cycling and detection 40 cycles of 15 Sec at 95°C, 45 Sec at 60°C for data collection step with the 7500 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA).

4.3.6 cDNA synthesis

Extracted RNA as previously described was converted to cDNA with High-Capacity cDNA Reverse Transcription Kits (Applied Biosystem) using random primers. Briefly, 20µL reactions containing 5 µL nucleic acids, 2 µL of 10 mM random primer, 1 µL multiScribe RT, 2.5 µL 10X RT buffer, 1 µL (0.1 M) DTT, 0.8 µL (10 mM) of each dNTPs, 1 µL (40 U/µL) RNaseOUT ribonuclease inhibitor (Invitrogen), and 6.7 µL nuclease-free water was incubated at 25°C for 10 min, 37°C for 120 min, with inactivation step of 85°C for 5 min.

4.3.7 Short and long amplicons: Two different strategies

We used two distinct strategies for the amplifications based on the low and high-copy number (RNA input base on the Ct value). Cruz et al, 2016, showed that it is possible to amplify whole DENV genome with two different strategies base on a high- copy and low-copy templates. Low-copy numbers can be amplified with 10 overlapping segments and high-copy numbers just with 5 overlapping amplicons. Therefore, the idea was tested in our case and was observed the same result as described by Cruz et al, 2016. Later in October of 2017 the same authors reported the same observation for ZIKV (54, 55).

Figure 7 – Short amplicon strategy for ZIKV full genome amplification
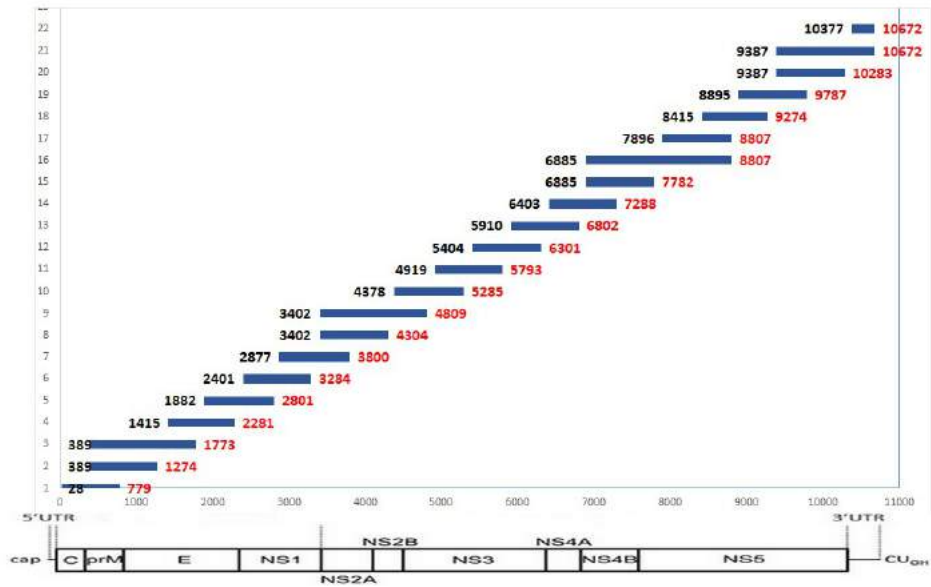
Figure 8 – Long amplicon strategy by just 6 amplicons
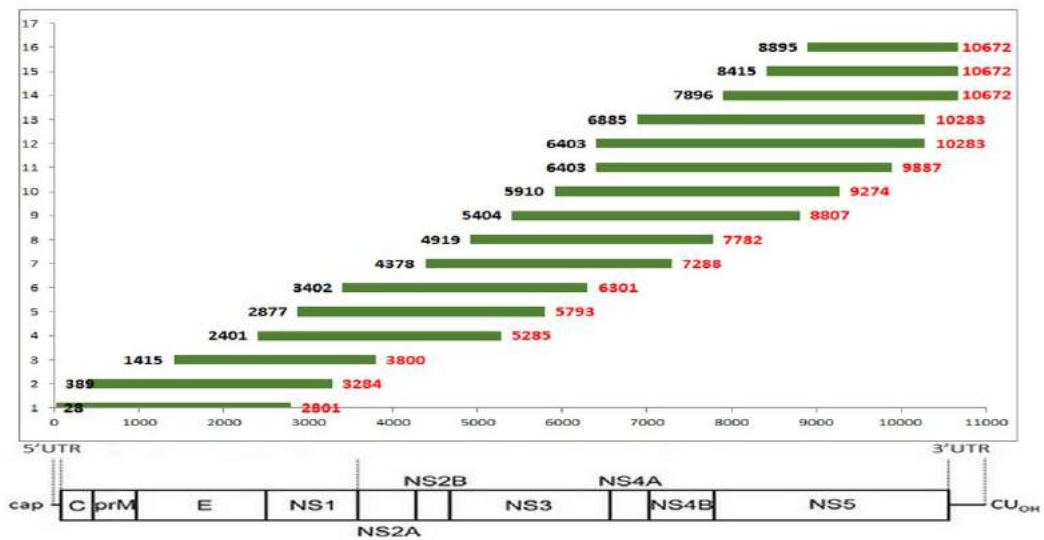
4.3.8 Amplication and amplicon generation

Amplicons were generated in 22 µL reactions containing 2 µL ZIKV cDNA previously described, 0.5 µL each (10 mM) forward and reverse primers, 0.1 µL PlatinumTaq DNA Polymerase High Fidelity (Invitrogen), 2.5 µL 10X PCR Buffer (reaching a final concentration of 1.13 x per reaction), 1 µL (50 mM) MgSO4, 0.5 µL (10 mM) dNTPs, and 14.9 µL nuclease-free water. The cycling conditions with MJ Research PTC-200 Peltier thermal cyclers from Bio-Rad consisted of Amplification was carried out as follows: 94 °C for 5 min, followed by 40 cycles at 94 °C for 15 s, 55 °C for 30 s, and 68 °C for 2 minutes and 30 seconds in the case of short amplicons and 4 minutes and 30 seconds in the case of long amplicons, followed by a final extension at 68 °C for 10 min. Amplicons were visualized on 1 % agarose gels submitted to electrophoresis at 90 Volts for 90 min in TAE buffer 1 %, utilizing the ladder of 100 bp (Invitrogen) in the case of short and 1 kb plus (Invitrogen) for long amplicons.

4.3.9 Limit of detection by convenience PCR

It was prepared a serial dilution of known titrated ZIKV (as previously described) from 10E9 to 10E5, spiked in serum which was already tested to be negative for ZIKV, to mimic clinical samples all carried out in triplicate. Viral RNA extraction was carried out with QIAamp Viral RNA Mini Kit (Invitrogen), following the manufacturer´s protocol. Purity and quality of extracted RNA were checked with a Nanodrop1000 Spectrophotometer (Thermos Fisher). qRT-PCR were done with Lanciotti primers the same as previously described. Each point was visualized on 1% agaose gels as previously described.

4.3.10 Standard curve for absolute quantitation

We prepared a triplicated serial dilution of already known titrated ZIKV. Subsequently RNA was extracted with QIAamp Viral RNA Mini Kit (Qiagen) and qRT-PCR carried out utilizing primers and probe from Lanciotti et al 2008. By doing so we can estimate the quantity of virus, according to the Ct value.

4.3.11 3' and 5' UTR regions

There are always some difficulties and challenges in the amplification of both 3' and 5' prime UTRs. Therefore, we tried to use different concentrations of enhancers such as DMSO, formamide, glycerol according to the literature (56 – 64). Successful amplifications were obtained by changing the forward primer for the 3´ UTR, as well as also starting the reactions after an incubation at 95°C for 5 minutes, also increasing annealing temperature from 50 to 55 C° followed by the remaining amplification cycles as previously described.

4.3.12 PCR product purification

There are different techniques available for the direct purification of PCR product in solution or from agarose gel bands. Generally, using these techniques we remove primers, salts, unincorporated nucleotides and also thermos-stable DNA polymerase, which may inhibit subsequent enzymatic reactions. It could also be useful for purifying cDNA for subsequent PCR reactions. It can be enzymatic like ExoSAP from Termo Fisher, using a silica column like the one we used, and also by simple steps of purification utilizing different alcohols like Ethanol and Isopropanol. For such a purpose, we used High Pure PCR cleanup kit from Roche according to the manufacturer's instructions. 20 µL of PCR product (amplicons) were adjusted to 100 µL with UltraPure DNase/RNase-free distilled water (Thermo Fisher Scientific) then 400 µL of binding buffer. Binding enhancer was not added to the reaction because of the size we were expecting to purify (>800 bp). Next, the mixture was added to the tubes with Silica and centrifuged for 1 minute. Subsequently, 300 µL and 400 µL of wash buffer were used respectively and centrifuged at 8000 RPM for a minute. Finally, eluted with 20 µL of elution buffer provided with the kit and centrifuged at 8000 RPM for a minute as well.

4.3.13 ZIKV full genome Sanger sequencing

Sequencing was done to verify the specificity of the primers and to check if complete genomes could be obtained. The cDNA was prepared as previously described. The primers used for sequencing PCR were diluted to a final concentration of 3 µM. Sequencing reactions was done utilizing ABI BigDye Terminator v3.1 (Applied Biosystem, Foster City, California, USA), containing 2.5 x Big dye 2 µL, 5x buffer 1 µL, primers (forward /reverse at once) 1 µL, nuclease free water 5 µL, DNA (Purified and quantified PCR product) 1 µL (At least 20 ng) for the final volume of 10 µL. It is Worthy to mention that all the amplicons (PCR product) were quantified by Qubit dsDNA HS Assay Kit (Thermo Fisher) after the purification and before the submission to the sequencing PCR. We adjusted the Nuclease free water, when was necessary to use more DNA (PCR product). The cycling conditions consisted of 96 °C 1 min, 25 cycles of 96 °C for 10 Sec, 50 °C for 15 Sec, 60 °C for 4 min, maintaining at 4 °C at the last cycle. For the Precipitation, we utilized X-terminator kit (Applied Biosystem, Foster City, California, USA) using 45 µL of SAM buffer, 5 µL of X-terminator, and 50 µL of the mixture per reaction. Then, 50 µL of the mixture of X-terminator was added to each well (MicroAmp Optical 96-Well Reaction Plate) covered by adhesive, then aluminum paper and shaking for 30 min on a shaker. The plate centrifuged at 3000 RPM for 3 min. We then transferred 10 µL of this mixture, to a new 96 plate, covered with septa. Sequencing was done on an ABI prism 3130 XL Genetic Analyser using the POP7 polymer according to the previously described by Sanger et al, 1977 (55).

Base calling, sequence assembly, and alignment, quality control were done with the CodonCode aligner software (CodonCode Corporation, Dedham, MA, USA).

4.3.14 Library preparation for Next generation sequencing

Double stranded DNA from PCR amplicons was used as the starting material to generate the Illumina library utilizing NEXTERA XT kit, following manufacture's recommendations and guidelines. As an input, this kit needs the RNA concentration of 1 ng, totally 5 µL of each sample. Initially, the first step is Tagmentation

(Fragmentation and tagging) of input DNA witch the DNA is fragmented in approximately 300 bp and also tagged the adaptors by the Transposase enzyme. The next step is PCR Amplification witch tagmented DNA is amplified via a limited-cycle PCR program, also adds indexes (i5 and i7) required for cluster formation. Next, the library DNA would be purified by AMPure XP beads from Beckman Coulter. Finally, libraries were diluted 1:1000 by library normalization buffer provided by the same kit. By this step we normalize the quantity of each library to ensure more equal library representation in your pooled sample.

Ready library was quantified in triplicate data points by qPCR using KAPA SYBR FAST qPCR kit (KAPA Biosystem). The quality of library and average size distribution were checked by Agilent High Sensitivity DNA kit (Agilent Technologies). The concentration of each library is calculated and adjusted by applying the formula provided by KAPA Biosystem library quantification kit as it is mentioned here, Library concentration = average concentration x (452/average fragment length) x 1000. Finally, normalized libraries were pooled together and sequencing with the NextSeq Reagents kit. The sequencing condition was 300 cycles, 2 x 150 bp, and final library concentration before the pool was 2 nM. (Illumin, USA).

4.3.15 Next-generation sequencing data analysis and processing

There are different software available for NGS data processing and analysis. Different pipelines can be designed. A reliable pipeline was designed and used as it can be observed in the Figure 7.

Figure 9– Next generation analysis pipeline
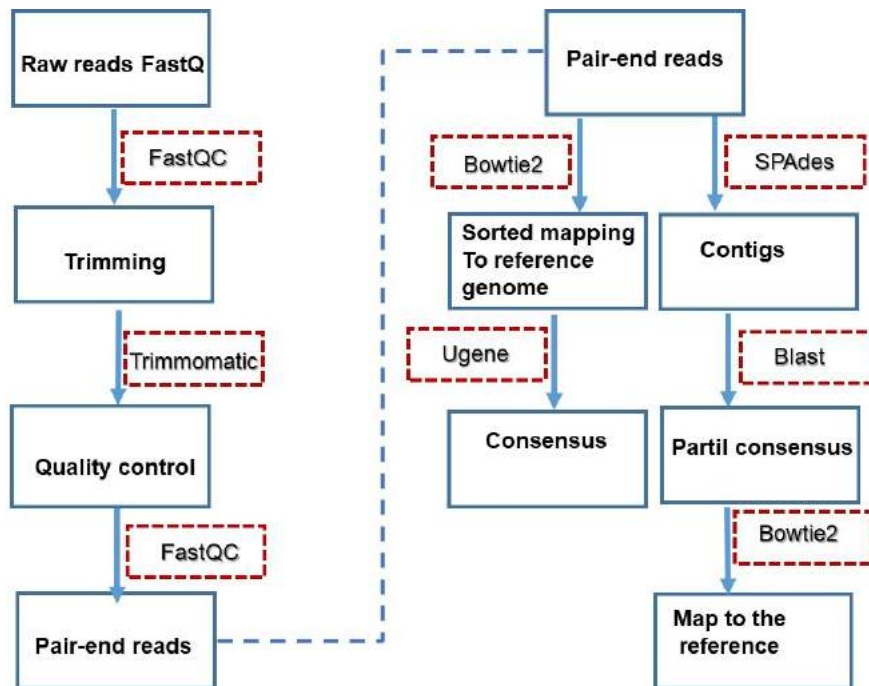
By checking the raw sequence data coming from high throughput sequencing machine, we were able to analyze the results quickly. For such a purpose, Fastq outputs were assessed for quality using FastQC V 0.11.6. The reports are in XML format and were saved and verified for the total sequences (Reads), sequence length, per base sequence quality, per sequence GC content, sequence length distribution, overrepresented sequences and adapter content (APPENDIX A).

4.3.16 Removing the adaptors and Trimming

Poor quality or technical sequences, such as adapters, were removed using Trimmomatic V 0.36. This software includes a variety of processing steps and algorithms for trimming and different filtering, such as size and quality. Some options and factors we implemented in the main script for better results. For the best options, we ran the same group of sequence reads using different options for downstream data processing. For instance, some of these factors can be observed bellow:

a) Quality scores were set to to Phred 33 (Probability of one error in 1000) as a kind of normalization for further trimming bases.

b) For each kit used for library preparation (Nextera XT), adapters were removed and they were further verified to ensure the successful removal.

c) Leading 20: Cut off low quality or N bases below a threshold quality from the start of reads. The default was 3 but we used 20 for having a better quality in reads.

d) Trailing: Cut off low quality or N bases below a threshold quality from the end of reads. The default was 3 but we used 20 for having the better quality in reads

e) SLIDINGWINDOW: Perform a sliding window trimming, cutting once the average quality within the window falls below a threshold. The default was 4:15 but we change it to 4:25. It means that the program scans the reads with a 4-bases wide sliding window, cutting the average quality per base drops below 25.

f) MINLEN: This option was used as the default value of 36. With this option, the sequences less than 36 bp were eliminated.

As a result of all these base trimming methods on the quality and size, we ended with files containing two single end and two pair end datasets. Once more, after trimming, the quality of the reads were checked by FastQC. The reads were assembled by mapping to the reference genome (It is used ZIKV isolate Paraiba Accession number of KX280026) utilizing Bowtie2 V 2.3.4 read alignment software. The resulting assemblies were checked using the IGV (Integrative Genomic Viewer) (65, 66) or Ugene (67, 68) or Tablet (69) programs. Furthermore, we also collected information such as a medium coverage, quality of each base, consensus sequence (APPENDIX A).

4.3.17 De novo Assembly

De novo assembly was done by SPAdes V 3.11 (70) that is an iterative short-read genome assembly module, based on the value of K-mers, selected on the basis of read length and data set type. Once again, after confirming the taxonomic identity of our query sequences, we aligned with a reference genome using Bowtie2 (71).

5. **Results**

5.1 Standard curves for absolute quantitation

Drawing a standard curve for absolute quantitation of an unknown amount is based on a known quantity. From an already titrated virus, we prepared 10-fold dilutions in triplicated 10 E6 to 10E2. RNA was extracted with QIAamp Viral RNA Mini Kit (Qiagen) and qRT-PCR carried out utilizing primers and probe from Lanciotti et al 2008, as previously described in details. Then the obtained Ct value is plotted against the logarithm of the corresponding template copy number. Each standard curve was generated by linear regression of the plotted points. Although each replicate had acceptable results with R^2 above 0.99, we considered the average of the tree 10-fold serial dilutions for further calculations.

Table 3 – Absolute quantification of ZIKV by Plaque Unit Assay

PFU/ ml	LOG PFU/ml	Curve 1	Curve 2	Curve 3	Average
600000	5.78	17.06	17.76	17.17	17.33
60000	4.78	19.2	20.57	20.75	20.17
6000	3.78	23.13	24.68	23.43	23.75
600	2.78	26.57	27.47	26.79	26.94
60	1.78	30.44	31.22	31.49	31.05
6	0.78	33.95	33.62	33.82	33.80

Figure 10 – Absolute quantification curve base on Plaque Unit Assay
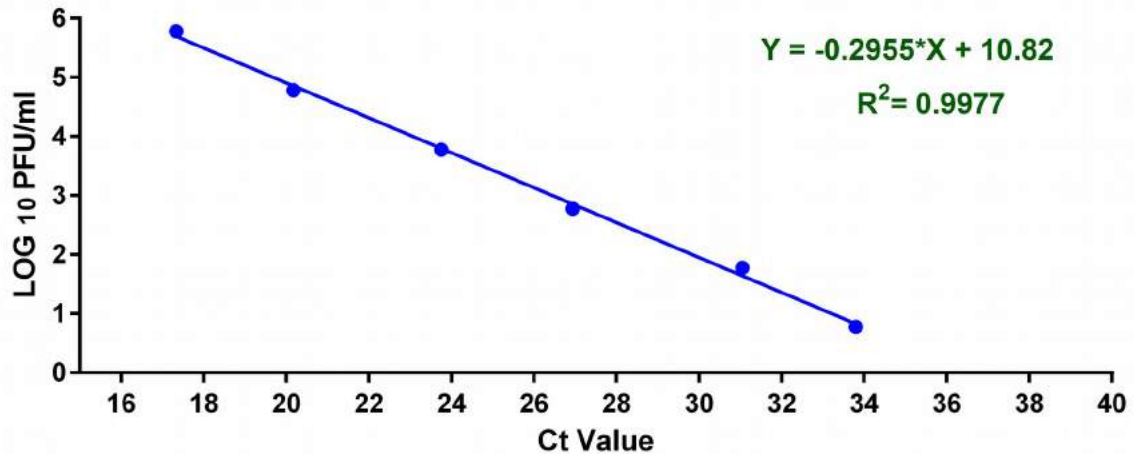
Given the standard curve, it was possible to make the relationship between the Ct value obtained and their respective RNA concentration. Using the Ct values we can estimate how many viral copy existed in a sample with the linear regression equation ($Y= a + b X$) where (Y) is the unknown concentration, (a) is the (Y) intercept, (b) is the slope and (X) is the obtained Ct value we can calculate unknown concentration base on the obtained Ct value.

5.2 Limit of detection (LD)

As previously described, we prepared a triplicate serial dilution of known titrated ZIKV from 10E9 to 10E5, spiked in serum that was already tested to be negative for ZIKV, to mimic clinical samples. Viral RNA was extracted with QIAamp Viral RNA Mini Kit (Qiagen) in accordance with the manufacturer's suggested protocol, posteriorly converted to cDNA using High-capacity cDNA reverse transcriptase (Invitrogen light). Conventional PCR was done following the same in-house protocol as usual, then PCR product was detected by 1% agarose gel electrophoresis, stained with ethidium bromide and observed under ultraviolet.

Figure 11 – The Limit of detection visualized by agarose gels

From left to right: 1 - 100 pb ladder (Invitrogen), 2 - ZIK-br-F1, ZIK-br-R 2 (751bp); 3 - ZIK-br-F 2, ZIK-br-R 3 (885 bp); 4 - ZIK-br-F 2, ZIK-br-R4 (1384 bp); 5 - ZIK-br-F 3, ZIK-br-R 5 (866bp); 6 - ZIK-br-F 4, ZIK-br-R 6 (919 bp); 7 - ZIK-br-F 5, ZIK-br-R 7 (884 bp); 8 - ZIK-br-F 6, ZIK-br-R 8 (924 bp); 9 - ZIK-br-F 7, ZIK-br-R 9 (903 bp); 10 - ZIK-br-F 7, ZIK-br-R 10 (1408 bp); 11 - ZIK-br-F 8 ,ZIK-br-R 11 (908 bp); 12 - ZIK-br-F 9, ZIK-br-R 12 (875 bp); 13 - ZIK-br-F 10, ZIK-br-R 13 (897 bp); 14 - ZIK-br-F 11, ZIK-br-R 14 (892 bp); 15 - ZIK-br-F 12, ZIK-br-R 15 (885 bp); 16 - ZIK-br-F 13, ZIK-br-R 16 (897 bp); 17 - ZIK-br-F 13, ZIK-br-R 17 (897 bp); 18 - ZIK-br-F 14, ZIK-br-R 17 (911 bp); 19 - ZIK-br-F 14, ZIK-br-R 18 (1378 bp); 20 - ZIK-br-F 16, ZIK-br-R 19 (892 bp); 21 - ZIK-br-F 17, ZIK-br-R 20 (896 bp); 22 - ZIK-br-F 16, ZIK-br-R 21 (1300 bp); 22 - ZIK-br-F 17, ZIK-br-R 21 (896 bp). The virus was used in this experiment was ZIKV-BR, spiked in ZIKV negative tested serum, The Ct value of virus was 15 and the titration of 5 E10 in PS (porcine stable kidney cell). The PCR products were detected by 1% Agarose Gel electrophoresis, stained with ethidium bromide and observed under ultraviolet light with the condition of 90 volts for 90 minutes TAE buffer 1%.

5.3 First near-full genome amplification.

After a long process of optimizations, for the first time near-full genome was amplified successfully according to the protocols has already described in details (Figure 12).

Figure 12 –The first successful nearly whole genome amplification

From left to right: 1 - 100 pb ladder (Invitrogen), 2 - ZIK-br-F1, ZIK-br-R2 (751bp); 3 - ZIK-br-F2, ZIK-br-R3 (885 bp); 4 - ZIK-br-F2, ZIK-br-R4 (1384 bp); 5 - ZIK-br-F3, ZIK-br-R5 (866bp); 6 - ZIK-br-F4, ZIK-br-R6 (919 bp); 7 - ZIK-br-F5, ZIK-br-R7 (884 bp); 8 - ZIK-br-F6, ZIK-br-R8 (924 bp); 9 - ZIK-br-F7, ZIK-br-R9 (903 bp); 10 - ZIK-br-F7, ZIK-br-R10 (1408 bp); 11 - ZIK-br-F8 , ZIK-br-R11 (908 bp); 12 - ZIK-br-F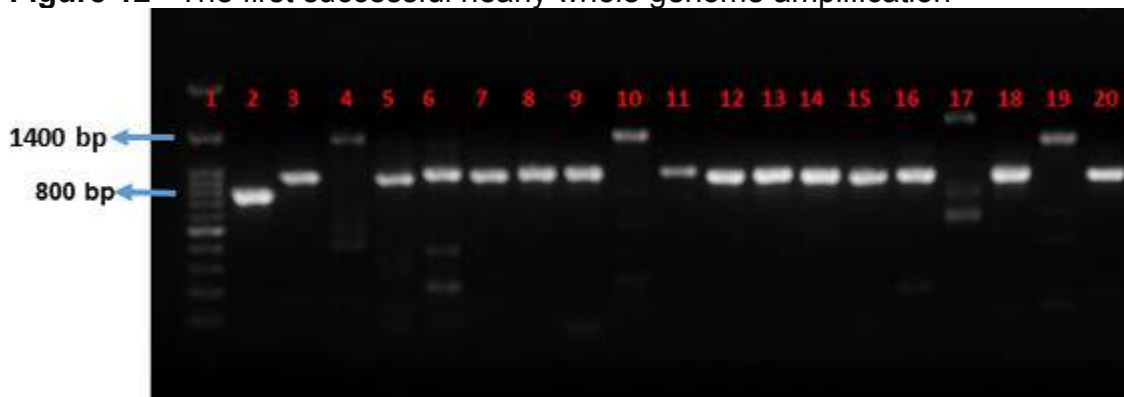 9, ZIK-br-R12 (875 bp); 13 - ZIK-br-F10, ZIK-br-R 13 (897 bp); 14 - ZIK-br-F11, ZIK-br-R14 (892 bp); 15 - ZIK-br-F12, ZIK-br-R15 (885 bp); 16 - ZIK-br-F13, ZIK-br-R16 (897 bp); 17 - ZIK-br-F13, ZIK-br-R17 (1900 bp); 18 - ZIK-br-F14, ZIK-br-R17 (911 bp); 19 - ZIK-br-F14, ZIK-br-R18 (1378 bp); 20 - ZIK-br-F17, ZIK-br-R20 (896 bp). The amplification was successful except 5' UTR. The virus was used in this experiment was ZIKV-BR, Ct value 15, with the titration of 5 E10 in PS (porcine stable kidney cell).

5.4 Amplification of 5' and 3' UTRs

The amplification of 3' UTR was not successful. At this point, we changed our strategy to solve the problem. First, we started to make the cDNA with the Specific primer (Reverse primer for this region) instead of Random primer provided by the High-capacity cDNA reverse transcription kit as previously described. In this case, nothing has changed. Then, we used PCR enhancers such as DMSO (5%), Formamid (5%), Betaine solution (0.8 – 1.6 M) as well as some different combination of forward primer with the same reverse primer. By doing so, we had amplification, but they were unexpected in specific bands. This is normal for such a region because there are repeated regions and also RNA makes secondary structures as previously

mentioned. Then by a simple increase of annealing temperature from 50 C° to 55 C°, and changing the forward primer the problem was solved (Figure 13).

Figure 13 – Different PCR enhancers for UTRs amplification
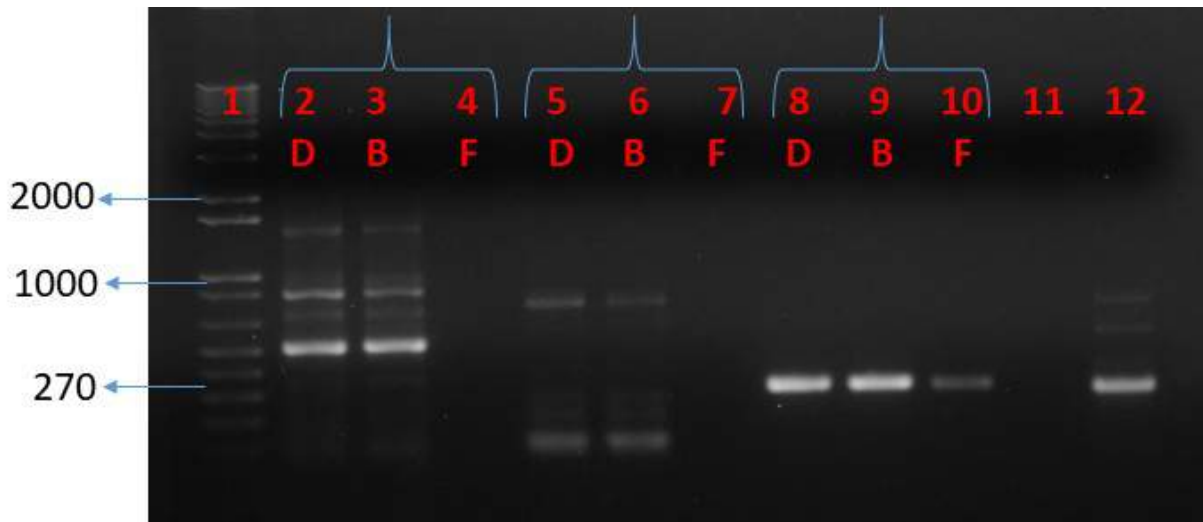
From left to right t: 1 kb Plus DNA ladder (Invitrogen), 2 - 2 - ZIK-br-F15, ZIK-br-R21 + DMSO (5%) - 2300 bp; 3 - ZIK-br-F15, ZIK-br-R21 + Betaine (1M) - 2300 bp; 4 - ZIK-br-F15, ZIK-br-R21 + Formamide (5%) - 2300 bp; 5 - ZIK-br-F17, ZIK-br-R21 + DMSO (5%) – 1300 bp; 6 - ZIK-br-F17, ZIK-br-R21 + Betaine (1M) - 1300 bp); 7 - ZIK-br-F17, ZIK-br-R21 +Formamide (5%)- 1300 bp; 8 - ZIK-br-F18, ZIK-br-R21 + DMSO (5%) - 295 bp; 9 - ZIK-br-F18, ZIK-br-R21 + Betaine (1M) - 295 bp; 10 - ZIK-br-F18, ZIK-br-R21 + Formamide (5%) -295 bp; 11 –Negative control (H2O), 12 – Positive control ZIKV-br-F1, ZIKV-br-R1 (290 bp). The same as before, the virus was used in this experiment was ZIKV-BR, Ct value 15, with the titration of 5 E10 in PS cells.

5.5 Near complete Sanger sequencing

22 PCR product (22 amplicons) were purified utilizing High Pure PCR cleanup kit from Roche, according to the manufacturer's recommendations and were eluted in 15 µL subsequently, 1 µL of purified DNA were quantified with Qubit 3.0 from thermos Fisher Scientific. Although the Kit recommendation is 5 to 20 ng depending on the size, in our experience we had a better result when 20 ng of DNA were submitted to the Sanger sequencing PCR using BigDye Terminator v3.1 (Applied Biosystem, Foster City, California, USA) as previously described in detail. As a result, with three plates the virus provided by the Evandro Chagas Institute (IEC) for scientific research was sequenced in our laboratory. The Blast result showed 100 % identity to the Paraiba-1 isolated. It was assembled from 2 contigs 10600 bp that is a near-full genome of ZIKV (Figure 14).

The base calling, assembly and trimming was done by DNA Sequence Assembler v4.36.0 (2013), from Heracle BioSoft, available at www.DnaBaser.com. It was optimized for normal sample. For automatic end trimming, it trims until there are more than 60 % good bases in a 16 bases window. Also base calling requires a quality of Phred higher than 20 in order to be considered good (trusted). The sequence assembler of the software considers 20-word size (the default) and 75 sample identity percent (default 80%), also minimum overlap of 25 bases, which is the default of the software. For automatic ambiguity correction, bases with Quality equal or higher than 22 are considered trusted. The assembly and chromatogram can be observed in Figure 15, 16.

Figure 14 – Short amplicons of IEC sequenced by Sanger

From left to right: 1 - 100 pb ladder (Invitrogen), 2 - ZIK-br-F1, ZIK-br-R2 (751bp); 3 - ZIK-br-F2, ZIK-br-R3 (885 bp); 4 - ZIK-br-F2, ZIK-br-R4 (1384 bp); 5 - ZIK-br-F3, ZIK-br-R5 (866bp); 6 - ZIK-br-F4, ZIK-br-R6 (919 bp); 7 - ZIK-br-F5, ZIK-br-R7 (884 bp); 8 - ZIK-br-F6, ZIK-br-R8 (924 bp); 9 - ZIK-br-F7, ZIK-br-R9 (903 bp); 10 - ZIK-br-F7, ZIK-br-R10 (1408 bp); 11 - ZIK-br-F8 ,ZIK-br-R11 (908 bp); 12 - ZIK-br-F 9, ZIK-br-R12 (875 bp); 13 - ZIK-br-F10, ZIK-br-R 13 (897 bp); 14 - ZIK-br-F11, ZIK-br-R14 (892 bp); 15 - ZIK-br-F12, ZIK-br-R15 (885 bp); 16 - ZIK-br-F13, ZIK-br-R16 (897 bp); 17 - ZIK-br-F13, ZIK-br-R17 (1900 bp); 18 - ZIK-br-F14, ZIK-br-R17 (911 bp); 19 - ZIK-br-F14, ZIK-br-R18 (1378 bp); 20 - ZIK-br-F16, ZIK-br-R19 (892 bp); 21 - ZIK-br-F17, ZIK-br-R20 (896 bp); 22 - ZIK-br-F17, ZIK-br-R21 (1300 bp) and 23 - Positive control consist of ZIKV-br-F1 and ZIKV-br-R1 (290 bp). The PCR products were detected by 1% Agarose Gel electrophoresis, stained with ethidium bromide and observed under ultraviolet light with the condition of 90 volts for 90 minutes TAE buffer 1%. Position based on KU940228.

Figure 15 – Schematic of genome assembly for Sanger sequencing

The assembly consists of different plate wells that each one has different forward or reverse primer. Then the software makes an assembly from each one, resulting the consensus sequence.

Figure 16 – Base quality of the sequences and chromatogram
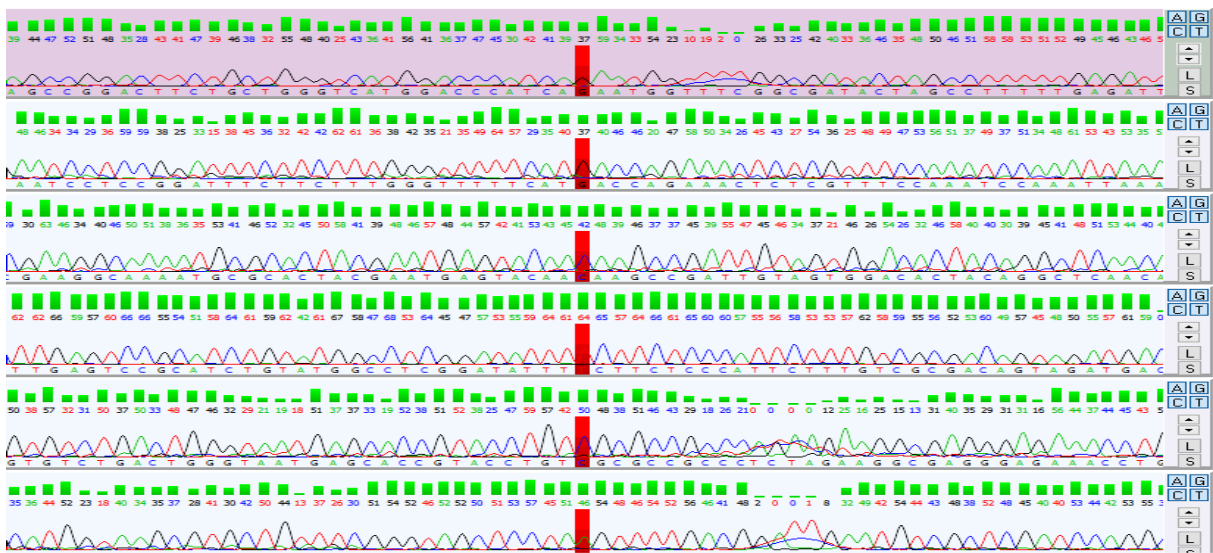
The chromatogram shows the quality of each base with the number written at the bottom of each one

5.6 Next generation sequencing

5.6.1 Sequencing of IEC isolated

The same PCR products, which were already purified (High Pure PCR cleanup kit from Roche) and also sequenced by Sanger (Figure 11), were undergoing the New Generation Sequencing (Illumina platform). All twenty-two amplicons were quantified by Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). Then, were normalized to the concentration of 1 ng then they were polled altogether. A library was prepared using the NEXTERA XT DNA library preparation kit as was described previously. Afterward, the quality of the library was analyzed by 2100 Bioanalyzer instrument from Agilent with the DNA microfluidic chip to check the average size of the sequences (Figure 13)

Figure 17 – Bioanalyzer result of short amplicons library of IEC
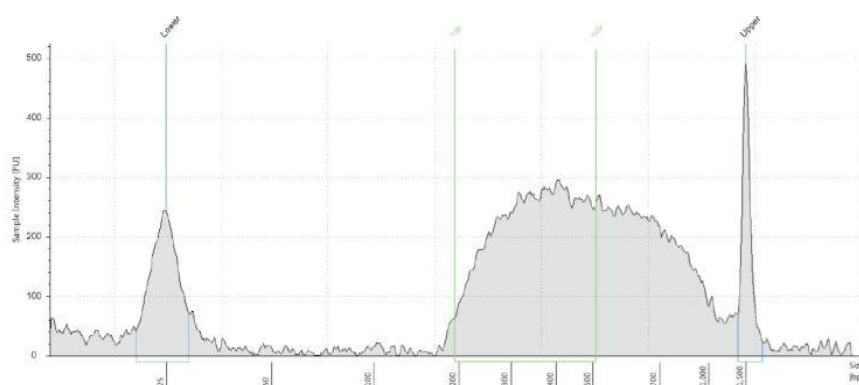
Region Table

From [bp]	To [bp]	Average Size [bp]	Conc. [pg/µl]	Region Molarity [pmol/l]	% of Total	Region Comment	Color
196	509	357	1630	7510	49.38		

The Average size of the constructed library is 357 that is optimized for Nextera Xt library preparation kit.

Finally, sequencing was done using the NextSeq sequencing reagents (Illumina), which the output range of 16.25 – 120 GB, 130 – 400 million reads per run, 2 x 150 bp maximum read length. The library was normalized to the final concentration of 2 nM, 300 cycles. We then added 5% of Phix and 1.8 pmol of the final concentration of library for cluster generation.

Figure 18 – Coverage of short amplicons sequenced by Illumina
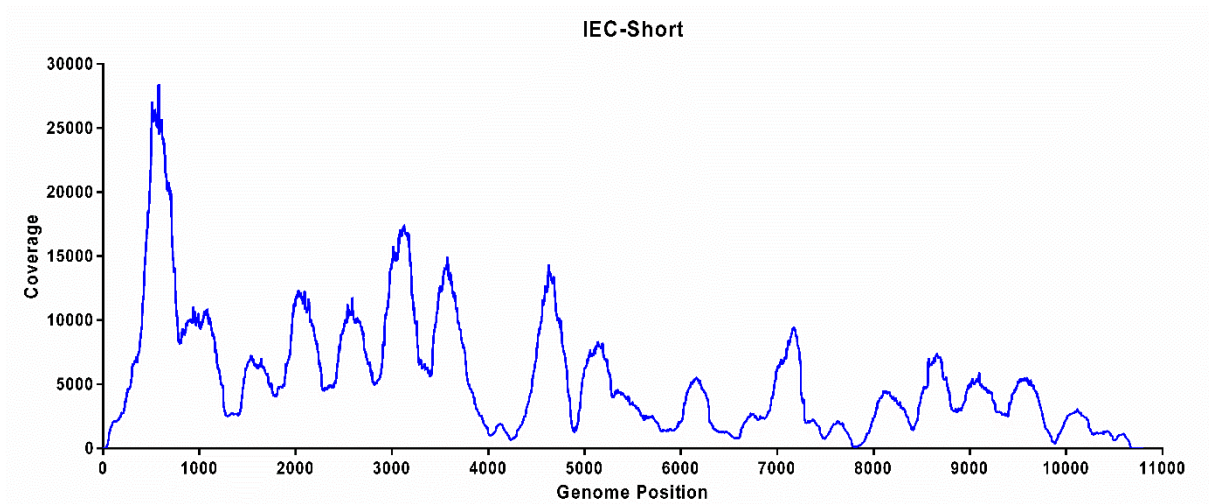
Total reads 602,263, after trimming 397,116, medium coverage 5,457 and it covers 99.11 % when was aligned and mapped to reference genome (Accession number of KX280026).

The long amplicons were produced by the protocol already described and submitted with the same conditions as short amplicons to the Illumina platform. The result of the agarose electrophoresis and also NGS coverage can be observed in figures 19..

**Figure 19 –**Long amplicons of IEC sequenced by Illumina
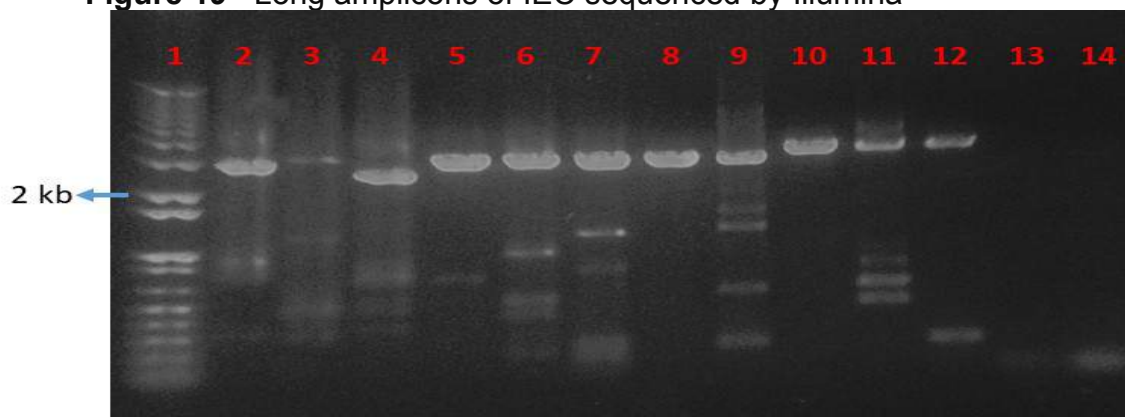
From left to right: 1 - 1 Kb Plus ladder (Invitrogen), 2 - ZIK-br-F1, ZIK-br-R6 (2781bp); 3 - ZIK-br-F2, ZIK-br-R7 (2895 bp); 4 - ZIK-br-F3, ZIK-br-R8 (2385 bp); 5 - ZIK-br-F5, ZIK-br-R11 (2885bp); 6 - ZIK-br-F6, ZIK-br-R12 (2917 bp); 7 - ZIK-br-F7, ZIK-br-R13 (2900 bp); 8 - ZIK-br-F8, ZIK-br-R15 (2911 bp); 9 - ZIK-br-F9, ZIK-br-R16 (2864 bp); 10 - ZIK-br-F10, ZIK-br-R17 (3403 bp); 11 - ZIK-br-F11 ,ZIK-br-R18 (3364 bp); 12 - ZIK-br-F12, ZIK-br-R19 (3384 bp); 13 - ZIK-br-F13, ZIK-br-R20 (3398 bp); 14 - ZIK-br-F14, ZIK-br-R21 (2776 bp). The PCR products were detected by 0.8% Agarose Gel electrophoresis, stained with ethidium bromide and observed under ultraviolet light with the condition of 90 volts for 100 minutes TAE buffer 1%.

As can be observed in Figure 8, the long amplicons made a long series of overlaps so it was not necessary to sequence all these amplicons. Moreover, the objective of having long amplicons is to be able to make the whole genome sequencing just with 6 of these amplicons. Therefore, just 6 of them were chosen and submitted to the Illumina platform (lane numbers 1, 4, 7, 10 and 13). Bioanalyzer result and genome coverage are presented by the figure 20.

Figure 20 – Bioanalyzer result of Long amplicons library of IEC

Region Table

From [bp]	To [bp]	Average Size [bp]	Conc. [pg/µl]	Region Molarity [pmol/l]	% of Total	Region Comment	Color
211	528	379	906	3970	42.10		

DNA size of the prepared library ranged from 211 to 528 bp with the average size of 379 bp. In both cases of short and long amplicons there were not observed any difference in size after library preparation and also tagmentation by the library preparation kit.

Figure 21 – Long amplicon NGS sequencing coverage of IEC
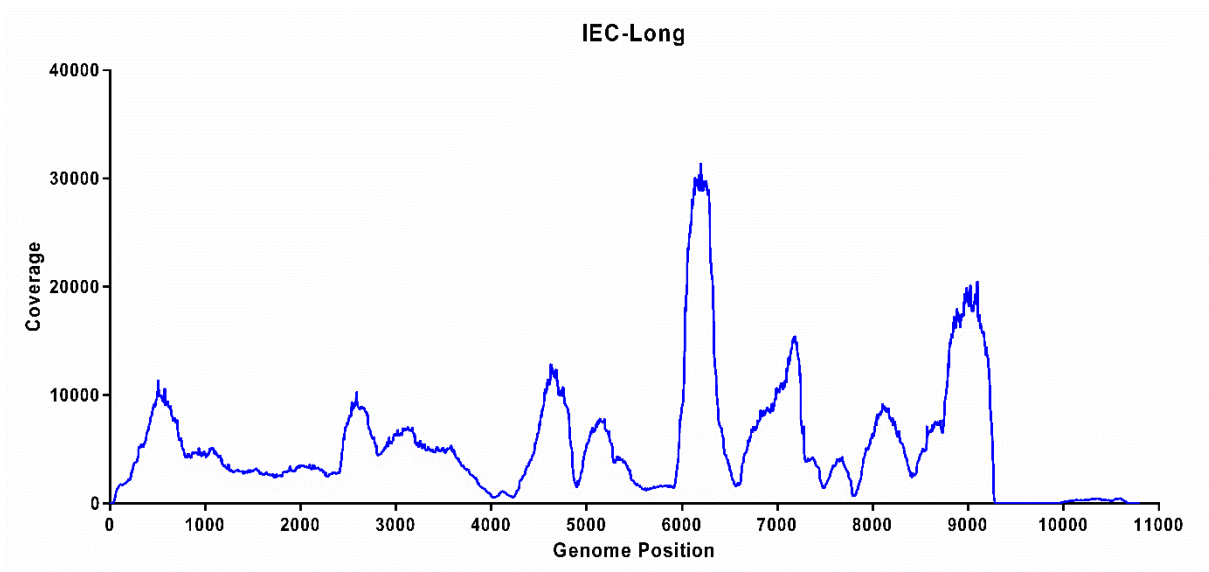
5.6.2 Next Generation Sequencing of P17, S007

Extraction, cDNA synthesis, PCR, qRT-PCR were done as previously described in detail (just short amplicons strategy) from the different time point samples collected from the same patient (Table 4). According to the result of RT-PCR they had different Ct value during two-month sample collection. These clinical samples were enriched by PCR prior to the Next Generation Sequencing by Illumina platform. Therefore, we could show the effect of Ct value of the sample collected from the same individual.

Table 4 – The samples were sequenced by NGS in this study

Name	Material	Collection date	Ct value
P17-A	Sperm	26/04/2016	19
P17-C	Sperm	10/05/2016	23
P17-H	Sperm	16/06/2016	24
P17-K	Sperm	05/07/2016	26
P17-S	Sperm	01/09/2016	33
P17-T	Sperm	06/09/2016	34

The sequencing of P17-T and P17-S were not successful as expected due to high Ct value. P17-S were sequenced, but unfortunately were not successful, So P17-T were eliminated from this study.

According to the absolute quantification curve, samples with Ct values of more than 33, would have approximately around 10,000 viruses/mL that would make the task almost impossible to amplify and sequenced them with the consideration of the volume is used in the extraction. Also, we should take into the consideration the fact of limitation of the samples in a case of clinical samples. According to our previous experience, even direct sequencing from clinical sample is possible with Ct value less than 24. For Cts above 24, we have to do some enrichment that can be cloning, PCR or other technic. The results of PCR and sequencing can be observed in Figures 18 and 19.
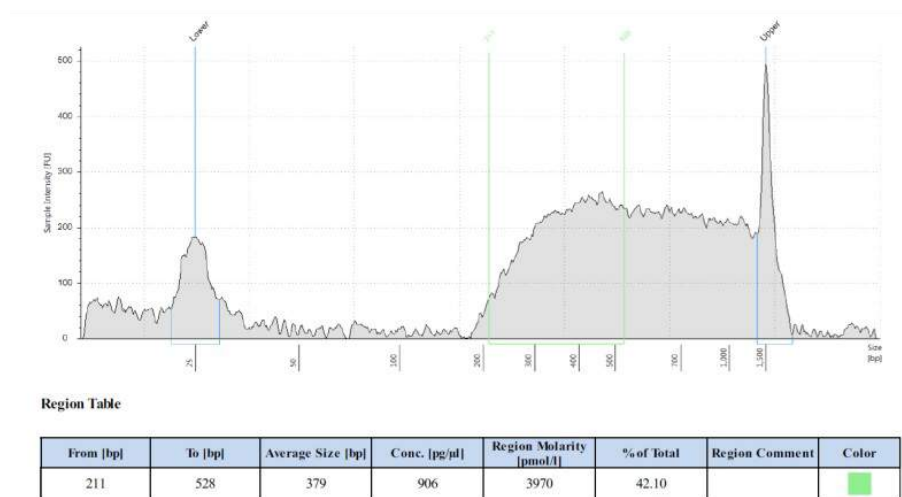
Figure 22 – P17 agarose electrophoresis

From left to right: 1 - 100 pb ladder (Invitrogen), 2 - ZIK-br-F1, ZIK-br-R 2 (751bp); 3 - ZIK-br-F 2, ZIK-br-R 3 (885 bp); 4 - ZIK-br-F 2, ZIK-br-R4 (1384 bp); 5 - ZIK-br-F 3, ZIK-br-R 5 (866bp); 6 - ZIK-br-F 4, ZIK-br-R 6 (919 bp); 7 - ZIK-br-F 5, ZIK-br-R 7 (884 bp); 8 - ZIK-br-F 6, ZIK-br-R 8 (924 bp); 9 - ZIK-br-F 7, ZIK-br-R 9 (903 bp); 10 - ZIK-br-F 7, ZIK-br-R 10 (1408 bp); 11 - ZIK-br-F 8 ,ZIK-br-R 11 (908 bp); 12 - ZIK-br-F 9, ZIK-br-R 12 (875 bp); 13 - ZIK-br-F 10, ZIK-br-R 13 (897 bp); 14 - ZIK-br-F 11, ZIK-br-R 14 (892 bp); 15 - ZIK-br-F 12, ZIK-br-R 15 (885 bp); 16 - ZIK-br-F 13, ZIK-br-R 16 (897 bp); 17 - ZIK-br-F 13, ZIK-br-R 17 (897 bp); 18 - ZIK-br-F 14, ZIK-br-R 17 (911 bp); 19 - ZIK-br-F 14, ZIK-br-R 18 (1378 bp); 20 - ZIK-br-F 16, ZIK-br-R 19 (892 bp); 21 - ZIK-br-F 17, ZIK-br-R 20 (896 bp); 22 - ZIK-br-F 16, ZIK-br-R 21 (1300 bp); 22 - ZIK-br-F 17, ZIK-br-R 21 (896 bp); 23 - ZIK-br-F 18, ZIK-br-R 21 (295 bp) and 24- ZIK-br-F1, ZIK-br-R1 (281bp) as a positive control. The PCR products were detected by 1% Agarose Gel electrophoresis, stained with ethidium bromide and observed under ultraviolet light with the condition of 90 volts for 90 minutes TAE buffer 1%.

Figure 23 – Coverage distribution of P17along side of its genome
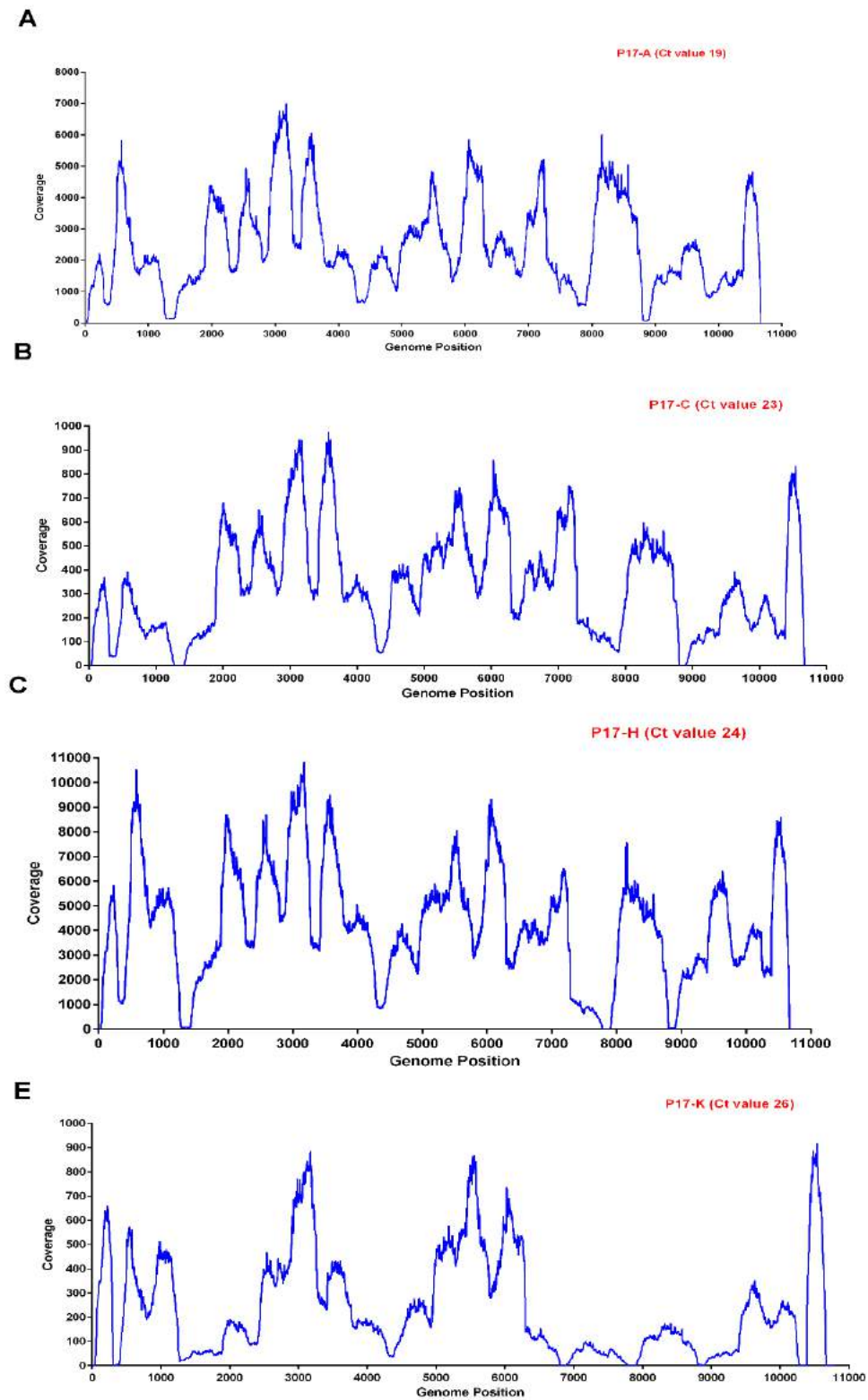
Figure 24 – Normalized coverage distribution of P17 (0 – 1)
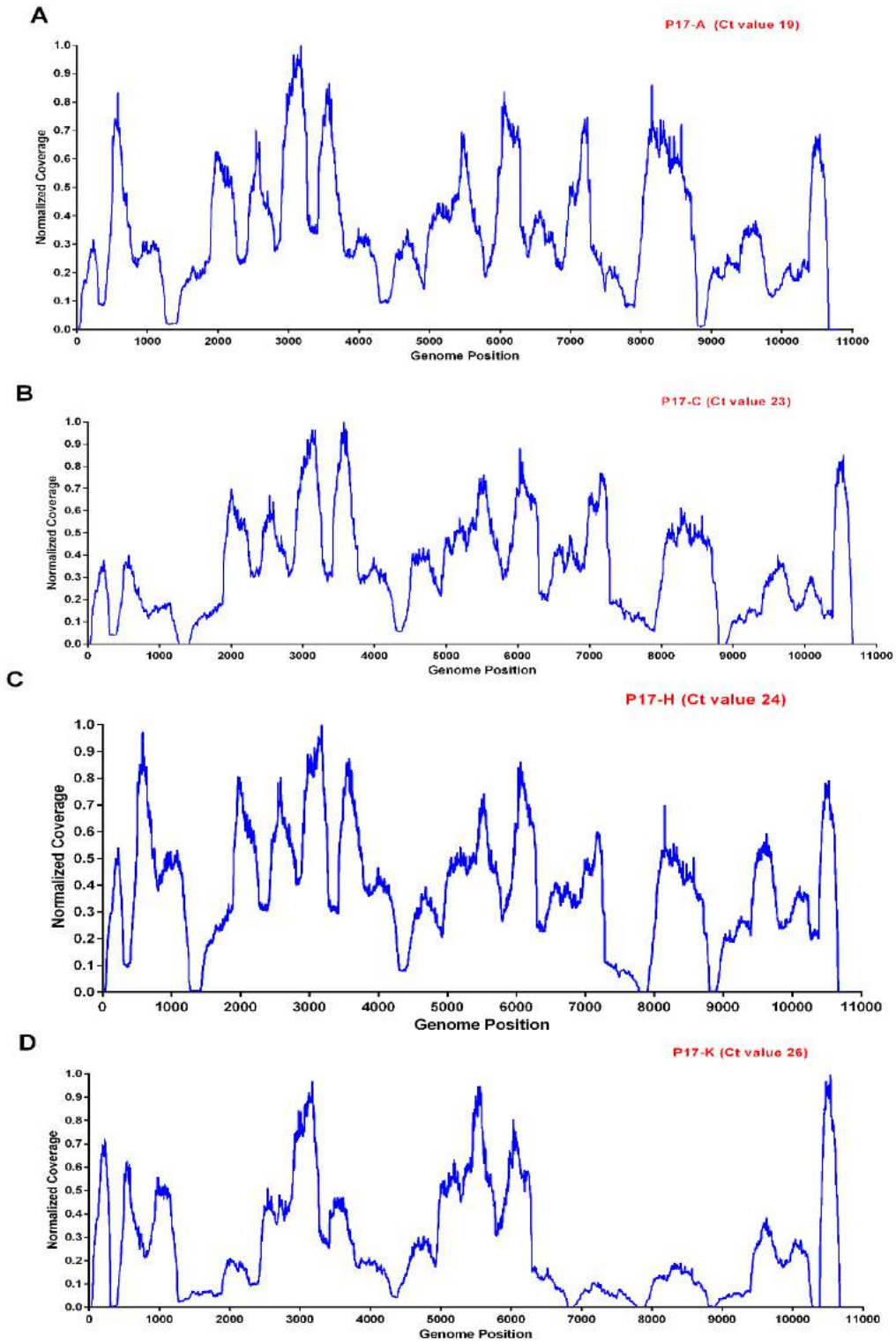
The sample S007 was sequenced even without amplification detectable with ethidium bromide in agarose. The results unfortunately were not successful, yielding a short sequence of the virus.

Table 5 summarizes the results obtained during the work presented in this project. It can be observed the average of coverage, maximum coverage, and the percent of genome coverage in the case of mapping against a reference genome. For the De novo assembly the number of contigs, the largest contig, the present of the desirable target genome when all the contigs were checked by Blast and finally the numbers of amplicons utilized for amplifying each sample. The sample S was not successfully mapped to the genome reference used due to the high Ct value, therefore in the case of the De Novo Assembly resulted a contig with 672 sequences with the 0 % similarity to our target (ZIKV). As we mentioned before, the sample S007 unfortunately was not proceeded successfully, neither in the case of mapping with reference genome, nor in the case of De Novo assembly. Due to the result of the sample P17-S, we did not proceed with the sample P17-T because the higher Ct value in comparison with the sample P17-S.

Table 5 – General aspect of all the samples sequenced by Illumina

Sample	Ct Value	Collected	Medium Cov x	Max Cov	Genome Cov (%)	Contigs	Largest	%ZIKV	Amplicons
P17-A	19	26/04/2016	2462	7004	98.45	351	2726	24.8	23
P17-C	23	10/05/2016	338	977	96.11	122	7842	2.47	23
P17-H	24	16/06/2016	4188	10832	99.27	562	5038	3.21	23
P17-K	26	07/05/2016	238	917	95.43	196	6519	3.57	23
P17-S	34	NA	NA	NA	NA	672	2131	0	23
P17-T	33	NA	NA	NA	NA	NA	NA	NA	23
S007	34	1/9/14	NA	NA	NA	576	2706	0	23
IEC-short	16	10/02/16	3267	28445	99.04	5	10744	40	23
IEC-long	16	10/02/16	3231	21506	98.00	102	1856	98	6

6. Discussion

Along the whole process from sample to the end there are some factors that can play a fundamental role and therefore change the final results. We will now discuss some of them in details.

6.1 Choice of clinical samples.

It is critical to do an early sample collection because usually, the symptoms appear 4-7 days after the onset with the higher titer of a viral copy in body fluid. However, following the appearance of symptoms viral load decreases. ZIKV is detectable in saliva and urine more than in the blood. In urine, viral load is higher than blood with a peak at around 5 to 7 days but is not detectable for more than 20 days after the clinical onset (72). ZIKV RNA was detectable in nasopharyngeal swabs while negative in serum (73, 74).

6.2 PCR inhibitors.

The effect of inhibitors in PCR reactions could be considered, especially in fresh urine, EDTA containing Vacutainers and in amplifications involving a low amount of viral copy or degraded genome. PCR inhibitors can be divided into two groups: organic and inorganic. Organic compounds, including bile salts, urea, phenol, ethanol, polysaccharides, sodium dodecyl sulfate (SDS), humic acids, tannic acid, melanin, different proteins, such as collagen, myoglobin, hemoglobin, lactoferrin, immunoglobulin G (IgG), Heparin, also proteinases (75, 76). Sometimes sample dilution can be a solution to reduce the effect of PCR inhibitors, but not in our case with a low amount of viral RNA (77, 78). Inorganics (*e.g.*, salts and metals) are less problematic in virus-containing samples subjected to nucleic acids extraction protocols we used. However, most of the known inhibitors are organic compounds.

6.3 Enhancers for better amplifications.

There are some enhancers and additives that can be added to cDNA and PCR reaction mixes, which increase efficiency and sensibility, improving the enzymatic activity of both, reverse transcriptase (RT) and DNA polymerase. In the case of an extremely low amount of viral genome amplification, these can be helpful. For instance, using 3.5 to 0.1 M of Betaine reduces Tm (melting temperature) facilitating GC-rich region amplification. BSA (bovine serum albumin) in a concentration of 0.01 µg/µL to 0.1 µg/µL is useful when attempting to amplify templates that contain PCR inhibitors, such as melanin. 2-10% of DMSO (dimethyl sulfoxide) reduces secondary structure and is particularly beneficial for better processing of GC-rich templates. But when used in concentrations above 10%, it will reduce polymerase activity. Formamide reduces also secondary structure and GC-rich templates. Non-ionic detergents such as Tween 20 also stabilize Taq polymerase and may also suppress the formation of the secondary structure of the concentration of 0.1-1%.

Moreover, choosing which cDNA synthesis kit should be used is also important. According to our own observation, we got better results when we make cDNA with the SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen) rather than using High-Capacity cDNA Reverse Transcription Kits (Applied Biosystem).

6.4 cDNA synthesize

Sometimes just a gene or specific region is subject to be amplified and not the whole genome. In this case, synthesizing the cDNA with specific reverse primer can be a solution for low input RNA can be convenient (79). Nested or Semi-nested PCR also can be favorable for better detection and amplification of low input genomic material.

Finally, sequencing DNA rather than RNA needs less effort, is less time consuming and reduces the cost noticeably. Moreover, by the implementation of these techniques we are allowed to increase the amount of genomic material that can be sequenced by Sanger sequencing, which is still routinely used especially in

the case of specific genes, genotyping, detection of SNV (single nucleotide variation), biological amplification by culture and cloning. Furthermore, it can be useful for the confirmation of PCR positive material, determine the source of an outbreak rapidly, understanding the molecular evolution of emerging viruses.

7. Future work

We gained experience during this project that was useful for designing a set of primers for African ZIKV one set for West of Africa and Nigeria and another set for Central Africa and Uganda. All the primers are available in our laboratory and they are in the optimization process. Most of the researchers are working with MR766 as a control or in comparison of African and Asian, which is circulating in Brazil, North America and recently Cabo Verde. There are some complete sequences of MR766 available in NCBI that are so different. Therefore, it is time to sequence this one and more African ZIKV in the context of our ongoing comparison of biological properties between low and high passage MR766 isolates. Further comparisons among ZIKV will be undertaken as part of the USP-IP-FIOCRUZ project in progress and the ongoing CNPq project #441105/2016-5.

8. References*

1. Dick GWA. Zika Virus (I). Isolations and serological specificity. Trans R Soc Trop Med Hyg. 1952; 46(5):509–20.
2. Dick GW. Zika virus (II). Pathogenicity and physical properties. Trans R Soc Trop Med Hyg. 1952; 46(5):521–34.
3. Smithburn KC. Neutralizing antibodies against certain recently isolated viruses in the sera of human beings residing in East Africa. J Immunol [Internet]. 1952; 69(2):223–34.
4. MacNamara FN. Zika virus: A report on three cases of human infection during an epidemic of jaundice in Nigeria. Trans R Soc Trop Med Hyg. 1954; 48(2):139–45.
5. Haddow AD, Schuh AJ, Yasuda CY, Kasper MR, Heang V, Huy R, et al. Genetic characterization of Zika virus strains: Geographic expansion of the asian lineage. PLoS Negl Trop Dis. 2012;6(2).
6. Faye O, Freire CCM, Iamarino A, Faye O, de Oliveira JVC, Diallo M, et al. Molecular Evolution of Zika Virus during Its Emergence in the 20th Century. PLoS Negl Trop Dis. 2014; 8(1):36.
7. Olson JG, Ksiazek TG, Suhandiman G, Triwibowo V. Zika virus, a cause of fever in central java, indonesia. Trans R Soc Trop Med Hyg. 1981; 75(3):389–93.
8. Kool JL, D P, Lanciotti RS, Pretrick M, Dubray C, Guillaumot L, et al. Zika Virus Outbreak on Yap Island, Federated States of Micronesia. N Engl J Med. 2009;2536–43.
9. Foy BD, Kobylinski KC, Foy JL, Blitvich BJ, Travassos da Rosa A, Haddow AD, et al. Probable Non–Vector-borne Transmission of Zika Virus, Colorado, USA. Emerg Infect Dis. 2011; 17(5):880-882. https://dx.doi.org/10.3201/eid1705.101939.
10. Roth A, Mercier A, Lepers C, Hoy D, Duituturaga S, Benyon E, et al. Concurrent outbreaks of dengue, chikungunya and Zika virus infections An unprecedented epidemic wave of mosquito-borne viruses in the Pacific 2012-2014. Eurosurveillance. 2014; 19(41):1–8.
11. Cao-Lormeau V-M, Musso D. Emerging arboviruses in the Pacific. Lancet. 2014; 384(9954):1571–2. Available from: http://dx.doi.org/10.1016/S0140-6736 (14)61977-2.
12. Besnard M, Lastère S, Teissier A, Cao-Lormeau VM, Musso D. Evidence of perinatal transmission of Zika virus, French Polynesia, December 2013 and February 2014. Eurosurveillance. 2014; 19(13):8–11.
13. WHO epidemical alert,Zika virus infection. 7 May 2015. Available from http://www.paho.org/hq/index.php?option=com_docman&task=doc_view&Itemid=270&gid=30075=en%20%28accessed%2002%20Feb%202016%29.
14. WHO emergencies Zika causality statement. 7 September 2016. Available from http://www.who.int/emergencies/Zika-virus/causality/en/
15. WHO media center. Fifth meeting of the Emergency Committee under the International Health Regulations (2005) regarding microcephaly, other neurological disorders and Zika virus, 18 November 2016.
16. CDC.2016 Web site. Division of vector born infectious diseases, arboviral diseases branch. Available at: www.cdc.gov/ncidod/dvbid/arbor/.

17. WHO. WHO Research and Development Blueprint: 2017 Annual review of diseases prioritized under the Research and Development Blueprint. 2017 ;(January):16. Available from: http://www.who.int/blueprint/what/research-development/2017-Prioritization-Long-Report.pdf?ua=1

18. ICTV. International comittee on taxonomy of viruses. Virus taxonomy. 2014. Release 2015, http://www.ictvonline.org/virustaxonomy.asp [accessed 22.12.16].

19. Kuno G, Chang GJ, Tsuchiya KR, Karabatsos N, Cropp CB. Phylogeny of the genus Flavivirus. J Virol [Internet]. 1998; 72(1):73–83. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9420202%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC109351

20. Pierson TC, Diamond MS. Flaviviruses. In: Knipe DM, Howley PM, editors. Fields virol. 6th ed. Philadelphia: Lippincott Williams & Will- kins; 2013. p. 747e94.

21. Avirutnan P, Fuchs A, Hauhart RE, Somnuke P, Youn S, Diamond MS, Atkinson JP: Antagonism of the complement component C4 by flavivirus nonstructural protein NS1. J Exp Med 2010, 207:793-806.

22. Donald CL, Brennan B, Cumberworth SL, Rezelj V V., Clark JJ, Cordeiro MT, et al. Full Genome Sequence and sfRNA Interferon Antagonist Activity of Zika Virus from Recife, Brazil. PLoS Negl Trop Dis [Internet]. 2016; 10(10): e0005048. Available from: http://dx.plos.org/10.1371/journal.pntd.0005048.

23. Funk A, Truong K, Nagasaki T, Torres S, Floden N, Balmori Melian E, et al. RNA structures required for production of subgenomic flavivirus RNA. J Virol. 2010; 84(21):11407–17.

24. Lindenbach BD, Murray CL, Thiel H-J, Rice CM: Flaviviridae: the viruses and their replication. In Fields Virology, edn 6. Edited by Knipe DM et al.: New York, NY: Lippincott Williams & Wilkins; 2013.

25. Hamel R, Dejarnac O, Wichit S, Ekchariyawat P, Neyret A, Luplertlop N, et al. Biology of Zika Virus Infection in Human Skin Cells. J Virol. 2015 Sep 1; 89(17):8880–96.

26. Assenberg R, Mastrangelo E, Walter TS, Verma A, Milani M, Owens RJ, et al. Crystal structure of a novel onformational state of the flavivirus NS3 protein: implications for polyprotein processing and viral replication. J Virol [Internet]. 2009; 83(24):12895–906.

27. Meertens L, Carnec X, Lecoin MP, Ramdasi R, Guivel-Benhassine F, Lew E, et al. The TIM and TAM Families of Phosphatidylserine Receptors Mediate Dengue Virus Entry. Cell Host Microbe. 2012 Oct 18; 12(4):544–57.

28. Nowakowski TJ, Pollen AA, Di Lullo E, Sandoval-Espinosa C, Bershteyn M, Kriegstein AR. Expression Analysis Highlights AXL as a Candidate Zika Virus Entry Receptor in Neural Stem Cells. Cell Stem Cell. 2016 May 5; 18(5):591–6.

29. Galán-Huerta KA, Rivas-Estilla AM, Martinez-Landeros EA, Arellanos-Soto D, Ramos-Jiménez J. The Zika virus disease: An overview. Med Univ [Internet]. 2016; 18(71):115–24. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1665579616300588

30. Villordo SM, Gamarnik A V. Genome cyclization as strategy for flavivirus RNA replication. Virus Res. 2009; 139(2):230–9.

31. Alvarez DE, De Lella Ezcurra AL, Fucito S, Gamarnik A V. Role of RNA structures present at the 3′UTR of dengue virus on translation, RNA synthesis, and viral replication. Virology. 2005; 339(2):200–12.

32. Khromykh A a, Meka H, Guyatt KJ, Westaway EG. Essential Role of Cyclization Sequences in Flavivirus RNA Replication Essential Role of Cyclization Sequences in Flavivirus RNA Replication †. J Virol. 2001; 75(14):6719–28.

33. Brinton MA, Dispoto JH. Sequence and secondary structure analysis of the 5′-terminal region of flavivirus genome RNA. Virology. 1988; 162(2):290–9.
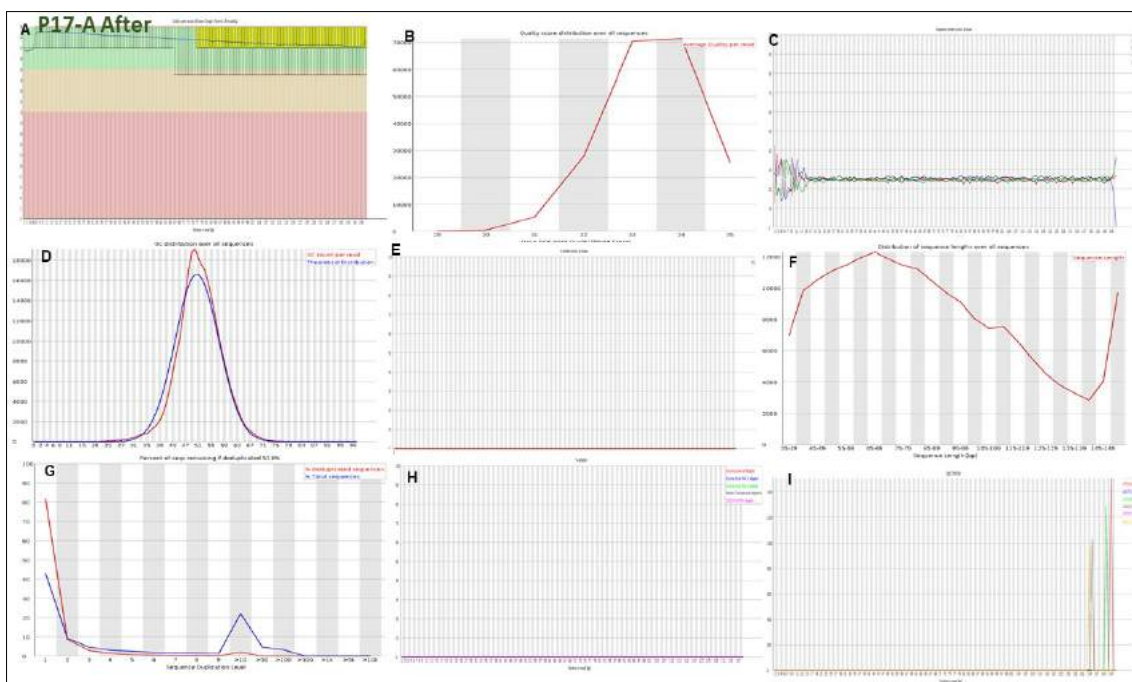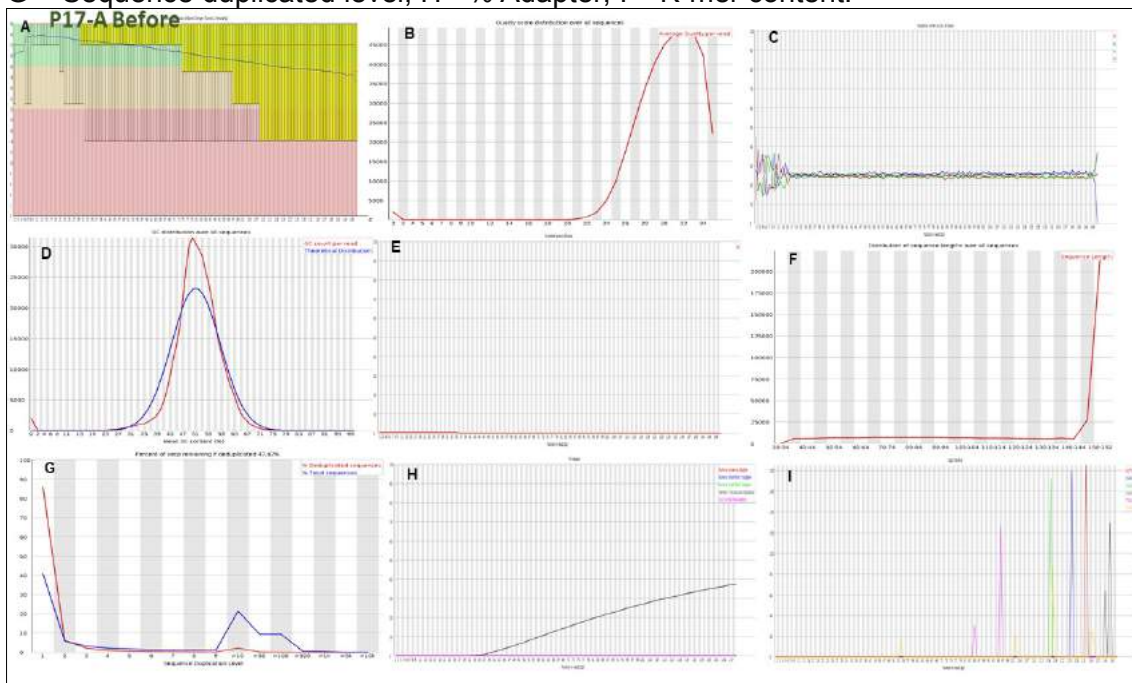
34. Hamel R, Dejarnac O, Wichit S, Ekchariyawat P, Neyret A, Luplertlop N, et al. Biology of Zika Virus Infection in Human Skin Cells. J Virol. 2015 Sep 1; 89 (17): 8880–96.

35. Meertens L, Carnec X, Lecoin MP, Ramdasi R, Guivel-Benhassine F, Lew E, et al. The TIM and TAM Families of Phosphatidylserine Receptors Mediate Dengue Virus Entry. Cell Host Microbe. 2012 Oct 18; 12 (4): 544–57.

36. Savidis G, McDougall WM, Meraner P, Perreira JM, Portmann JM, Trincucci G, et al. Identification of Zika Virus and Dengue Virus Dependency Factors using Functional Genomics. Cell Rep. 2016 Jun 28; 16 (1): 232–46.

37. Nowakowski TJ, Pollen AA, Di Lullo E, Sandoval-Espinosa C, Bershteyn M, Kriegstein AR. Expression Analysis Highlights AXL as a Candidate Zika Virus Entry Receptor in Neural Stem Cells. Cell Stem Cell. 2016 May 5; 18 (5): 591–6.

38. Musso D, Roche C, Nhan T-X, Robin E, Teissier A, Cao-Lormeau V-M. Detection of Zika virus in saliva. J Clin Virol [Internet]. 2015; 68:53–5.

39. Stiasny K, Fritz R, Pangerl K, Heinz FX. Molecular mechanisms of flavivirus membrane fusion. Amino Acids. 2011;41(5):1159–63.

40. Vázquez-Calvo Á, Saiz JC, McCullough KC, Sobrino F, Martín-Acebes MA. Acid-dependent viral entry. Virus Res [Internet]. 2012;167(2):125–37.

41. Brasil P, Sequeira PC, Freitas ADA, Zogbi HE, Calvet GA, Souza RV De, et al. Guillain-Barré syndrome associated with Zika virus infection. Lancet [Internet]. 2016; 387(10026):1482. Available from: http://dx.doi.org/10.1016/S0140-6736 (16)30058-7

42. Lanciotti RS, Lambert AJ, Holodniy M, Saavedra S, del Carmen Castillo Signor L. Phylogeny of zika virus in western Hemisphere, 2015. Emerg Infect Dis. 2016; 22(5):933–5.

43. Pyke AT, Moore PR, Hall-mendelin S, Mcmahon JL. Isolation of Zika Virus Imported from Tonga into. 2018;7–12.

44. Bonaldo MC, Ribeiro IP, Lima NS, dos Santos AAC, Menezes LSR, da Cruz SOD, et al. Isolation of Infective Zika Virus from Urine and Saliva of Patients in Brazil. PLoS Negl Trop Dis. 2016;10(6):1–17.

45. Barzon L, Pacenti M, Berto A, Sinigaglia A, Franchin E, Lavezzo E, et al. Isolation of infectious Zika virus from saliva and prolonged viral RNA shedding in a traveller returning from the dominican republic to Italy, January 2016. Eurosurveillance. 2016; 21(10):1–5.

46. Harrower J, Kiedrzynski T, Baker S, Upton A, Rahnama F, Sherwood J, et al. Sexual Transmission of Zika Virus and Persistence in Semen, New Zealand, 2016. Emerg Infect Dis [Internet]. 2016; 22(10):1855–7.

47. Nicastri E, Castilletti C, Liuzzi G, Iannetta M, Capobianchi MR, Ippolito G. Persistent detection of Zika virus RNA in semen for six months after symptom onset in a traveller returning from Haiti to Italy, February 2016. Eurosurveillance. 2016;21(32):1–4

48. Bandeira AC, Campos GS, Rocha VFD, Souza BS de F, Soares MBP, Oliveira AA, et al. Prolonged shedding of Chikungunya virus in semen and urine: A new perspective for diagnosis and implications for transmission. ID Cases [Internet]. 2016; 6:100–3.

49. Miner JJ, Sene A, Richner JM, Ebel GD, Diamond MS, Apte RS, et al. Zika Virus Infection in Mice Causes Panuveitis with Shedding of Virus in Tears. Cell Rep [Internet]. 2016; 16:1–11.

50. Lessler J, Chaisson LH, Kucirka LM, Bi Q, Grantz K, Salje H, et al. Assessing the Global Threat from Zika Virus. Science

51. Yoon H, Leitner T. 2014. PrimerDesign-M: a multiple-alignment based multiple-primer design tool for walking across variable genomes. Bioinformatics, Dec. 17. PMID: 25524896

52. Brodin J, Krishnamoorthy M, Athreya G, Fischer W, Hraber P, Gleasner C, Green L, Korber B, Leitner T. 2013. A multiple-alignment based primer design algorithm for genetically variable DNA targets. BMC Bioinformatics 14:255.
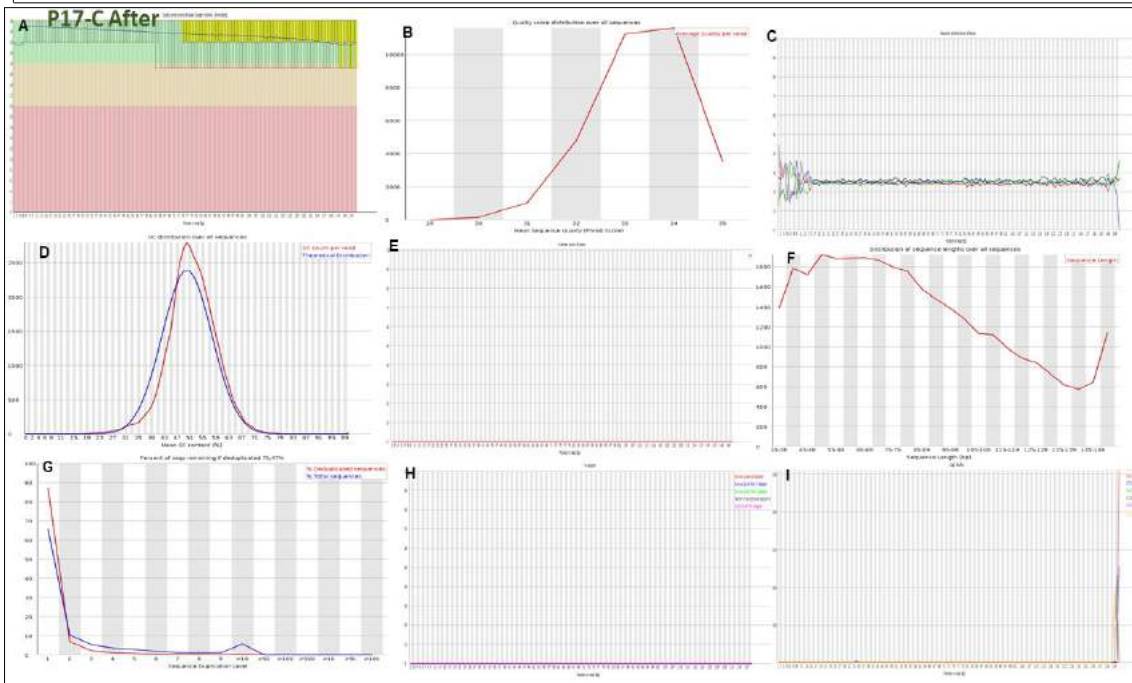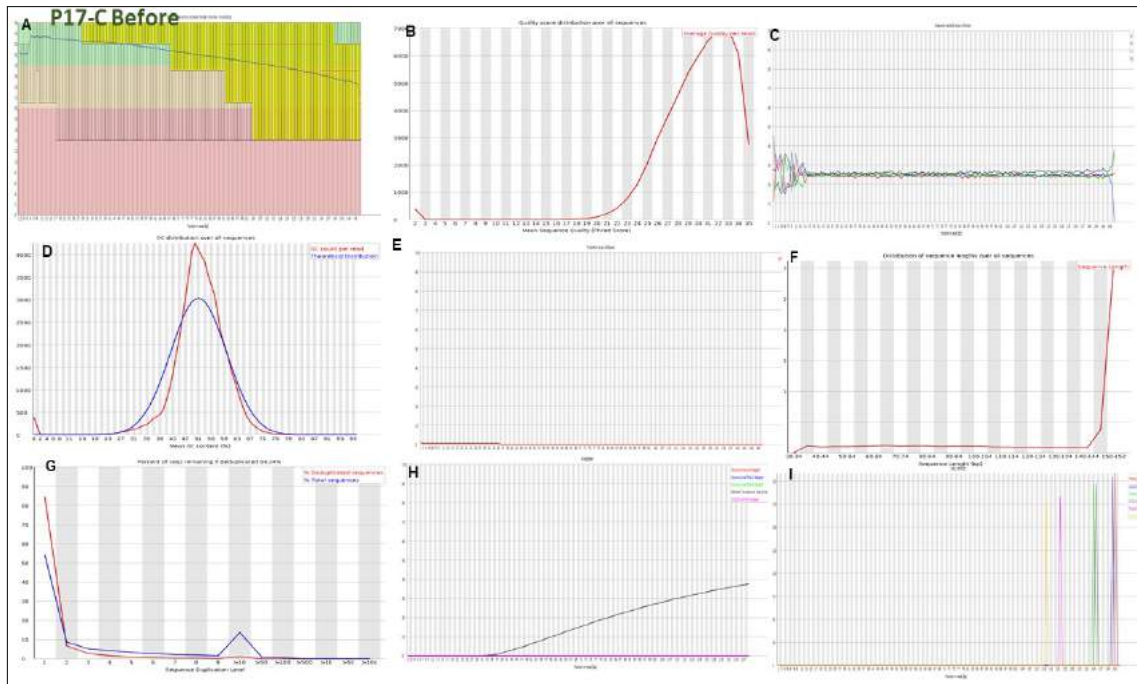
53.   Lanciotti RS, Kosoy OL, Laven JJ, Velez JO, Lambert AJ, Johnson AJ, et al. Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. Emerg Infect Dis. 2008; 14(8):1232–9.

54.   Cruz CD, Torre A, Troncos G, Lambrechts L, Leguia M. Targeted full-genome amplification and sequencing of dengue virus types 1–4 from South America. J Virol Methods [Internet]. 2016;235:158–67.

55.   Leguia M, Cruz CD, Felices V, Torre A, Troncos G, Espejo V, et al. Full-genome amplification and sequencing of Zika viruses using a targeted amplification approach. J Virol Methods [Internet]. 2017; 248:77–82. Available from: http://www.sciencedirect.com/science/article/pii/S0166093417302264

56.   Chakrabarti R, Schutt CE. The enhancement of PCR amplification by low molecular-weight sulfones. Gene. 2001; 274(1–2):293–8.

57.   Chakrabarti R, Schutt CE. Novel sulfoxides facilitate GC-rich template amplification. Biotechniques. 2002; 32(4):866–74.

58.   Chakrabarti R, Schutt CE. The enhancement of PCR amplification by low molecular weight amides. Nucleic Acids Res. 2001; 29(11):2377–81.

59.   Catalog U. Perfect Match PCR Enhancer Instruction Manual. 600129.

60.   Chang A, Ostrove JM, Bird RE. Development of an improved product enhanced reverse transcriptase assay. J Virol Methods. 1997; 65(1):45–54.

61.   Chou Q, Russell M, Birch DE, Raymond J, Bloch W. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. Nucleic Acids Res. 1992; 20(7):1717–23.

62.   Farell EM, Alexandre G. Bovine serum albumin further enhances the effects of organic solvents on increased yield of polymerase chain reaction of GC-rich templates. BMC Res Notes [Internet]. 2012; 5(1):257.

63.   Kang J, Myung SL, Gorenstein DG. The enhancement of PCR amplification of a random sequence DNA library by DMSO and betaine: Application to in vitro combinatorial selection of aptamers. J Biochem Biophys Methods. 2005; 64(2):147–51.

64.   Musso M, Bocciardi R, Parodi S, Ravazzolo R, Ceccherini I. Betaine, dimethyl sulfoxide, and 7-deaza-dGTP, a powerful mixture for amplification of GC-rich DNA sequences. J Mol Diagn [Internet]. 2006; 8(5):544–50.

65.   James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. Integrative Genomics Viewer. Nature Biotechnology 29, 24–26 (2011)

66.   Helga Thorvaldsdóttir, James T. Robinson, Jill P. Mesirov. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics 14, 178-192 (2013).

67.   Okonechnikov K, Golosova O, Fursov M, the UGENE team. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 2012 28: 1166-1167. doi:10.1093/bioinformatics/bts091

68.   Golosova O, Henderson R, Vaskin Y, Gabrielian A, Grekhov G, Nagarajan V, Oler AJ, Quiñones M, Hurt D, Fursov M, Huyen Y. Unipro UGENE NGS pipelines and components for variant calling, RNA-seq and ChIP-seq data analyses. PeerJ 2014 2: e644. doi:10.7717/peerj.644

69.   Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, Cardle L, Shaw PD and Marshall D. 2013. Using Tablet for visual exploration of second-generation sequencing data. Briefings in Bioinformatics 14(2), 193-202.

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

71.   Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012, 9:357-359.

72. Gourinat AC, O'Connor O, Calvez E, Goarant C, Dupont-Rouzeyrol M. Detection of zika virus in urine. Emerg Infect Dis. 2015; 21(1):84–6.
73. Atkinson B, Hearn P, Afrough B, Lumley S, Carter D, Aarons EJ, et al. Detection of zika virus in semen. Emerg Infect Dis. 2016; 22(5):940.
74. Yurumez Y, Yavuz Y, Yucel M, Cetinkaya Z, Ciftci IH. Reactions and complications to bites. Neth J Med. 2016; 74(3):142–3.
75. Schrader C, Schielke A, Ellerbroek L, Johne R. PCR inhibitors - occurrence, properties and removal. J Appl Microbiol. 2012; 113(5):1014–26.
76. Al-Soud WA, Rådström P. Purification and Characterization of PCR-Inhibitory Components in Blood Cells Purification and Characterization of PCR-Inhibitory Components in Blood Cells. J Clin Microbiol. 2001; 39(2):485–93.
77. King CE, Debruyne R, Kuch M, Schwarz C, Poinar HN. A quantitative approach to detect and overcomePCR inhibition in ancient DNA extracts. Biotechniques. 2009; 47(5):941–9.
78. Warren J. Joseph Warren, Ph.D., F-ABC. A Rev PCR Inhib It's Implic Hum Identity Test. 2012;
79. Liles C, Kumar MA, Weinshenker D. Use of Gene-Specific Primer Cocktails for First-Strand cDNA Synthesis with a Reverse Transcriptase Kit. Am Biotechnol Lab. 2004; 2(December):20–1.

APPENDIX A - General profile of samples before and after trimming

A - Per base sequence quality; B – Per sequence quality scores (Average quality per read); C – Sequence content across all bases (G, C, T and A); D - GC distribution over all sequences; E – Per base N Content; F - Distribution of sequence lengths over all sequences; G – Sequence duplicated level; H - % Adapter; I – K-mer content.

P17-C Before

P17-C After

P17-H Before

P17-H After

P17-K Before

P17-K After

P17-S Before

P17-S After

S007 Before

S007 After

IEC-Short Before

IEC-Short After

IEC-Long Before

IEC-Long After
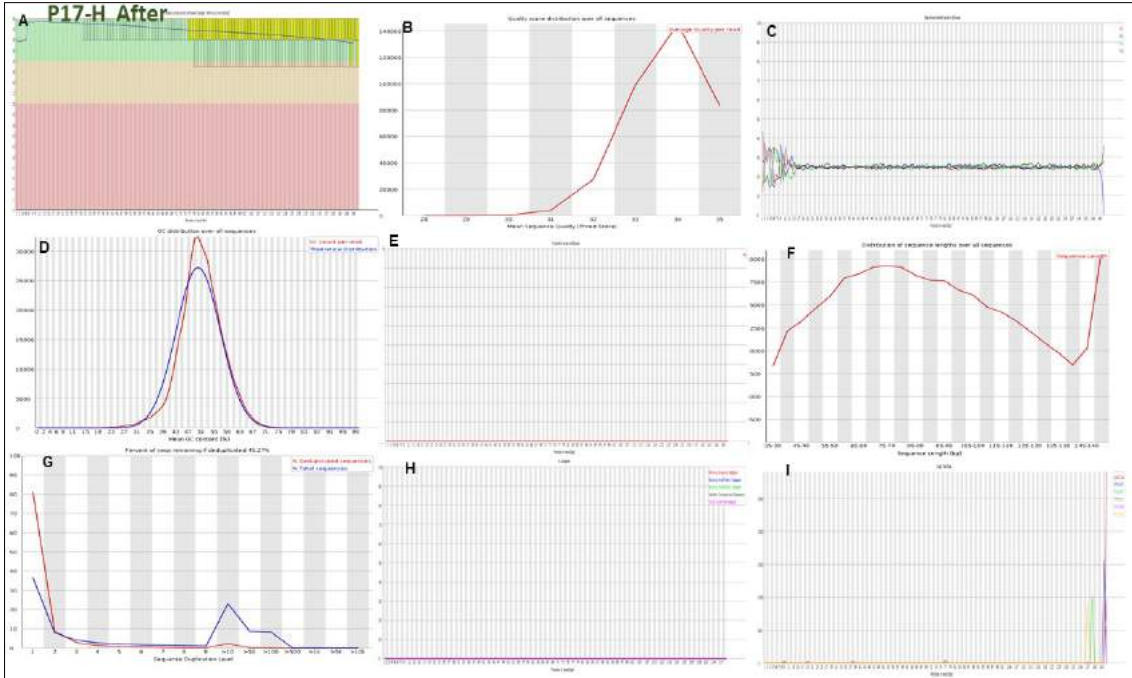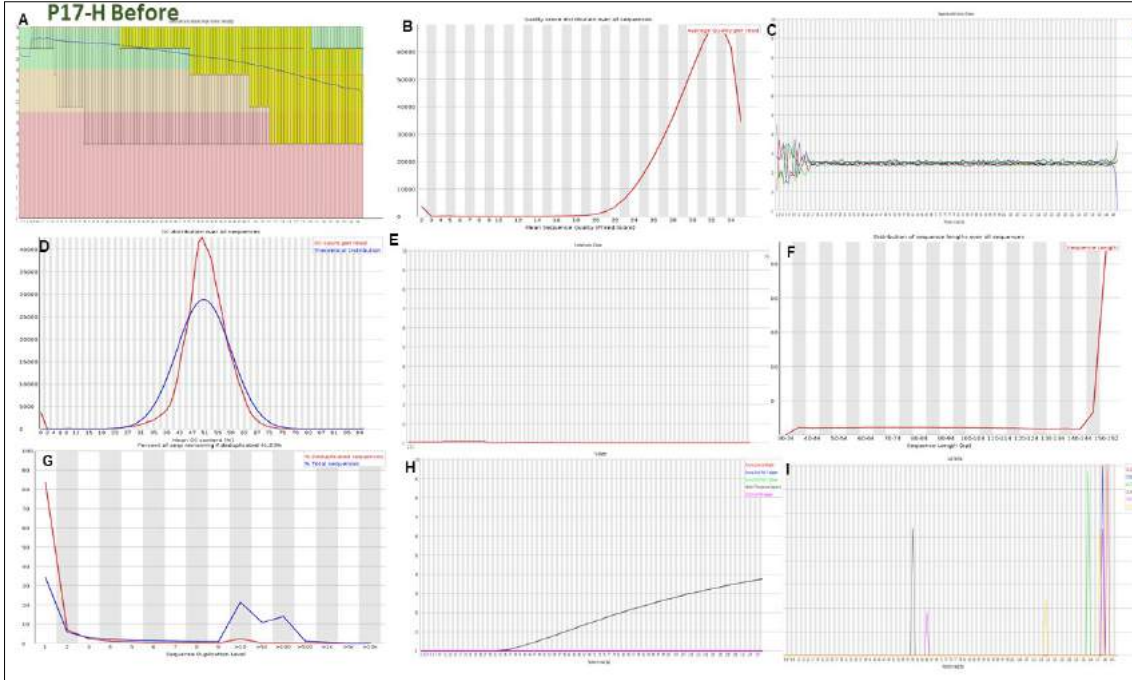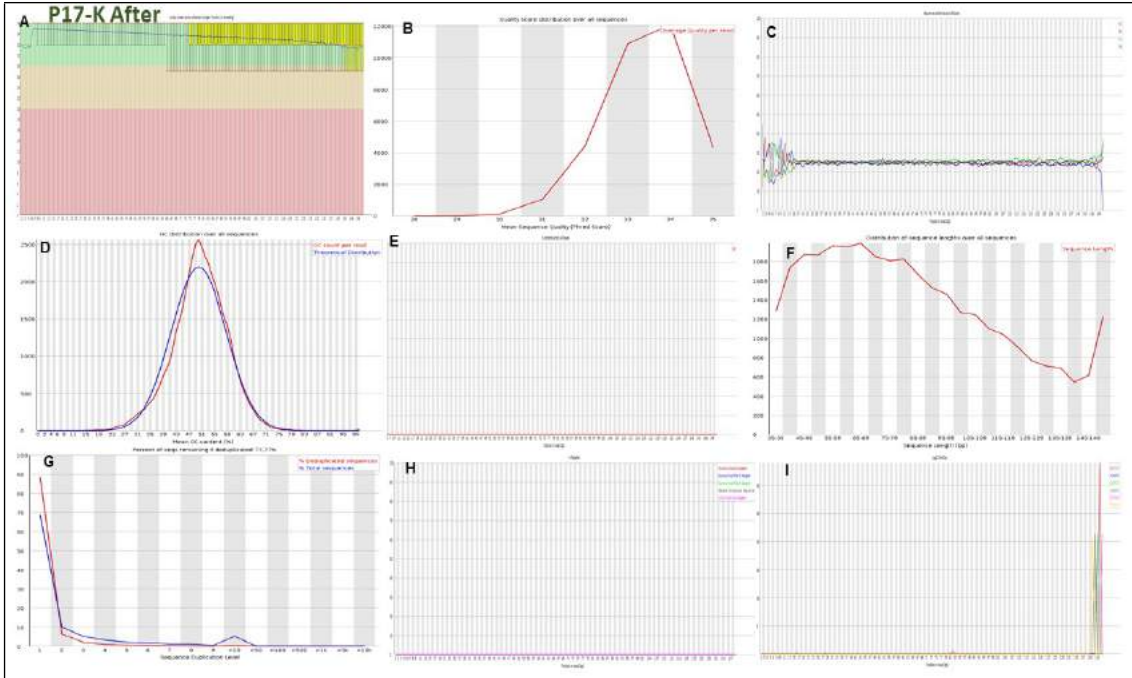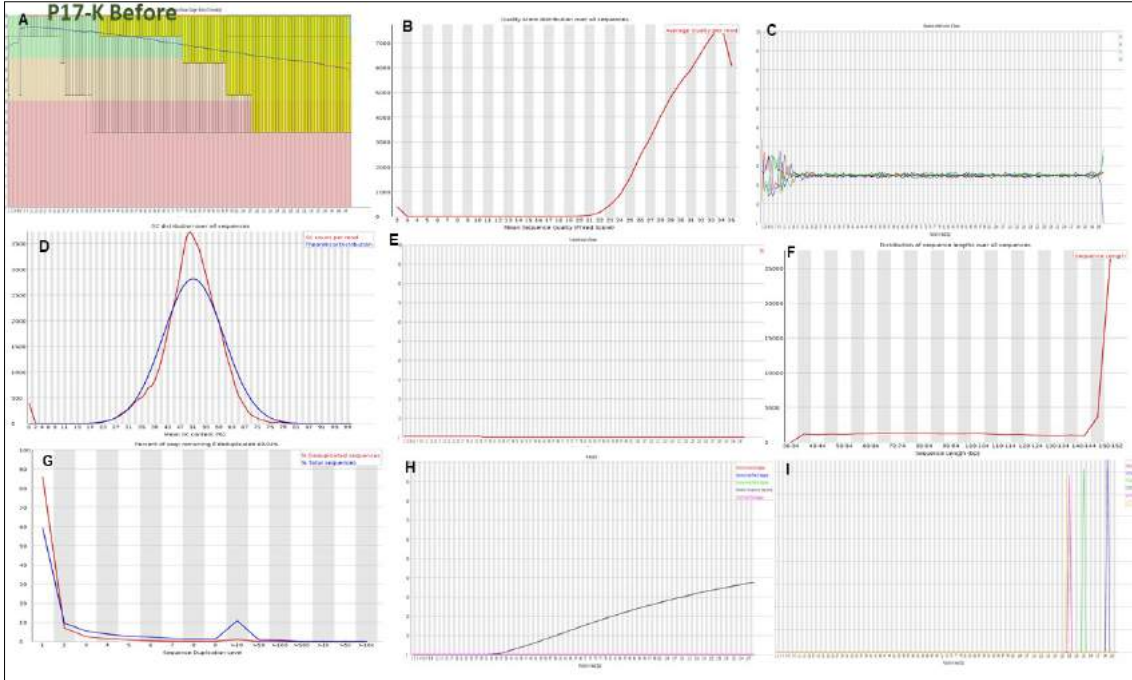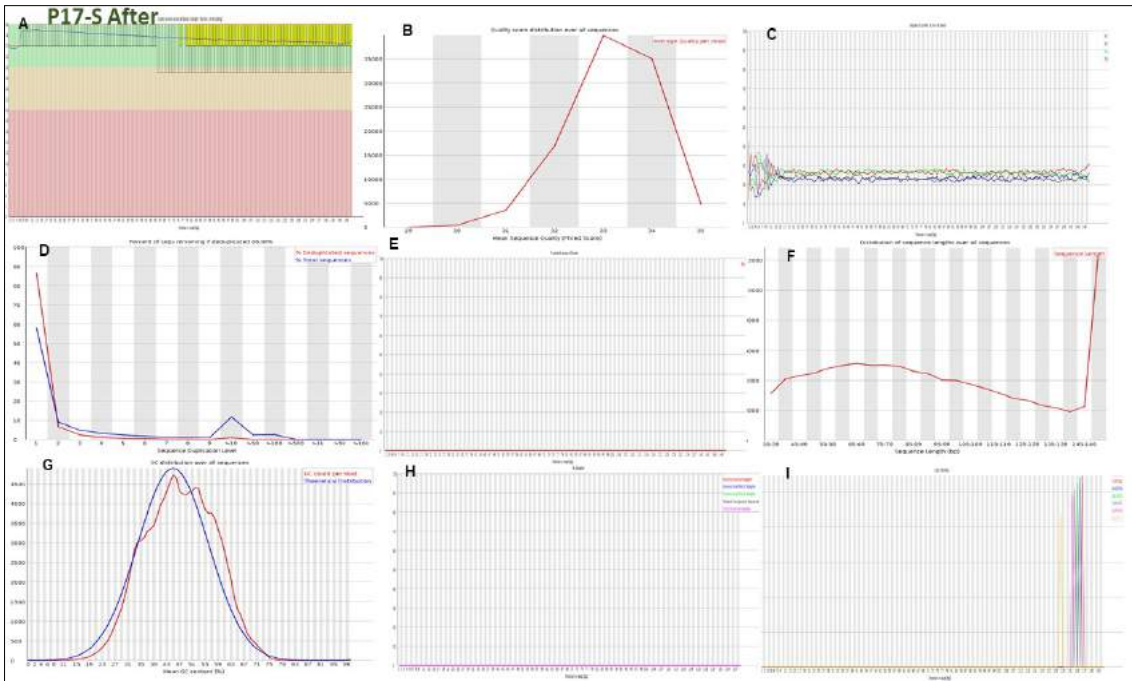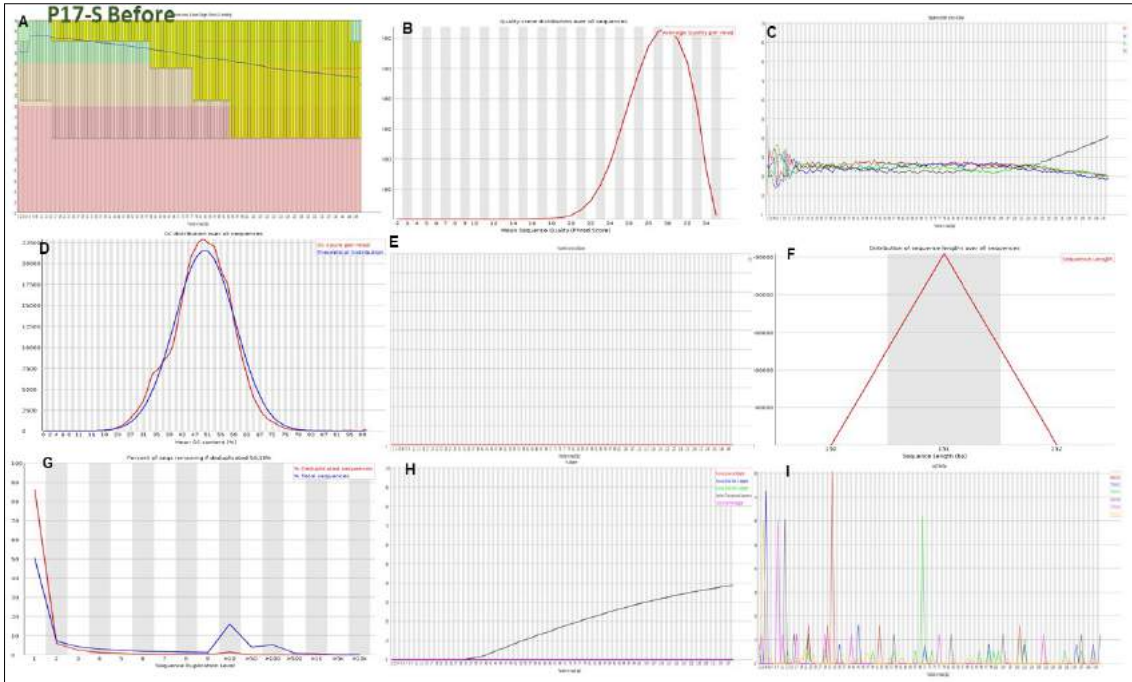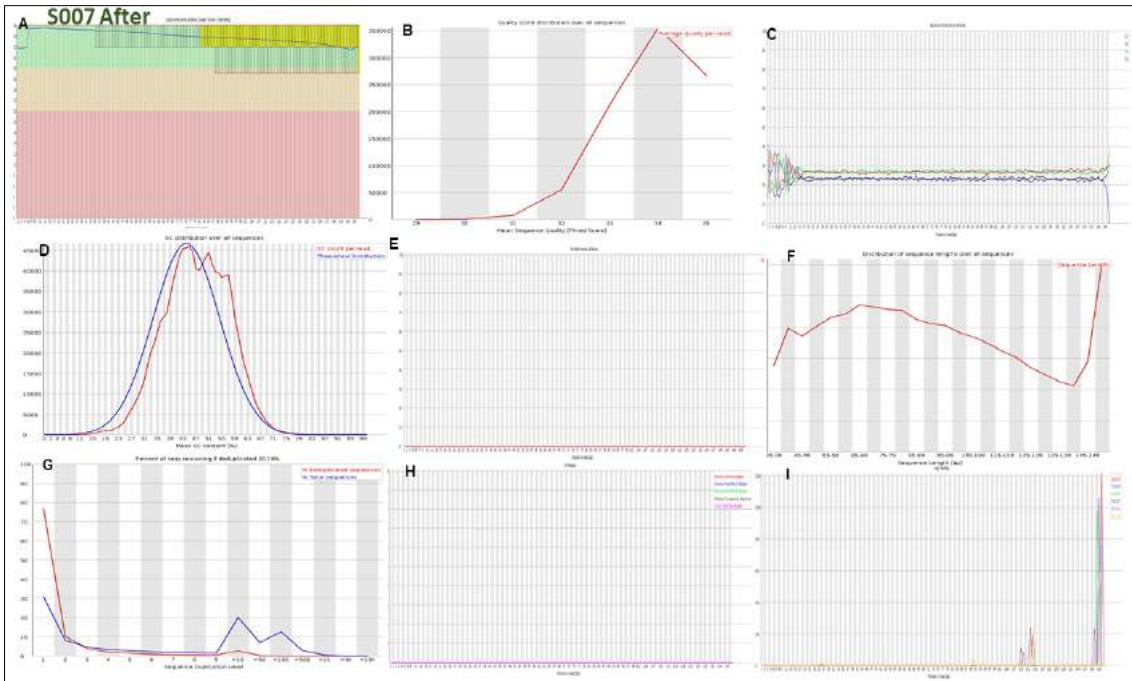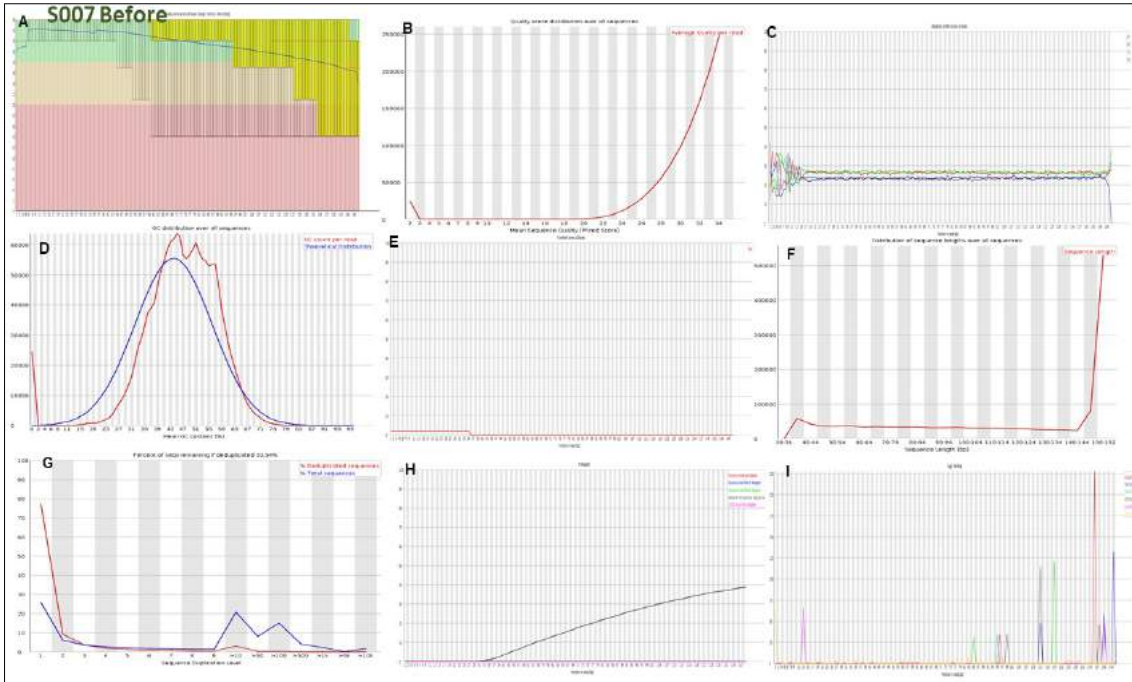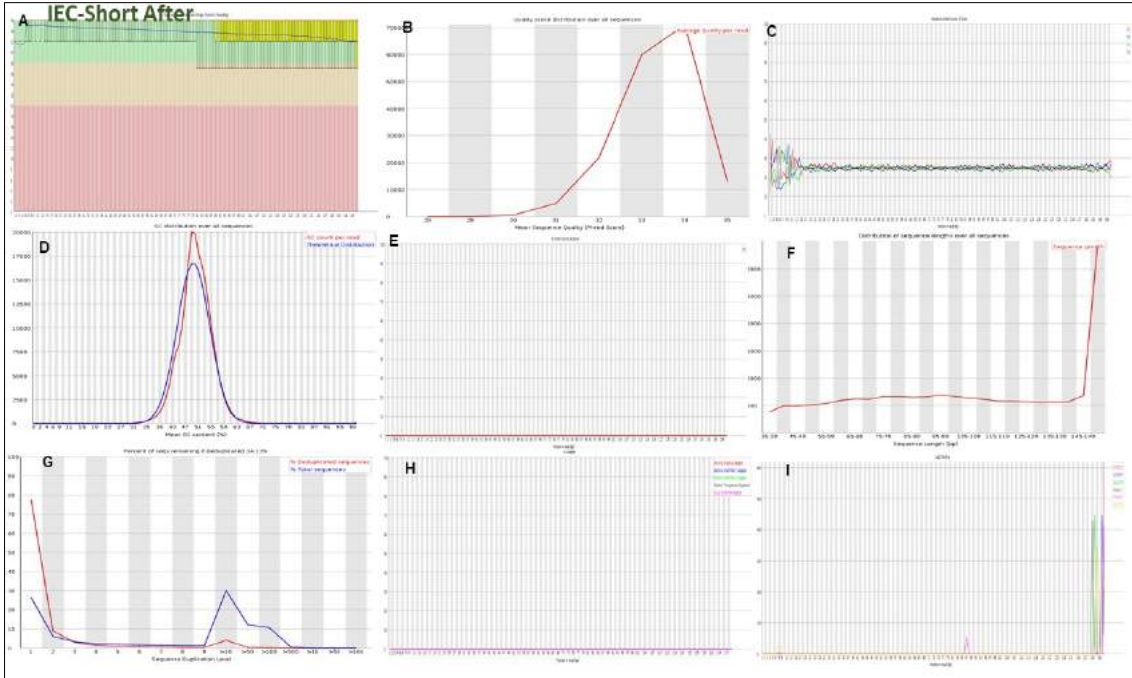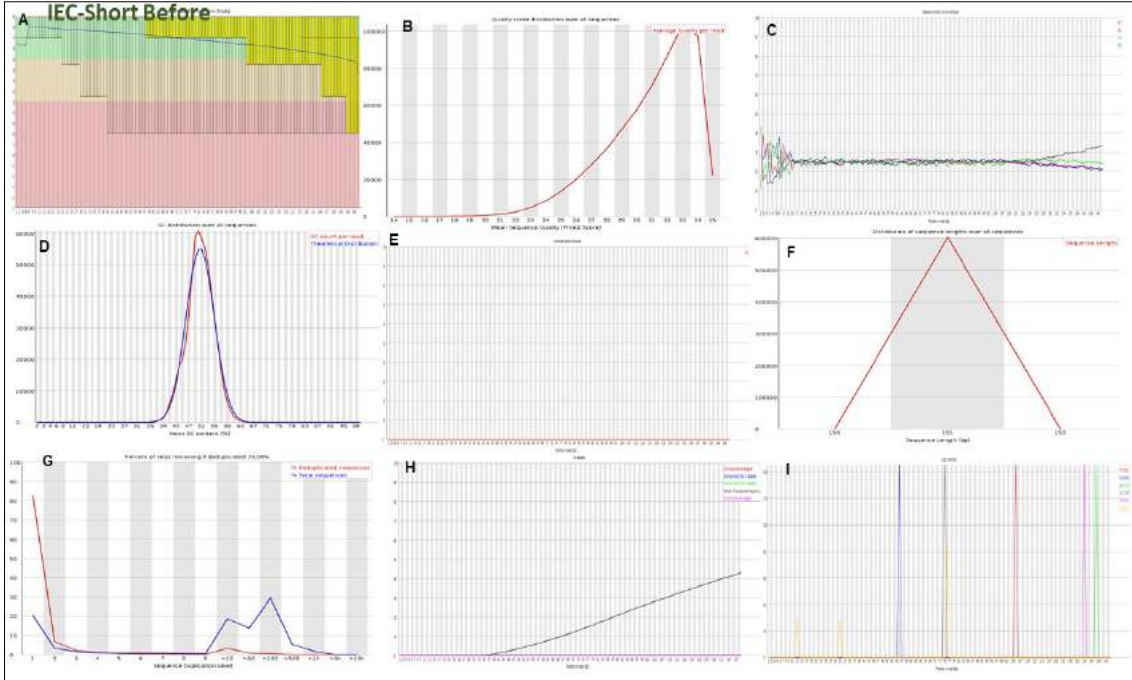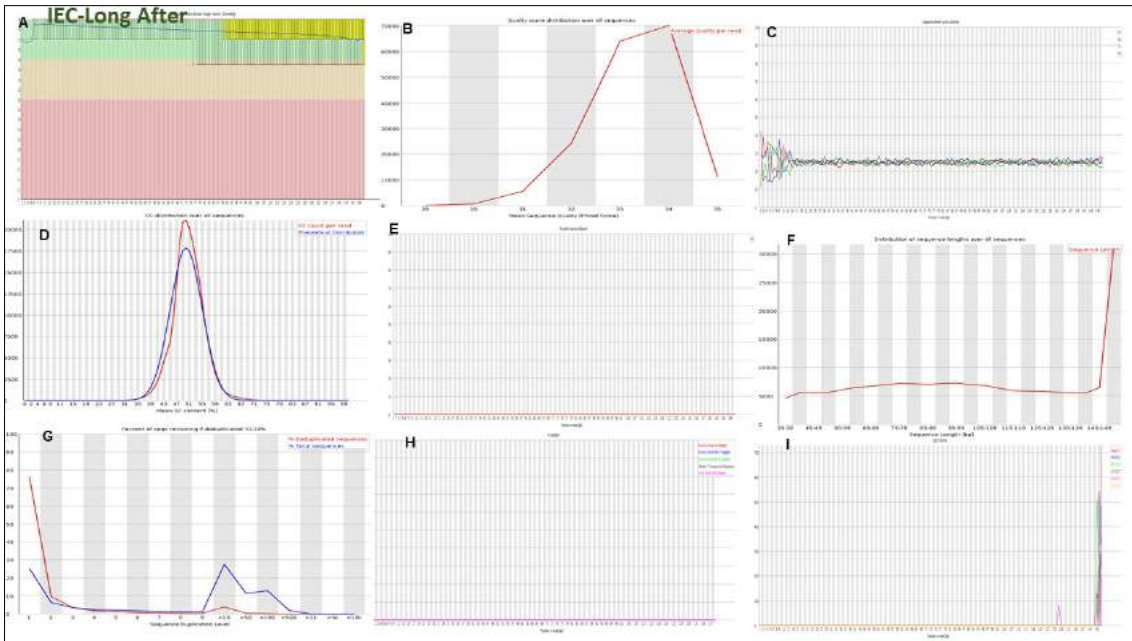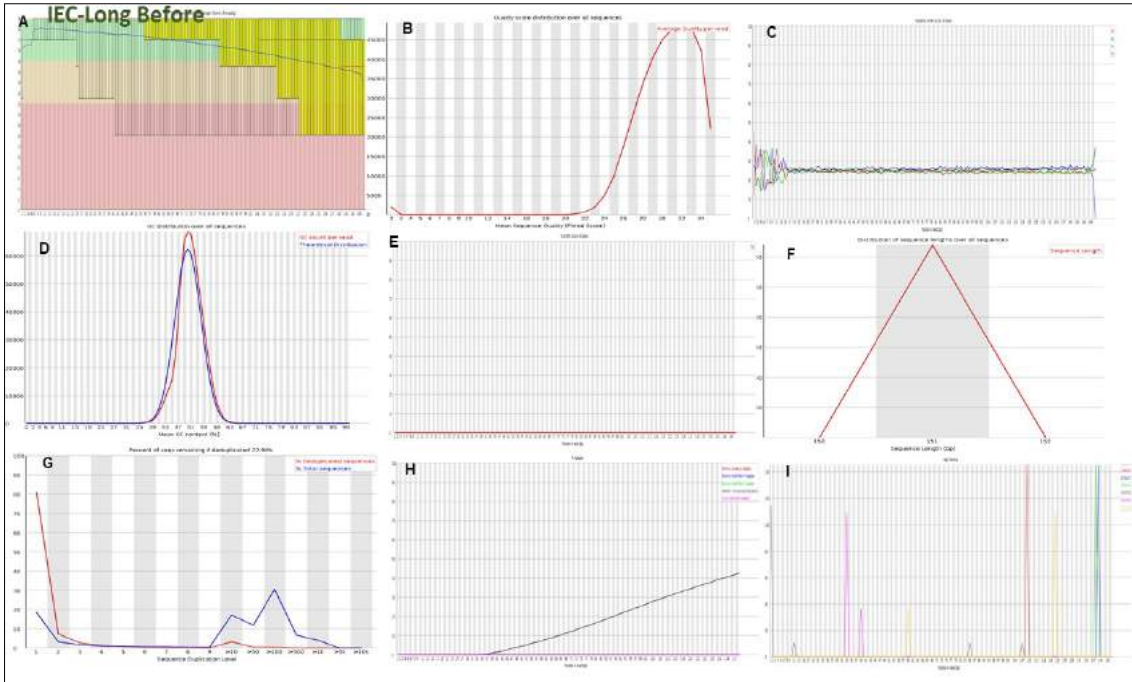
APPENDIX B - Primers Used in Short and Long amplicons

Primers used for short amplicons generation

Set	Primer Name	Sequence 5'-3'	Position	Amplicon (bp)
1	ZIKbr-F1	GACAGTTCGAGTTTGAAGCGAA	28-49	751
	ZIKbr-R2	CTAGTGGAATGGGAGGGGAGC	759-779	
2	ZIKbr-F2	TCAATGCTAGGAAGGAGAAGAAGA	389-412	885
	ZIKbr-R3	CCTCTGTCCACTAACGTTCTTTTG	1251-1274	
3	ZIKbr-F2	TCAATGCTAGGAAGGAGAAGAAGA	389-412	1384
	ZIKbr-R4	GTGAACTGCTCCTTCTTGACTC	1752-1773	
4	ZIKbr-F3	AGCACAGTGGGATGATTGTTAATG	1415-1438	866
	ZIKbr-R5	CTCCAACTGATCCAAAGTCCCAG	2259-2281	
5	ZIKbr-F4	GTGTCATACTCCTTGTGTACTGC	1882-1904	919
	ZIKbr-R6	GGCAATCTCTGTGGACCTCTC	2781-2801	
6	ZIKbr-F5	GGGTCTGAACACAAAGAATGGATC	2400-2423	884
	ZIKbr-R7	CCTTTCATTTGGGTCCTGTAGC	3263-3284	
7	ZIKbr-F6	ATAACAGCTTTGTCGTGGATGGTG	2876-2899	2076
	ZIKbr-R8	CTGACTTTGAATGCCGCTATCAG	3778-3800	
8	ZIKbr-F7	GAAGGGTGATCGAGGAATGGT	3401-3421	881
	ZIKbr-R9	ATCTCTATATCTGCCTTGGCGAA	4282-4304	
9	ZIKbr-F7	GAAGGGTGATCGAGGAATGGT	3401-3421	1408
	ZIKbr-R10	ATCAAGTCTCCCTTCACCGCTT	4788-4809	
10	ZIKbr-F8	CATTGAAAGAGCAGGTGACATCA	4377-4399	908
	ZIKbr-R11	GTTGGAGCTAAGATCACGGTG	5265-5285	
11	ZIKbr-F9	GGAGAGAGAGCGAGGAACATC	4918-4938	875
	ZIKbr-R12	CTCTGTCTCAAAAGTCTTTCTGCT	5770-5793	
12	ZIKbr-F10	ACCTTCACTTCACGTCTACTACAG	5404-5427	897
	ZIKbr-R13	CAAAGCACCATCTTCTATCTGTGT	6278-6301	
13	ZIKbr-F11	ATACTTGATGGCGAGAGAGTCATT	5910-5934	892
	ZIKbr-R14	GAAACACAACAATGAGGACACATG	6779-6802	
14	ZIKbr-F12	AGAGTTTGTTCAGATCATGCGG	6403-6424	885
	ZIKbr-R15	CTGGGATCAAGTACATGTAGTGC	7266-7288	
15	ZIKbr-F13	AGTAGGTCTTCTGGGCTTGATTAC	6885-6908	897
	ZIKbr-R16	CTCTTCTCTGCACACCTCGGT	7762-7782	
16	ZIKbr-F13	AGTAGGTCTTCTGGGCTTGATTAC	6885-6908	1922
	ZIKbr-R17	CAGGAAGAGACCATGCTCATAAC	8785-8807	
17	ZIKbr-F14	CATTGATCTTGGATGTGGCAGAG	7896-7918	911
	ZIKbr-R17	CAGGAAGAGACCATGCTCATAAC	8785-8807	
18	ZIKbr-F15	AGTGAAATATGAGGAGGATGTGGA	8415-8438	859
	ZIKbr-R18	CAGTGTCATCTGCATACATCCTTC	9251-9274	
19	ZIKbr-F16	AGGGGCAATATTTGAAGAGGAAAA	8895-8918	892
	ZIKbr-R19	ACTTCTTCCCAGTTGTCCCATC	9767-9787	
20	ZIKbr-F17	CCAAAACAAAGTGGTAAAGGTCCT	9387-9410	896
	ZIKbr-R20	CGCACCATGTTGACTGTGTTTT	10262-10283	
21	ZIKbr-F17	CCAAAACAAAGTGGTAAAGGTCCT	9387-9410	1285
	ZIKbr-R21	CTCTAACCACTAGTCCCTCTTCTG	10649-10672	
22	ZIKbr-F18	AGCACCAATCTTAATGTTGTCAGG	10377-10400	295
	ZIKbr-R21	CTCTAACCACTAGTCCCTCTTCTG	10649-10672	

Primers used for Long amplicons generation

Set	Primer Name	Sequence 5'-3'	Position	Amplicon (bp)
1	ZIKbr-F1	GACAGTTCGAGTTTGAAGCGAA	28-49	2773
	ZIKbr-R6	GGCAATCTCTGTGGACCTCTC	2781-2801	
2	ZIKbr-F2	TCAATGCTAGGAAGGAGAAGAAGA	389-412	2895
	ZIKbr-R7	CCTTTCATTTGGGTCCTGTAGC	3263-3284	
3	ZIKbr-F3	AGCACAGTGGGATGATTGTTAATG	1415-1438	2385
	ZIKbr-R8	CTGACTTTGAATGCCGCTATCAG	3778-3800	
4	ZIKbr-F5	GGGTCTGAACACAAAGAATGGATC	2400-2423	2885
	ZIKbr-R11	GTTGGAGCTAAGATCACGGTG	5265-5285	
5	ZIKbr-F6	ATAACAGCTTTGTCGTGGATGGTG	2876-2899	2917
	ZIKbr-R12	CTCTGTCTCAAAAGTCTTTCTGCT	5770-5793	
6	ZIKbr-F7	GAAGGGTGATCGAGGAATGGT	3401-3421	2900
	ZIKbr-R13	CAAAGCACCATCTTCTATCTGTGT	6278-6301	
7	ZIKbr-F8	CATTGAAAGAGCAGGTGACATCA	4377-7399	2911
	ZIKbr-R15	CTGGGATCAAGTACATGTAGTGC	7266-7288	
8	ZIKbr-F9	GGAGAGAGAGCGAGGAACATC	4918-4938	2864
	ZIKbr-R16	CTCTTCTCTGCACACCTCGGT	7762-7782	
9	ZIKbr-F10	ACCTTCACTTCACGTCTACTACAG	5404-5427	3403
	ZIKbr-R17	CAGGAAGAGACCATGCTCATAAC	8785-8807	
10	ZIKbr-F11	ATACTTGATGGCGAGAGAGTCATT	5910-5934	3364
	ZIKbr-R18	CAGTGTCATCTGCATACATCCTTC	9251-9274	
11	ZIKbr-F12	AGAGTTTGTTCAGATCATGCGG	6403-6424	3384
	ZIKbr-R19	ACTTCTTCCCAGTTGTCCCATC	9767-9787	
12	ZIKbr-F13	AGTAGGTCTTCTGGGCTTGATTAC	6885-6908	3398
	ZIKbr-R20	CGCACCATGTTGACTGTGTTTT	10262-10283	
13	ZIKbr-F14	CATTGATCTTGGATGTGGCAGAG	7896-7918	2776
	ZIKbr-R21	CTCTAACCACTAGTCCCTCTTCTG	10649-10672	
14	ZIKbr-F15	CTGGGATCAAGTACATGTAGTGC	8415-8438	2257
	ZIKbr-R21	CTCTAACCACTAGTCCCTCTTCTG	10649-10672	
15	ZIKbr-F16	AGGGGCAATATTTGAAGAGGAAAA	8895-8918	1777
	ZIKbr-R21	CTCTAACCACTAGTCCCTCTTCTG	10649-10672	
16	ZIKbr-F1	GACAGTTCGAGTTTGAAGCGAA	28-49	281
	ZIKbr-R1	TCTATTGATGAGACCCAGTGATGG	286-309	

APPENDIX C-Published Manuscript

Full length article

Outbreak of chikungunya virus in a vulnerable population of Sergipe, Brazil—A molecular and serological survey

Marielton dos Passos Cunha[a], Cliomar Alves dos Santos[b], Daniel Ferreira de Lima Neto[a], Alessandra Soares Schanoski[c], Shahab Zaki Pour[a], Saulo Duarte Passos[d], Mércia Simone Feitosa de Souza[e], Danuza Duarte Costa[b], Paolo Marinho de Andrade Zanotto[a,*]

[a] Laboratory of Molecular Evolution and Bioinformatics, Department of Microbiology, Biomedical Sciences Institute, University of São Paulo, São Paulo, Brazil
[b] Health Foundation Parreiras Horta (FSPH), Central Laboratory of Public Health (LACEN/SE), State Secretary for Health, Sergipe, Brazil
[c] Bacteriology Laboratory II, Butantan Institute, São Paulo, Brazil
[d] Department of Pediatrics, Jundial Medical Faculty, São Paulo, Brazil
[e] Coordination of the Nucleus of Communicable Diseases, State Secretary for Health, Sergipe, Brazil

ARTICLE INFO

ABSTRACT

Background: Chikungunya virus (CHIKV) is a re-emerging arbovirus that is causing outbreaks in several countries of the Americas. The virus was introduced in Brazil in 2014, and since then, several Brazilian states have notified autochthonous cases.
Objectives: Provide additional evidence on a CHIKV outbreak and an outline of the laboratory and clinical profile of symptomatic patients in Sergipe, Brazil.
Study design: In February 2016, we collected 142 serum samples from symptomatic patients for arboviruses in Sergipe, Brazil. All samples were submitted to qRT-PCR for the emerging arboviruses circulating in Brazil – ZIKV, CHIKV, and DENV – and later submitted to the immunoenzymatic assay. RNA positive samples were randomly selected and sequenced for characterization of the genotype involved in the outbreak.
Results: Our study had 75.35% (107/142) positivity for CHIKV infection, with all age groups and genera being equally infected. The virus was identified in 11 of the 13 cities studied in that state, including the ECSA genotype. Importantly, fever was the only statistically significant symptoms for CHIKV infection ($p < 0.05$), while asthenia was significantly associated with symptomatic patients that were CHIKV-negative ($p < 0.05$).
Conclusions: Our findings support the importance of fever as a clinical marker and contribute to molecular and serological surveillance data, which may help in the understanding of CHIKV circulation, emergence and clinical description.

1. Background

Chikungunya virus (CHIKV: *Togaviridae: Alphavirus*) is a mosquito-borne alphavirus mainly transmitted in urban and peri-urban areas by *Aedes* mosquitoes [1,2]. It was first described during a "dengue-like" outbreak in Tanzania between 1952 and 1953 [3]. In 2005, CHIKV caused a remarkable outbreak in La Réunion, a Madagascar island, with more than a third of its population testing positive for the infection [4]. The virus then spread across Asia, Europe, Pacific islands and the Americas, causing large outbreaks [2,5–8].

CHIKV genome is made of a single-stranded, positive-sense RNA of approximately 12 kb, which contains two open reading frames (ORFs) encoding for non-structural and structural proteins, respectively, all of which are arranged as a 5′-cap-nsP1-nsP2-nsP3-nsP4-(junction-region)-C-E3-E2-6K-E1-poly(A)-3′ [9]. CHIKV is classified into three genotypes based on its genetic diversity and the initial geographical distribution of its lineages that include West African, Asian, and East-Central-South-Africa (ECSA) genotypes [1].

CHIKV causes a self-limiting disease characterized by fever, asthenia, arthralgia, myalgia, headache and rash [1]. The acute phase of infection lasts one week and is characterized by the onset of fever and high viral loads. After the end of viremia, IgM becomes detectable, followed by severe myalgias and arthralgias [1]. The symptoms disappear over a period of time ranging from days to years [10,11].

The first autochthonous transmission of CHIKV in Brazil was described on 13 of September 2014, in the north region of Brazil caused by Asian genotype, and seven days later, autochthonous cases of ECSA genotype was confirmed in the Northeast of Brazil [12]. In the

following years, the ECSA genotype continued to be identified in the Northeast, as well as in the Southeast region of Brazil [12–14]. In Brazil, CHIKV was associated with severe and fatal cases [15,16], posing not only a major threat to public health but also, a serious economic burden due to the prolonged disability and high medical costs imposed. Moreover, this situation is made worse by the lack of available treatments and/or vaccines against this virus.

2. Objectives

The present study aimed to provide additional evidence on a CHIKV outbreak and an outline of the laboratory and clinical profile of symptomatic patients during an outbreak in the state of Sergipe, Brazil. This state is located in northeastern Brazil, a region considered endemic for dengue virus (DENV) and zika virus (ZIKV) in previous years [17].

3. Study design

3.1. Study location

The state of Sergipe is located in the Northeast Region and is surrounded by the Atlantic Ocean to the east and adjoins the states of Bahia to the west and south, and Alagoas to the north (Fig. 1). It is the smallest State in Brazil, occupying a total area of 21 915,116 km^2, and has an estimated population of 2,068,017 inhabitants. The climate is characterized as tropical in the coastal region, and also as semi-arid and tropical savannah in the western region of the state.

3.2. Ethical statement

The samples analyzed in this study were collected during an on-going project for arbovirus research in Sergipe, Brazil approved by the Ethics Committee of the Department of Microbiology of the Institute of Biomedical Sciences of the University of São Paulo (Protocol 1284/CEPSH – CAAE: 54937216.5.0000.5467). All participating subjects were asked to sign an informed consent and were subjected to a questionnaire that asked about clinical symptoms and socio-demographic information after their acceptance to participate in the study.

3.3. Patients

This was an observational cross-sectional study. It was conducted with sera samples obtained from 142 patients with symptoms of acute disease presenting an "arbovirus-like" infection. Samples were collected in February 2016 in Sergipe state, Brazil. All patients were attended at basic health units in Sergipe. All samples were kept in storage at −80 °C until processing.

3.4. Molecular characterization

Nucleic acid extraction was performed using the QIAamp Viral RNA Mini Kit (Qiagen, Valencia, CA, USA) and carried out according to the manufacturer's instructions. Molecular detection of DENV, CHIKV, and ZIKV was performed with the use of the Quantitative RT-PCR ReadyMix™ reagents (Sigma-Aldrich, St. Louis, MO, USA). Primers/probes used for these viruses were previously described for DENV [18], ZIKV [19] and for CHIKV [20]. qRT-PCR reactions consisted of a step of reverse transcription at 44 °C for 30 min of the enzyme activation at

1 - Aracajú
2 - Barra dos coqueiros
3 - Campo do Brito
4 - Carira
5 - Frei Paulo
6 - Itabaiana
7 - Macambira
8 - Malhador
9 - Moita Bonita
10 - Pinhão
11 - Salgado
12 - São Cristovão
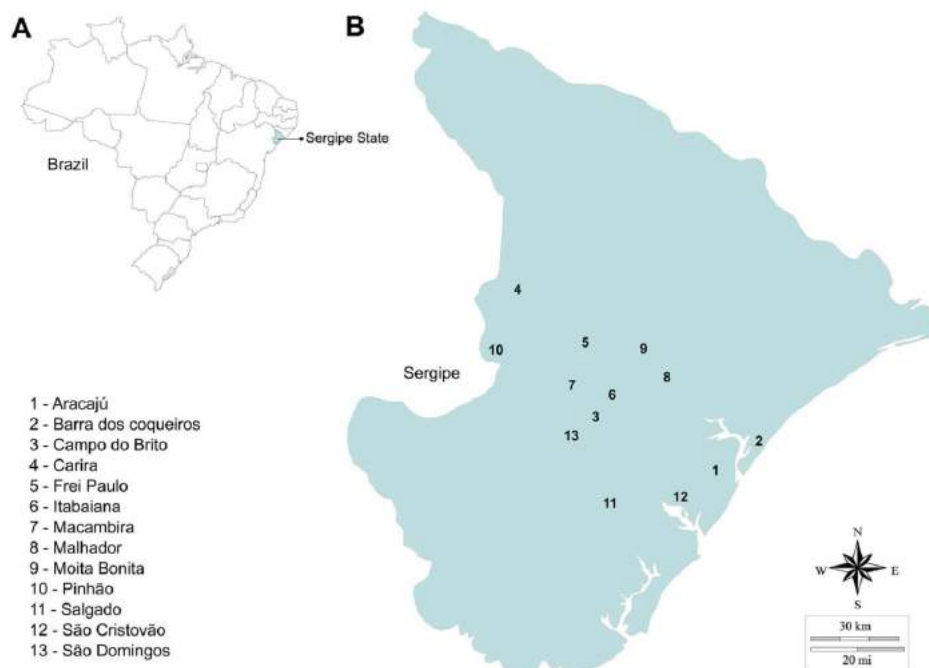13 - São Domingos

Fig. 1. Map of the state of Sergipe and its localization in Brazil with the origin location of all patients.
(A) Map with the location of the state of Sergipe in the Brazilian territory. (B) The state of Sergipe is located in the Northeast Region and is bounded by the Atlantic Ocean to the east and the states of Bahia to the west and south, and Alagoas to the north.
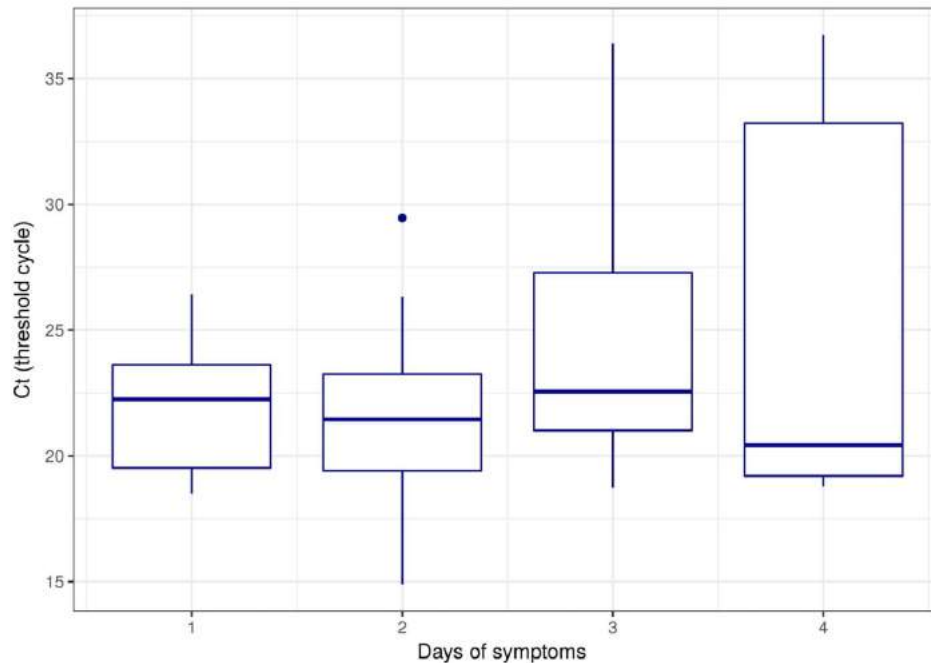
Fig. 2. Threshold cycle according to days of symptoms in positive patients.
The initial phase of infection had low dispersion and low Ct values, with an increase in both parameter values for Ct and dispersion, during later stages of viremia. Boxplots represent the 75th percentile, median, 25th percentile and the whiskers extend to the highest and lowest value in the 1.5x interquartile range.

94 °C for 2 min, and 40 cycles at 94 °C for 15 s and 60 °C for 1 min for hybridization and extension with the use of ABI7500 equipment (Thermo Fisher Scientific, Waltham, MA, USA).

3.5. Serological characterization

Sera samples were evaluated with a commercial semi-quantitative ELISA kit (enzyme-linked immunosorbent assay), which detects anti-CHIKV IgM and IgG antibodies. All procedures were carried out according to the manufacturer's instructions (Euroimmun, Lubeck, Germany). Briefly, sera were diluted 1:101 in sample buffer and incubated at 37 °C for 60 min in a microplate well; a calibrator, positive and negative controls, provided by the manufacturer, was used. The optical density (O.D.) was measured in an Epoch Microplate Spectrophotometer (BioTek, Vermont, USA) and the results were calculated according to the manufacturer's instructions. Values < 0.8 were regarded as negative, ≥ 0.8 to < 1.1 as borderline, and ≥ 1.1 as positive.

3.6. Phylogenetic characterization

To determine circulating CHIKV genotypes, positive samples were subjected to a reverse transcription reaction with High-Capacity cDNA Reverse Transcription Kits (Applied Biosystem, Foster City, CA, USA) following the manufacturer's recommendations. The Invitrogen protocol for PCR amplification was executed with minor modifications. Briefly, reactions were conducted in 22 μL total containing 2 μL CHIKV cDNA, 1 μL each (10 μM) forward and reverse primers previously described [21], 0.3 μL (5 U/μL) Platinum Taq DNA Polymerase High Fidelity (Invitrogen, Carlsbad, CA, USA), 2.5 μL 10 x PCR Buffer

(reaching a final concentration of 1.13 x per reaction), 1 μL (50 mM) $MgSO_4$, 1 μL (10 mM) dNTPs, and 13.2 μL nuclease-free water. Amplification was carried out as follows: 94 °C for 3 min, followed by 40 cycles at 94 °C for 15 s, 56 °C for 30 s, and 68 °C for 4 min, followed by a final extension at 68 °C for 10 min.

Sequencing reactions were performed with the BigDye terminator v.3.1 cycle sequencing kit (Applied Biosystems) using specific primer pairs in the Genetic analyzer automated sequencer "Applied Biosystems PRISM" 3130xl (Applied Biosystems). Chromatograms were analyzed with CodonCode Aligner 3.7.1 (Sequence Assembly and Alignment Software – CodonCode Corporation) with a *Phred* quality score of 20 as the cut-off for trimming of low-quality sequences. Sequences were aligned using Clustal X2 [22] and the alignment was manually edited using AliView [23]. Viral phylogenies based on the partial E1-3'NTR region were estimated using Maximum Likelihood (ML) considering the nucleotide substitution model general time-reversible with gamma-distributed rate variation (GTR + G) and 10,000 replicates implemented in FastTree 2 [24]. The final tree was then visualized and plotted using FigTree v.1.4.2 (http://tree.bio.ed.ac.uk).

3.7. Data analysis

The available information from the patient's medical records was analyzed together with the molecular and serological data. Statistical analysis was performed using the online contingency table analysis from the GraphPad Prism website assuming a two-tailed Fisher's exact test. Statistical significance was assessed, considering 95% confidence intervals, and p values below 0.05.

Table 1

Molecular and serological markers (viral RNA, IgM and IgG) against CHIKV in Sergipe-Brazil, February 2016 (n = 142).

Markers	Positive (n = 107)	Negative (n = 35)
Viral RNA	91	0
IgM	16	0
IgG	6	0
IgM and IgG	6	0

4. Results

The CHIKV prevalence in the samples was 75.35% (107/142) considering the results of qRT-PCR and/or IgM-ELISA. 64.1% (91/142) of the patients were positive for viral RNA, with low dispersion and Ct (threshold cycle) values at the beginning of infection, with an increase in both parameters in the cases with longer periods after the onset of symptoms (Fig. 2). Moreover, 11.3% (16/142) were only positive for IgM in the whole cohort (Table 1). There were no statistically significant differences between the ages and sexes of the CHIKV positive individuals, but the majority of samples were of female patients, 61.7% (66/107), and in general patients were between 15–30 years old, 44.9% (48/107) (Table 2). We found CHIKV circulation in eleven out of thirteen cities in Sergipe, shown in Table 3. Symptomatic patients had statistically significant (p < 0.05) association with both fever and asthenia. Importantly, fever associated with CHIKV-positive cases, while asthenia associated with CHIKV-negative cases. However, reported symptoms such as joint pain, hemorrhagic signs, headache, rash, signs of meningoencephalitis, itching, retro-orbital pain, abdominal pain, myalgia, and vomiting had no significant association with acute CHIKV infection.

Serological results showed that 4.2% (6/142) patients were IgG positive, and all of them were also IgM positive (Table 1). No DENV or ZIKV genomic RNA was detected in all sera samples.

A phylogenetic tree, including sequences from Sergipe and representative sequences for all the known genotypes, indicated that the Sergipe sequences belonged to a monophyletic group, characterized as ECSA and were related to sequences isolated in previous years in Northeast region of Brazil between 2014–2016 (Fig. 3).

Table 3

Comparison of total cases (n = 142) with positive cases for CHIKV infection (n = 107) in the sampled cities of Sergipe-Brazil.

Location	Cases (n = 142)	Positive cases (n = 107)
Aracajú	64	44
Barra dos coqueiros	1	1
Campo do Brito	8	7
Carira	2	1
Frei Paulo	3	3
Itabaiana	17	13
Macambira	4	4
Malhador	4	3
Moita Bonita	11	11
Pinhão	19	15
Salgado	1	0
São Cristovão	3	0
São Domingos	5	5
TOTAL	142	107

5. Discussion

Here we present clinical, serological and molecular data showing a high prevalence of CHIKV genotype ECSA in symptomatic patients from a susceptible population spread over a considerable geographical region of Sergipe. CHIKV was first identified in Brazil in September 2014, with two distinct introductions, one in Oiapoque-Amapá in the North Region caused by the Asian genotype, and another in Feira de Santana-Bahia in the Northeast region of Brazil, caused by the ECSA genotype [12]. In 2015, the Northeast region of Brazil was hit by a large outbreak caused by the ZIKV associated with the co-circulation of the CHIKV and DENV [13], with cases of co-infections [25].

The State of Sergipe is located in a region with intense DENV [26] and ZIKV [17] circulation in previous years, due to the presence of *Aedes aegypti*, which is the main vector responsible for the circulation of these viruses in the urban settings [27,28]. In this study, we highlight a high positivity of CHIKV genomic RNA detection in symptomatic patients in the state of Sergipe during the summer of 2016. This was the first study that showed the circulation of the virus in susceptible population in Sergipe. Moreover, it revealed that the virus was widely distributed in that State, being identified in 11 of the 13 cities sampled.

Table 2

Comparison of individuals with CHIKV infection (viral RNA and/or IgM) with individuals without CHIKV infection in Sergipe-Brazil (n = 142).

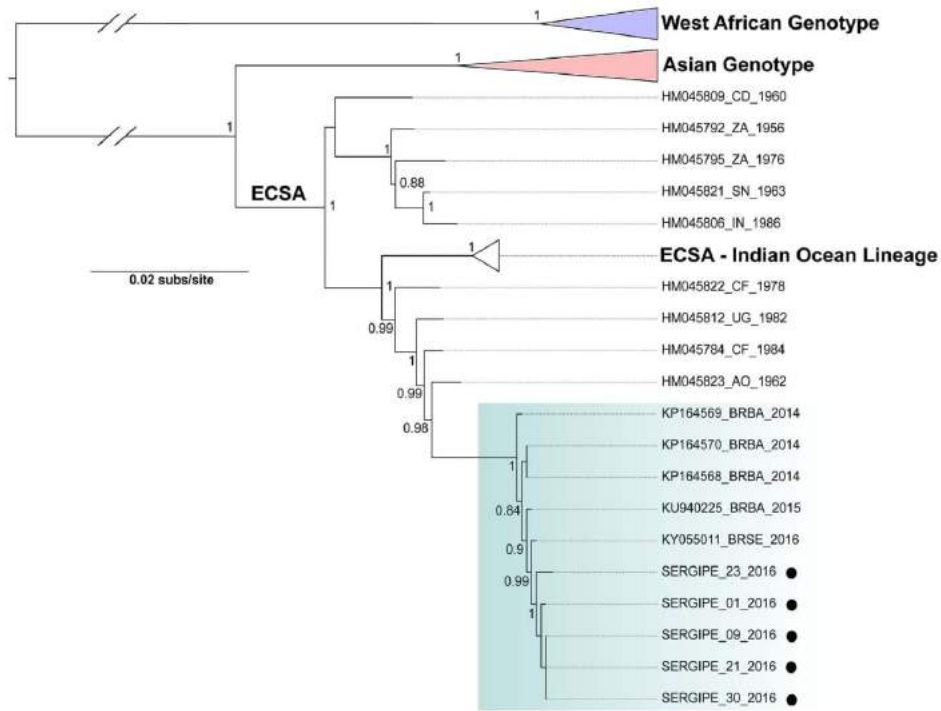|	Positive (n = 107)		Negative (n = 35)		P value*
	n	%	n	%	
Sex					
Female	66	61.7	21	60	1.0
Age groups (years)					
0–15	4	3.7	2	5.7	0.64
> 15–30	48	44.9	16	45.7	1.0
> 30–45	28	26.2	8	22.9	0.82
> 45–60	21	19.6	7	20	1.0
> 60	6	5.6	2	5.7	1.0
Clinical manifestations					
Fever	71	66.4	7	20	< 0.0001
Asthenia	36	33.6	19	54.3	0.04
Joint pain	98	91.6	30	85.7	0.33
Hemorrhagic Signs	6	5.6	0	0	0.33
Headache	93	86.9	33	94.3	0.36
Abdominal pain	11	10.3	6	17.1	0.37
Rash	31	29	7	20	0.38
Signs of Meningoencephalitis	1	0.9	1	2.9	0.43
Itching	23	21.4	10	28.6	0.49
Retro-orbital pain	30	28	12	34.3	0.52
Vomiting	20	18.7	5	14.3	0.62
Myalgia	74	69.2	25	71.4	1.0

*Fisher's exact test.

Fig. 3. Maximum likelihood phylogeny of CHIKV partial E1-3'NTR region sequences.
The dataset consist in 36 CHIKV sequences with: West African genotype (n = 5), Asian genotype (n = 7), ECSA genotype (n = 19) and the sequences generated in this study (n = 5). All sequences used are labeled in the format: accession number/country/year of isolation. Bootstrap scores are shown next to well-supported nodes and the phylogeny was midpoint rooted. The Brazilian sequences belonging to the ECSA genotype are highlighted in blue, and the sequences characterized in the present study from Sergipe-2016 are highlighted with the dot in black color.

It is noteworthy that CHIKV virus caused a major outbreak in Salvador, capital of the State of Bahia at the end of 2015–2016, and apparently, 16.7% of symptomatically infected patients were PCR positive [29]. Moreover, in a seroepidemiological study carried out in the district of Chapada-Bahia, IgM and IgG antibodies were detected in 20% of the sampled individuals, indicating that virus had already spread in that region [30].

The absence of association of infection with the age and gender was characteristic of emerging viruses spreading among an immune-naïve population [12]. Our findings agree with those of Cunha et al. [30], which reported fever among seroconverters in Brazil. Nevertheless, we further observed that fever was the symptom associated with the acute, initial phase of infection, which was also consistently reported by previous studies in other countries [31,32]. Recent studies suggest that during the beginning of the infection, there are differences in gene expression, of several pathways involved in the biosynthesis of molecules related to antiviral responses (e.g., immunological, inflammation, energy deprivation, and joint pain), correlating with symptoms shown at the acute phase of infection [32,33].

Taken together, symptoms that are collectively associated with CHIKV infections were present in the majority of our cases. Crucially, only fever was statistically associated with the infection, while symptoms such as joint pain (98/107), headaches (93/107) and myalgia (74/107) did not show statistical significance. We argue that another agent causing asthenia was co-circulating with CHIKV in these localities.

The phylogenetic characterization of the sequences from the state of Sergipe indicated that they were closely related, given the high bootstrap support, to other isolates characterized during previous years in the Northeast region of Brazil [12,25,28,34]. These results showed that the ECSA genotype is an important genotype associated with the emergence and persistence of the cases of Chikungunya fever in this region of Brazil. Specific and sensitive laboratorial diagnostic of different arboviruses is a goal to be achieved. Nevertheless, it is important to account for clinical signs and symptoms as relevant in helping the attending physician to choose among diagnostic tools [13,32].

Competing interests

None.

Ethical approval

This study is part of an ongoing project for arbovirus research in Sergipe, Brazil approved by the Ethics Committee of the Department of

Microbiology of the Institute of Biomedical Sciences of the University of São Paulo (Protocol 1284/CEPSH – CAAE: 54937216.5.0000.5467).

References

[1] S.C. Weaver, M. Lecuit, Chikungunya virus and the global spread of a mosquito-borne disease, N. Engl. J. Med. 372 (2015) 1231–1239, http://dx.doi.org/10.1056/NEJMra1406035.

[2] M. Dupont-rouzeyrol, V. Caro, L. Guillaumot, M. Vazeille, E. D'Ortenzio, J. Thiberge, N. Baroux, A. Gourinat, M. Grandadam, A.-B. Failloux, Chikungunya virus and the mosquito vector aedes aegypti in New Caledonia (South Pacific Region), Vector-Borne Zoonotic Dis. 12 (2012) 1036–1041, http://dx.doi.org/10.1089/vbz.2011.0937.

[3] M.C. Robinson, An epidemic of virus disease in Southern Province, Tanganyika territory, in 1952–1953, Trans. R. Soc. Trop. Med. Hyg. 49 (1955) 28–32, http://dx.doi.org/10.1016/0035-9203(57)90022-6.

[4] P. Renault, J. Solet, D. Sissoko, E. Balleydier, S. Larrieu, L. Filleul, C. Lassalle, J. Thiria, E. Rachou, H. De Valk, D. Ilef, M. Ledrans, I. Quatresous, P. Quenel, V. Pierre, A major epidemic of chikungunya virus infection on Réunion Island France, 2005–2006, Am. J. Trop. Med. Hyg. 77 (2007) 727–731.

[5] H.C. Hapuarachchi, K.B.A.T. Bandara, S.D.M. Sumanadasa, M.D. Hapugoda, Y.L. Lai, K.S. Lee, L.K. Tan, R.T.P. Lin, L.F.P. Ng, G. Bucht, W. Abeyewickreme, L.C. Ng, Re-emergence of chikungunya virus in south-east Asia: virological evidence from Sri Lanka and Singapore, J. Gen. Virol. 91 (2010) 1067–1076, http://dx.doi.org/10.1099/vir.0. 015743-0.

[6] I. Leparc-Goffart, A. Nougairede, S. Cassadou, C. Prat, X. De Lamballerie, Chikungunya in the americas, Lancet 383 (2014) 514, http://dx.doi.org/10.1016/S0140-6736(14)60185-9.

[7] C. Lahariya, S.K. Pradhan, Emergence of chikungunya virus in Indian subcontinent after 32 years: a review, J. Vector Borne Dis. 43 (2006) 151–160.

[8] G. Rezza, L. Nicoletti, R. Angelini, R. Romi, A.C. Finarelli, M. Panning, P. Cordioli, C. Fortuna, S. Boros, F. Magurano, P. Angelini, M. Dottori, G.C. Majori, A. Cassone, Infection with chikungunya virus in Italy: an outbreak in a temperate region, Lancet 370 (2007) 1840–1846.

[9] J.H. Strauss, E.G. Strauss, The alphaviruses: gene expression, replication, and evolution, Microbiol. Rev. 58 (1994) 491–562.

[10] D. Vassilopoulos, L.H. Calabrese, Virally associated arthritis 2008: clinical, epidemiologic, and pathophysiological considerations, Arthritis Res. Ther. 10 (2008) 1–8, http://dx.doi.org/10.1186/ar2480.

[11] E. Javelle, T. Hua, I. Leparc-goffart, H. Savini, F. Simon, Inflammation of the external ear in acute chikungunya infection: experience from the outbreak in Johor Bahru, Malaysia, 2008, J. Clin. Virol. 59 (2014) 270–273, http://dx.doi.org/10.1016/j.jcv.2014.01.011.

[12] M.R.T. Nunes, N.R. Faria, J.M. de Vasconcelos, N. Golding, M.U. Kraemer, L.F. de Oliveira, R.S.S. Azevedo, D.E.A. da Silva, E.V.P. da Silva, S.P. da Silva, V.L. Carvalho, G.E. Coelho, A.C.R. Cruz, S.G. Rodrigues, J.L.S.G. Vianez Jr, B.T.D. Nunes, J.F. Cardoso, R.B. Tesh, S.I. Hay, O.G. Pybus, P.F.C. Vasconcelos, Emergence and potential for spread of Chikungunya virus in Brazil, BMC Med. 13 (2015) 102, http://dx.doi.org/10.1186/s12916-015-0348-x.

[13] C.W. Cardoso, I.A.D. Paploski, M. Kikuti, M.S. Rodrigues, M.M.O. Silva, G.S. Campos, S.I. Sardi, U. Kitron, M.G. Reis, G.S. Ribeiro, Outbreak of exanthematous illness associated with zika, chikungunya, and dengue viruses, Salvador, Brazil, Emerg. Infect. Dis. 21 (2015) 2274–2276, http://dx.doi.org/10.3201/eid2112.151167.

[14] C.F. Estofolete, A.C.B. Terzian, R. Parreira, A. Esteves, L. Hardman, G.V. Greque, P. Rahal, M.L. Nogueira, Clinical and laboratory profile of zika virus infection in dengue suspected patients: a case series, J. Clin. Virol. 81 (2016) 25–30, http://dx.doi.org/10.1016/j.jcv.2016.05.012.

[15] SVS-MS, Monitoramento dos casos de dengue, febre de chikungunya e febre pelo vírus Zika até a Semana Epidemiológica 1 (2017), p. 2016 ISSN 2358-9450.

[16] T.S.L. Keesen, R.P. De Almeida, B.M. Gois, R.F. Peixoto, A.S.C. Pachá, F.C.F. Vieira, M. Paixão, R. Cazzaniga, R.J. Boyton, D.M. Altmann, Guillain-Barré syndrome and arboviral infection in Brazil, Lancet Infect. Dis. 17 (2017) 693–694, http://dx.doi.org/10.1016/S1473-3099(17)30333-X.

[17] SES-SE, Atualização da situação epidemiológica da Dengue, Chikungunya, Zika e dos casos de Microcefalia em Sergipe, (2016).

[18] D. Wagner, K. De With, D. Huzly, F. Hufert, M. Weidmann, S. Breisinger, S. Eppinger, W.V. Kern, T.M. Bauer, Nosocomial acquisition of dengue, Emerg. Infect. Dis. 10 (2004) 1872–1873, http://dx.doi.org/10.3201/eid1010.031037.

[19] R.S. Lanciotti, O.L. Kosoy, J.J. Laven, J.O. Velez, A.J. Lambert, A.J. Johnson, S.M. Stanfield, M.R. Duffy, Genetic and serologic properties of zika virus associated with an epidemic, Yap State, Micronesia, 2007, Emerg. Infect. Dis. 14 (2008) 1232–1239, http://dx.doi.org/10.3201/eid1408.080287.

[20] P. Patel, A.A. El Wahed, O. Faye, P. Prüger, M. Kaiser, S. Thaloengsok, S. Ubol, A. Sakuntabhai, I. Leparc-goffart, F.T. Hufert, A.A. Sall, M. Weidmann, M. Niedrig, A field-deployable reverse transcription recombinase polymerase amplification assay for rapid detection of the chikungunya virus, PLoS Negl. Trop. Dis. 10 (2016) e0004953, http://dx.doi.org/10.1371/journal.pntd.0004953.

[21] K.A. Stapleford, G. Moratorio, R. Henningsson, R. Chen, S. Matheus, A. Enfissi, D. Weissglas-Volkov, O. Isakov, H. Blanc, B.C. Mounce, M. Dupont-Rouzeyrol, N. Shomron, S. Weaver, M. Fontes, D. Rousset, M. Vignuzzi, Whole-genome sequencing analysis from the chikungunya virus caribbean outbreak reveals novel evolutionary genomic elements, PLoS Negl. Trop. Dis. 10 (2016) e0004402, http://dx.doi.org/10.1371/journal.pntd.0004402.

[22] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. Mcgettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, Clustal W and clustal X version 2.0, Bioinformatics 23 (2007) 2947–2948, http://dx.doi.org/10.1093/bioinformatics/btm404.

[23] A. Larsson, AliView: a fast and lightweight alignment viewer and editor for large datasets, Bioinformatics 30 (2014) 3276–3278, http://dx.doi.org/10.1093/bioinformatics/btu531.

[24] M.N. Price, P.S. Dehal, A.P. Arkin, FastTree 2–approximately maximum-likelihood trees for large alignments, PLoS One 5 (2010) e9490, http://dx.doi.org/10.1371/journal.pone.0009490.

[25] S.I. Sardi, S. Somasekar, S.N. Naccache, A.C. Bandeira, L.B. Tauro, G.S. Campos, C.Y. Chiu, Coinfections of zika and chikungunya viruses in bahia, Brazil, identified by metagenomic next-generation sequencing, J. Clin. Microbiol. 54 (2016) 2348–2353, http://dx.doi.org/10.1128/JCM.00877-16.

[26] J.A.B. Alves, J.R. Santos, E.N. De Mendonça, A.C.F. Abud, M.S. Nunes, R. Fakhouri, A. Dorcas, A.D.D.M. Inagaki, M. Marchioro, A.R. Antoniolli, Epidemiological aspects of dengue in Aracaju, state of Sergipe, Brazil, Rev. Soc. Bras. Med. Trop. 44 (2011) 670–673.

[27] L.S. Marteis, L.M. Steffler, K.C.G.M. de Araújo, R.L.C. dos Santos, Identification and spatial distribution of key premises for Aedes aegypti in the Porto Dantas neighborhood, Aracaju Sergipe State, Brazil, 2007–2008, Cad. Saude Publica. 29 (2013) 368–378.

[28] A.L. Costa-da-silva, R.S. Ioshino, V. Petersen, A.F. Lima, M.P. Cunha, M.R. Wiley, J.T. Ladner, K. Prieto, G. Palacios, D.D. Costa, L. Suesdek, P.M.D.A. Zanotto, M.L. Capurro, First report of naturally infected Aedes aegypti with chikungunya virus genotype ECSA in the Americas, PLoS Negl. Trop. Dis. 11 (2017) e0005630.

[29] C.W. Cardoso, M. Kikuti, Prates A.P.P.B, I.A.D. Paploski, L.B. Tauro, M.M.O. Silva, P. Santana, M.F.S. Rego, G. Reis, U. Kitron, G.S. Ribeiro, Unrecognized emergence of chikungunya virus during a zika virus outbreak in salvador, Brazil, PLoS Negl. Trop. Dis. 11 (2017) e0005334, http://dx.doi.org/10.1371/journal.pntd.0005334.

[30] R.V. Cunha, K.S. Trinta, C.A. Montalbano, M.V.F. Sucupira, M.M. De Lima, E. Marques, I.H. Romanholi, J. Croda, Seroprevalence of chikungunya virus in a rural community in Brazil, PLoS Negl. Trop. Dis. 11 (2017) e0005319, http://dx.doi.org/10.1371/journal.pntd.0005319.

[31] S.P. Manimunda, P. Vijayachari, R. Uppoor, A.P. Sugunan, S.S. Singh, S.K. Rai, A.B. Sudeep, N. Muruganandam, I.K. Chaitanya, D.R. Guruprasad, Clinical progression of chikungunya fever during acute and chronic arthritic stages and the changes in joint morphology as revealed by imaging, Trans. R. Soc. Trop. Med. Hyg. 104 (2010) 392–399, http://dx.doi.org/10.1016/j.trstmh.2010.01.011.

[32] J. Shrinet, J.S. Shastri, R. Gaind, N.S. Bhavesh, S. Sunil, Serum metabolomics analysis of patients with chikungunya and dengue mono/co-infections reveals distinct metabolite signatures in the three disease conditions, Sci. Rep. 6 (2016) 1–12, http://dx.doi.org/10.1038/srep36833.

[33] J.A.C. Wilson, N.A. Prow, W.A. Schroder, J.J. Ellis, E. Cumming, L.J. Gearing, Y.S. Poo, A. Taylor, P.J. Hertzog, F. Di Giallonardo, L. Hueston, R. Le Grand, B. Tang, T.T. Le, J. Gardner, S. Mahalingam, P. Roques, P.I. Bird, A. Suhrbier, RNA-Seq analysis of chikungunya virus infection and identification of granzyme A as a major promoter of arthritic inflammation, PLoS Pathog. 13 (2017) e1006155, http://dx.doi.org/10.1371/journal.ppat.1006155.

[34] A.C. da Costa, J. Thézé, S.C.V. Komninakis, R.L. Sanz-Duro, M.R.L. Felinto, L.C.C. Moura, I.M.O. Barroso, L.E.C. Santos, M.A.L. Nunes, A.A. Moura, J. Lourenço, X. Deng, E.L. Delwart, M.R.A.S. Guimarães, O.G. Pybus, E.C. Sabino, N.R. Faria, Emergence and persistence of the Chikungunya virus East-Central-South-African genotype in Northeastern Brazil, bioRxiv (2017) 111948.

APPENDIX D - Submitted Manuscript

Clinical Infectious Diseases
A cryptic outbreak of Human Parvovirus B19 hidden by Dengue fever
--Manuscript Draft--

Manuscript Number:	CID-89222
Full Title:	A cryptic outbreak of Human Parvovirus B19 hidden by Dengue fever
Short Title:	Parvovirus B19 outbreak hidden by dengue
Article Type:	Major Article
Corresponding Author:	Paolo Marinho de Andrade Zanotto, D.Phil. Biomedical Sciences Institute - ICB II Sao Paulo, Sao Paulo BRAZIL
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Biomedical Sciences Institute - ICB II
Corresponding Author's Secondary Institution:	
First Author:	Nicholas Di Paola
First Author Secondary Information:	
Order of Authors:	Nicholas Di Paola
	Flávio S Mesquita
	Danielle Bruna Leal de Oliveira
	Christian Julian Villabona-Arenas
	Shahab Zaki Pour
	Carla de Sousa-Capra
	Karime Balarini
	Celso Ricardo Theoto Pereira da Fonseca
	Paolo Marinho de Andrade Zanotto
Order of Authors Secondary Information:	
Manuscript Region of Origin:	BRAZIL
Abstract:	Seasonal outbreaks of Dengue often result in hundreds of dengue-suspected cases where a clinical diagnosis cannot be confirmed. Usually, during large outbreaks of Dengue and other pathogens that can cause acute-febrile illnesses, the search for secondary pathogens with similar disease outcomes is rare. Using Total RNA sequencing, we discovered a cryptic outbreak of Parvovirus B19 in dengue-suspected patients that occurred from November 2013 to February 2014. Of the 182 cases investigated, 63% were viremic for the B19 virus. Moreover, we found that over 43% of infected patients had no serological evidence of prior infection. Parvovirus B19 is a typical childhood infection, yet we observed that 82% of the infected patients were adults. Additionally, we perceived that infected adults had significantly higher presentations of myalgia than in children. We also obtained VP1/VP2 gene nucleotide sequences from 43 patients. Our results support the utility of next generation sequencing for symptomatic patients with unknown etiologies during seasonal outbreaks of Dengue and other arborviruses. Our findings could improve the vigilance of hospitals and laboratories by raising awareness to co-circulating pathogens like Parvovirus B19 that may be hidden in plain sight.
Suggested Reviewers:	Neal S Young youngns@nhlbi.nih.gov Considered an expert of Parvovirus B19

	Erik D Heegaard e.heegaard@immi.ku.dk Has written several articles and reviews about Parvovirus B19
	Larry J Anderson larry.anderson@emory.edu Specializes in the clinical aspects of Parvovirus B19
	Jianming Qiu jqiu@kumc.edu Recently wrote a large review on parvoviruses, particularly Parvovirus B19
Opposed Reviewers:	

Cover Letter

INSTITUTO DE CIÊNCIAS BIOMÉDICAS-USP

Excelência em Ensino e Pesquisa

Homepage: www.icb.usp.br

DEPARTAMENTO DE
MICroBiologia
UNIVERSIDADE DE SÃO PAULO

Cidade Universitária – Ed. Biomédicas II
Av. Prof. Lineu Prestes, 1374 – Cep. 05508-900 São Paulo, SP -
Brasil
Telefone :(55) (11) 3091-7259 – telefax : (55) (11) 3091 7354
Homepage: www.icb.usp.br/bmm
e-mail: microsec@icb.usp.br

São Paulo, January 17th, 2018.

Dear *Clinical Infectious Diseases* Editors,

We are submitting the manuscript entitled "**A cryptic outbreak of Human Parvovirus B19 hidden by Dengue fever**" by Di Paola *et al.* to *Clinical Infectious Diseases*.

In this manuscript, we report the first outbreak of Parvovirus B19 coinciding with an outbreak of dengue fever. We investigated 183 dengue-suspected cases and found that 63% were viremic for Parvovirus B19. Parvovirus B19 infection is a typical childhood illness, yet we observed that 82% were adults. Interestingly, infected adults had significantly higher presentations of myalgia than in children. We also obtained VP1/VP2 gene sequences from 43 patients.

We believe that our findings may be of particular interest to your readers since we provide an example that a medically important virus like Parvovirus B19 can be neglected by passive surveillance systems during a dengue aoutbreak. Our findings support the need for greater attention in surveillance systems to viruses showing similar clinical presentations.

Yours faithfully,

Prof. Paolo Zanotto, *D.PHIL.*

Laboratory of Molecular Evolution & Bioinformatics - LEMB
Department of Microbiology

Biomedical Sciences Institute - ICB II
University of Sao Paulo - USP
Av. Prof. Lineu Prestes, 1374
Sao Paulo - SP - Brazil - 05508-000
Phone: +55 11 3091-8453
FAX: +55 11 3091-7354
E-mail: pzanotto@usp.br

P.S. All authors have seen and approved the manuscript, contributed significantly to the work. The manuscript has not been previously published nor is not being considered for publication elsewhere. We agree to cover the costs of reproducing our two color figures.

Suggested Reviewers:

Neal S. Young
youngns@nhlbi.nih.gov
Hematopoiesis and Bone Marrow Failure Laboratory, NIH

Erik D. Heegaard
e.heegaard@immi.ku.dk.
Departments of Clinical Microbiology and Pathology
University State Hospital
Rigshospitalet, Copenhagen, Denmark

Larry J. Anderson
larry.anderson@emory.edu
Division of Infectious Disease
Department of Pediatrics
Emory University School of Medicine

Jianming Qiu
jqiu@kumc.edu
Department of Microbiology
Molecular Genetics and Immunology
University of Kansas Medical Center

1   **A cryptic outbreak of Human Parvovirus B19 hidden by Dengue fever**

2

3   Nicholas Di Paola[1#], Flávio S Mesquita[1], Danielle Bruna Leal de Oliveira[1], Christian

4   Julián Villabona-Arenas[2], Shahab Zaki Pour[1], Carla de Sousa-Capra[3], Karime Balarini[4],

5   Celso Ricardo Theoto Pereira da Fonseca[4] and Paolo Marinho de Andrade Zanotto[1*]

6

7   [1]Department of Microbiology, Institute of Biomedical Sciences, University of Sao Paulo,

8   Brazil; [2]Laboratoire TransVIHMI, Institut de Recherche pour le Développement, and

9   University of Montpellier, Montpellier, France; [3]Office of Epidemiological Surveillance,

10   Department of Health of Guarujá, Guarujá, São Paulo, Brazil; [4]Clinical Laboratory

11   Analysis Center, ITAPEMA, Guarujá, Brazil

12

13   *Corresponding author: pzanotto@usp.br

14   #Alternate corresponding author: nicholasdipaola@gmail.com

15

16   **Keywords:** Parvovirus B19, hidden outbreak, next generation sequencing, dengue fever,

17   passive surveillance

18

19   **Running title:** Parvovirus B19 outbreak hidden by dengue

20   **Author summary:** We investigated symptomatic patients suspected of dengue fever and

21   discovered a hidden outbreak of Parvovirus B19. Our results bring attention to the

22   limitations of passive surveillance systems, where medically important viruses could be

23   neglected.

24 **Abstract.**

25   Seasonal outbreaks of Dengue often result in hundreds of dengue-suspected cases

26   where a clinical diagnosis cannot be confirmed. Usually, during large outbreaks of

27   Dengue and other pathogens that can cause acute-febrile illnesses, the search for

28   secondary pathogens with similar disease outcomes is rare. Using Total RNA sequencing,

29   we discovered a cryptic outbreak of Parvovirus B19 in dengue-suspected patients that

30   occurred from November 2013 to February 2014. Of the 182 cases investigated, 63%

31   were viremic for the B19 virus. Moreover, we found that over 43% of infected patients

32   had no serological evidence of prior infection. Parvovirus B19 is a typical childhood

33   infection, yet we observed that 82% of the infected patients were adults. Additionally, we

34   perceived that infected adults had significantly higher presentations of myalgia than in

35   children. We also obtained VP1/VP2 gene nucleotide sequences from 43 patients. Our

36   results support the utility of next generation sequencing for symptomatic patients with

37   unknown etiologies during seasonal outbreaks of Dengue and other arboviruses. Our

38   findings could improve the vigilance of hospitals and laboratories by raising awareness to

39   co-circulating pathogens like Parvovirus B19 that may be hidden in plain sight.

40

41 **Introduction**

42   National passive surveillance systems generally rely on weekly clinical reports to

43   administrators from a large network of hospitals and laboratories [1]. However, high

44   costs, poor clinical diagnostic accuracy, and general underreporting can limit their

Introduction

41

42   National passive surveillance systems generally rely on weekly clinical reports to

43 administrators from a large network of hospitals and laboratories [1]. However, high

44 costs, poor clinical diagnostic accuracy, and general underreporting can limit their

45 effectiveness. As an example, a previous study found that for every 20 dengue patients,

46 only 1 had been reported to Brazil's Notifiable Diseases Information System (SINAN) as

47 having dengue [2]. Most Dengue diagnoses are made on clinical and epidemiological

48 grounds where mildly symptomatic and non-specific febrile cases are often discarded [2–

49 4].

50   Few epidemiological studies have investigated or assessed the simultaneous

51 occurrence of different pathogens with similar disease outcome [5–8]. In this case,

52 passive surveillance underestimates Dengue's true rate of infection and ignores the

53 possibility of infection from related viruses *i.e.,* Mayaro, Saint-Louis encephalitis or Zika,

54 but also for distantly related viruses such as Parvovirus B19 (B19V) [5,9,10]. B19V

55 infections have been associated with both mild (asymptomatic, Fifth disease, febrile non-

56 specific flu-like illness, rash) and severe (chronic arthralgia, transient aplastic crisis,

57 anemia, hepatitis, myocarditis, and autoimmune syndromes) clinical manifestations [11].

58   Acute parvovirus B19 infections are characterized by two phases of disease. The

59 first phase of illness consists of a prodromal non-specific febrile illness that can present

60 with fever, myalgia, headache, and chills [12,13], which is a similar disease outcome in

61 Dengue-infected patients [14]. However, the Parvovirus B19 disease is usually associated

62 with and clinically diagnosed by the second phase of the illness, which may include the

63 hallmark presentations of erythema infectiosum (*i.e.,* Fifth disease), various cutaneous

64 and popular rashes, and arthralgia (joint pain, arthritis) [13]. Generally, B19 disease is an

65 asymptomatic or mild pediatric illness with far fewer diagnosed cases in adults; previous

66 childhood infections are thought to provide life-long immunity. However, infections in

67 immunosuppressed individuals and patients with other medical complications can be

68 susceptible to a range of severe disease manifestations and relapsing B19V infections

69 [15,16]. The spread and infection of B19V can occur through respiratory secretions,

70 blood-derived products, vertically (from mother to child) and nosocomial transmission

71 [17–21].

72   Here, we investigated dengue-suspected cases and report the cryptic outbreak of

73 Human Parvovirus B19 during a dengue virus outbreak in Guarujá, a touristic city

74 embedded in Brazil's South-Atlantic rainforest and disposed to seasonal dengue

75 outbreaks [22–24]. Guarujá is adjacent to Santos, the largest seaport in Brazil with

76 intense daily commuting with the city of São Paulo, the largest conurbation in South

77 America. We found that 63.2% of patients were viremic for B19V and many had clinical

78 outcomes indistinguishable to dengue fever. It was previously unknown that B19V was

79 compromising dengue diagnostics and our findings provide a warning that there may be

80 other, more severe pathogens could be missed if passive surveillance systems remain as

81 they are.

82

83 **Results**

84 **RNA Sequencing of Pooled Serum Samples**

85   From November 2013 to February 2014 during a dengue outbreak, 182 serum

86 samples were collected from patients who visited the city's public health clinics with

87   symptoms of an acute febrile illness, with some (34.2%) diagnosed with dengue fever but

88   many were undiagnosed. We found that in all these cases, there was no evidence for

89   DENV infection or any other nationally prevalent arboviruses such as Zika virus,

90   Chikungunya virus, Mayaro virus or Yellow Fever virus. From here, samples were

91   grouped into 24 pools and were subjected to total RNA sequencing to screen for other

92   viruses with a similar disease outcome to Dengue. Using reads from individual pools, we

93   were able to *de novo* assemble a complete or nearly complete Parvovirus B19 genome

94   from 10 of the 24 pools (see Supplementary Figure 1). Albeit at a lower frequency, we

95   also found sequences from GB virus C, Hepatitis C virus, Hepatitis A virus, and many

96   different groups of phages and retroviruses (see Supplementary Material).

97

98   **PCR amplification of Parvovirus B19**

99   Once we identified B19V we designed specific primers for the VP1/VP2 region of

100   B19V. Of the 182 patients, 115 (63.2%) were PCR positive for the B19 virus (Table 1).

101   The median age of these patients was 27. In fact, nearly 60% of the patients were

102   between 18 and 39 years of age.

103   From the patients where DNA was amplified by PCR, eight were also positive for

104   the B19V IgM antibody. Only four samples were PCR negative and IgM positive. A total

105   of 50 patients (27.5%) with B19V DNA were negative for both B19V IgM and IgG,

106   suggesting a first time exposure. Additionally, 47 (25.8%) patients were PCR positive,

107   IgM negative and IgG positive, which is a serological profile that has been associated to a

108   prior, persistent or a reoccurring infection [25–28]. Only 23 patients (12.6%) were

109   negative for all three diagnostic assays for B19V.

110   As contaminations have been reported as a problem for B19V diagnostics, we

111   took precautions to ensure the reliability of our results. For each sample, PCR assays

112   were performed in duplicate and on separate days to avoid false-positive DNA

113   amplifications. Prior to and upon completion of both assays, we sterilized all materials

114   and used new reagents as well as PCR specific negative controls for each PCR run and

115   associated gel. We also tested 35 serological samples of febrile patients from another

116   Brazilian state, Sergipe, that were suspected of Chikungunya [29], but were found

117   negative for both B19V PCR and IgM antibodies.

118   We then retrieved the patient's date of attendance at a public clinic from 111 of

119   the 115 patients positive for B19V DNA. Cases ranged from November 4th, 2013 to

120   February 25th, 2014. We mapped the B19V positive patient's date of attendance by week

121   to discover whether was the time where case incidence was highest (Figure 1A). Peak

122   incidence occurred during the week of November 10th, 2013. We also obtained the

123   addresses for 68 of the PCR positive patients and mapped the distribution of cases around

124   the city (Figure 1C). We found that the majority of cases were in the Santo Antônio

125   district where the main administrative, mass transportation and commercial facilities are

126   located. Highly populated districts as Morrinhos (6 cases), Enseada (7 cases) and Vila

127   Zilda (7 cases) also had a large proportion of localized cases.

128

129   **Clinical presentations of B19 positive patients**

130   For 70 patients, we were able to retrieve a more detailed medical record with a

131   range of incomplete information including symptoms, hematology, complications,

132   diagnoses, and drug recommendations. Several medical records were discarded because

133 they were either lacking data or were completely illegible. However, all patients included

134 in our study presented one or more dengue-like symptom(s), but not all medical records

135 included patient's symptoms during their attendance (see footnote of Table 1).

136 Nevertheless, we annotated the symptoms and complications presented from 56 patients

137 who were also viremic for B19V during their clinical attendance (Table 2 and Table 3).

138 Fever (70%), myalgia (61%) and headaches (43%) were the most prevalent clinical

139 presentations in patients. Only 12 of these patients (21%) had presentations of arthralgia

140 (6 patients) or rash (6 patients). Eight patients (14.3%) described retro-orbital pain at

141 their attendance, a symptom to the best of our knowledge that has no current association

142 with acute or persistent parvovirus B19 infection.

143   We also obtained data from patients with complications such as vomiting and

144 diarrhea. Additionally, 25 medical records included patient platelet count information that

145 helped us determine the prevalence of thrombocytopenia (platelets < 150,000/mm^3) in

146 viremic patients. In total, 15 patients (60%) presented with thrombocytopenia during their

147 clinical attendance.

148   Because people can be persistently infected by B19V, the sole existence of B19V

149 DNA in a patient's serum can be misleading for a clinical diagnosis of an acute infection

150 without any supporting clinical information, especially if paired sampling or a

151 quantitative PCR or ELISA analysis were not done [30–33]. The presence of the B19V

152 IgM antibody is a strong indicator of an acute infection [34,35]. Therefore, we looked to

153 see if these patients presented different symptoms in relation to their serological profile

154 (Table 2) by analyzing the occurrence of patient symptoms and clinical complications

155 with their serology. Only two patients with a detailed medical record were IgM positive.

156   Nevertheless, 22 patients (with medical records) were IgM negative and IgG positive,

157   which could indicate a persistent or reoccurring infection [34,36,37]. The other 24

158   patients (with medical records) were negative for both antibodies, suggesting a first-time

159   and recent infection. Given that 50.4% of our (PCR only or PCR and IgM positive)

160   patients appeared to be acutely infected it is fair to assume that we sampled from an

161   ongoing B19V outbreak. We did not find any significant differences between serological

162   profiles and clinical presentations.

163   Next we compared clinical presentations by the age and sex of the patient (Table

164   3). We grouped patients by those who were 5-17 years of age (children) and those who

165   were over 18 (adults) at the time of their attendance. Notably, we found that the

166   prevalence of myalgia was significantly higher in adult patients compared to younger

167   patients (Fisher's exact test, p-value = 0.02). Likewise, we also observed that only adults

168   presented with symptoms of arthralgia, odynophagia (difficulty swallowing), and

169   diarrhea. In disagreement to some previous works [38,39], we did not see any significant

170   differences in symptoms between men and women. However, we did find that only men

171   reported having diarrhea at their clinical attendance (Fisher's exact test, p-value = 0.02).

172

173   **Phylogeny of Parvovirus B19**

174   To explore the genetic diversity of B19V in Guarujá, we amplified a small region

175   (sequences ranged from 427 to 560 nucleotides in length) of the VP1/VP2 gene that has

176   been used for genotyping in other works [40,41]. We then aligned our sequences

177   (Genbank accession numbers: MG765329-MG765371) with all publically available

178   B19V sequences with the country of isolation and collection date available from

179   Genbank. Our phylogenetic estimation showed that our sequences (orange tips) are most

180   similar to the B19V Genotype 1 sequences (Figure 2). It is possible that there was more

181   than one emergence of B19V in Guarujá. Our sequences were genetically similar to two

182   sequenced isolates from the city of São Paulo: one from 2010 (Genbank accession:

183   KC013305) and one from 2009 (Genbank accession: KC013344). We also indicated

184   (violet tips) all published Brazilian and non-Brazilian sequences (light blue tips) of

185   B19V, which are widely spread across Genotype 1 and Genotype 3B. Genotype 2

186   sequences from Brazil are only available for the NS1 gene.

187

188   **Discussion**

189   In this study, we retrospectively studied dengue-suspected cases and found a large

190   prevalence (63.2%) of B19V in patient's serum. Despite several seroprevalence studies

191   [42,43] and some sporadic cases in the state of Rio de Janeiro [44], the public attention

192   for Parvovirus B19 in Brazil is limited. To our knowledge, B19V is rarely tested for or

193   even diagnosed in public health centers. Moreover, its febrile prodromal disease (*i.e.*, pre-

194   rash and pre-arthritis symptoms) is often mistaken for more prevalent pathogens such as

195   dengue.

196   Dengue infections occur all year round in Brazil, but explosive outbreaks tend to

197   happen during the summer months of November to April [22–24]. A high number of

198   dengue cases lead to clinical diagnoses based on epidemiological grounds, which we

199   showed here can lead to cryptic outbreaks of pathogens with similar disease outcomes.

200   We found that during the 2013/2014 B19V outbreak in Guarujá, infections peaked in the

201   late spring and early summer months of November through January (Figure 1), which

202   agreed with other reports of erythema infectiosum outbreaks [35,39,45–47], and occurred

203   in a similar time frame of locally reported dengue outbreaks [22,24]. Furthermore,

204   clinical and hematological manifestations of B19V infected patients were also found to

205   be similar to classical dengue disease presentations like fever, headache, myalgia, and

206   retro-orbital pain [14]. This is especially true for the eight patients who presented with

207   retro-orbital pain—a symptom that has no known association with the Parvovirus B19

208   disease—and with thrombocytopenia, a common feature and diagnostic marker of

209   dengue.

210   Since Parvovirus B19's discovery, diagnostic methods for detecting acute

211   infections in serum have involved electron microscopy, direct DNA hybridization,

212   antigen enzyme-linked immunosorbent assays (ELISA) *i.e.* detection of IgM, IgG, IgA,

213   and NS1-specific antibodies, PCR and hemagglutination [36,47–57]. Detection of B19V

214   DNA in serum is circumstantial evidence for an infection, but cannot alone determine

215   whether the viremia in question is acute or persistent because of its prolonged presence

216   after infection [34,58]. The lack of any IgM or IgG antibodies in 50 symptomatic cases

217   could be a strong indication for a first time infection, which is surprising since the

218   majority of patients were adults and B19V exposure typically happens during childhood

219   [59]. Moreover, we only detected IgM antibodies in 12 cases, which is much less than the

220   number of PCR positive cases. In these cases, it is possible that the majority of patients

221   were attended very early in their illness and before their IgM antibody response could be

222   detected [13,21].

223   We also found 47 symptomatic patients who were PCR and IgG positive, but IgM

224   negative. We are aware of four explanations that could account for this: *i*) an acute

225 infection where the patient could not elicit an IgM response, *ii*) a prolonged acute

226 infection that lingered after the disappearance of IgM antibodies and was not completely

227 neutralized by IgG [13], *iii*) a persistent infection where a secondary viral reactivation or

228 relapsing infection led to secondary clinical manifestations [25–28], and *iv*) the presence

229 of low levels of B19V DNA from a previous infection that have no correlation to the

230 current clinical manifestations [34]. Even with clinical records for some patients, we were

231 uncertain how to determine which of the above-mentioned possibilities were most likely.

232 None of the patients in this study were pregnant or younger than 5 years of age.

233 However, it is of high concern that B19V infections in pregnant women are linked to

234 hydrops fetalis (*i.e.*, fluid collects in two or more fetal compartments), miscarriages,

235 microcephaly, hydrocephalus, contractures, and intracranial and hepatic calcification

236 [19,60]. Nevertheless, few studies have investigated the prevalence of B19V in babies

237 with birth defects suspected of ZIKV congenital syndrome in Brazil [61,62]. This is

238 especially worrisome as congenital infections of ZIKV and B19V can both cause hydrops

239 fetalis and microcephaly [19,60,63,64]. Furthermore, Brazil harbors all three genotypes

240 of *Erythroparvovirus*, as well as recombinant variants, attesting to the large endemic

241 diversity within the country and abundant case studies of severe complications associated

242 with the virus (Figure 2) [40,41,65].

243 Next generation sequencing approaches have led to better awareness and

244 detection of several pathogens; its success in accomplishing this has been demonstrated

245 over the last decade [5,66–68]. In this case, it allowed us to discover what underlying

246 etiological agents were behind a large number of sick patients. This outbreak served as an

247 example that current passive surveillance measures could be vulnerable to missing a mild

248 pathogen like B19V, which in turn, also raises concerns that we could be overlooking

249 more severe pathogens too. Here, we found high amounts of B19V in dengue-suspected

250 patients where no diagnosis could be previously confirmed. Significantly, we found that

251 despite a negative DENV diagnosis, 34% of patients with medical records were still

252 clinical diagnosed for dengue (Table 1). In one case, an official case notification for

253 dengue without any diagnostic evidence was found on the patient's medical record. This

254 further illustrates that even known viruses could circulate camouflaged underneath

255 diagnostic detection. Although relatively expensive, we advocate for the situational use

256 of NGS for symptomatic but negatively diagnosed patients during large seasonal

257 outbreaks of dengue or other febrile illnesses. Our findings provide an example that a

258 medically important virus like Parvovirus B19 can be neglected by passive surveillance

259 systems during a dengue outbreak. Crucially, it demands greater attention in surveillance

260 systems to viruses showing similar clinical presentations.

261

262 **Materials and Methods**

263 **Study site.** The coastal city of Guarujá (23°59′37″S 46°15′23″W) is located in the

264 southeastern region of the state of São Paulo, Brazil (Figure 1B). The city is surrounded

265 by a tropical rain forest that results in the city's high humidity, precipitation, and

266 temperature. Annual temperatures average at 24.7 °C, with a minimum of 18 °C and a

267 maximum of 31.3 °C. Precipitation is highest during the summer in the month of

268 February (average rainfall of 413 mm) and lowest during the winter in the month of

269 August (average rainfall of 156 mm). Annual rainfall is ~3,400 mm. The estimated

270 population in 2013 was 306,683 and the human population density was ~2,000

271   inhabitants per km^2. The city's main economic sources are acquired through the

272   commercial use of its seaports and from tourism during the summer months.

273

274   **Sample collection.** Samples were collected from November 10th, 2013, to February 25th

275   of 2014. We collected 182 serum samples from patients who showed signs of acute-

276   febrile illness, but tested negative for all dengue-specific ELISA and PCR assays.

277   Patients of any age with symptoms and clinical signs of dengue disease were examined in

278   Primary Health or Emergency Care Units. Patients were included who presented with the

279   following symptoms and complications: fever, joint-pain (arthralgia), headache, rash,

280   retro-orbital pain, muscle pain (myalgia), difficulty swallowing (odynophagia), vomiting,

281   diarrhea, and thrombocytopenia.

282

283   **Next-generation sequencing.** Viral RNA was extracted from the nasopharyngeal

284   aspirates using the QIAamp Viral RNA Mini Kit (Qiagen; Valencia, CA), and purified

285   with DNAse I and concentrated using the RNA Clean & ConcentratorTM-5 Kit (Zymo

286   research; Irvine, CA). The paired-end RNA libraries were constructed and validated

287   using the TruSeq Stranded Total RNA HT Sample Prep Kit (Illumina; San Diego, CA)

288   was used to construct and validate paired-end RNA libraries. Sequencing was done at the

289   Core Facility for Scientific Research – University of São Paulo (CEFAP-USP/GENIAL)

290   using the Illumina NextSeq platform. Each sample was barcoded individually, which

291   allowed separation of reads for each patient. Short, unpaired reads/bases, and low-quality

292   reads were removed using Trimmomatic v0.36 [69]. Paired-end reads (Qphred>33) were

293   assembled *de novo* with SPAdes 3.10 using default parameters [70]. Contigs over 500 nts

294 (nucleotides) were searched against the viral nucleotide refseq Genbank repository using

295 BLAST.

296

297 **Molecular diagnostic assays for Parvovirus B19.** Samples were extracted on the

298 NUCLISENS easyMag platform (bioMerieux). The eluate, containing RNA and total

299 DNA, was stored in a freezer at -70 °C until the time of its use. PCR, followed by a semi-

300 nested reaction was performed targeting the VP1/VP2 region using primers designed by

301 Durigon *et al*., 1996 and Erdman *et al*., 1993 [71,72]. The two primers used in the PCR

302 were P12f (nt 4127-4148; 5'- CAGCCATACCACCACTGGGACA-3') and P17r (nt

303 4824-4803; 5'- TTACGCATCCTGGCTGAGGGCA-3'), with a fragment of

304 approximately 697 base pairs; semi-nested primers were P13f (nt 4214-4237; 5'-

305 GACAAAGAGTATCAGCAAGGAGTG-3') and P17r (nt 4824-4803; 5'-

306 TTACGCATCCTGGCTGAGGGCA-3'), with final product of approximately 610 base

307 pairs. PCR and the semi-nested reactions were performed using a reaction mixture

308 containing 5.0 µL of DNA, 5.0µL of PCR buffer reaction 10× concentration (50 mM

309 Tris-HCl [pH 9.0], 1.0µL of $MgCl_2$ (50 mM KCl, 20 mM [$(NH_4)_2SO_4$]), 1.0µL of 2.5

310 mM of each dNTPs, 1.0µL of 10pmol of each primer, 0.3µL of 5U/µL of taq DNA

311 polymerase (platinum taq DNA polymerase, Invitrogen, CA, USA) and water, resulting

312 in a final volume of 50µL. The amplification was performed in a thermocycler with the

313 following conditions: a cycle of 94 °C for 5 min for the activation of Hf, followed by 40

314 cycles of denaturation at 94 °C for 30 seconds, Hybridization at 55 °C for 30 seconds and

315 extension at 72 °C for 45 seconds and finally a final extension cycle at 72 °C for 7

316 minutes. The semi-nested PCR reaction was performed using the same PCR conditions.

317   As control, a well-known sample was used, and DNA/RNAse free water was used as the

318   negative control. Detection of the PCR/semi-nested amplified product was performed by

319   2% agarose gel electrophoresis (Life Techonologies corporation) in TAE buffer [0.5x],

320   stained with 0.5µg/ml ethidium bromide. In 3.0µL of bromophenol blue, 10.0µL of the

321   sample was mixed and submitted to electrophoresis in a horizontal vessel for 60 minutes

322   at 100V.

323

324   **Sanger sequencing.** Samples that were found positive for B19V using a semi-nested

325   PCR were selected for downstream sequencing. Purification of the PCR product was

326   performed with the ExoSap kit (exonuclease I - Amersham Pharmacia Biotech),

327   according to the manufacturer's instructions. After purification of the semi-nested

328   product, the amplified VP1/VP2 region using P13f/P17r primers was sequenced using the

329   "ABI PRISM Dye™ Terminator Cycle Sequencing Ready Reaction kit" ("Big Dye" -

330   Applied Biosystems, Inc., USA), using 3.0µL of BigDye® Terminator v1.1/3.1

331   Sequencing Buffer (5X), 1.0µL of Terminator Ready Reaction Mix and 1.0µL of the

332   primer. The enzymatic extension was performed on GeneAmp PCR System 9700

333   thermocycler for 25 cycles of 96 °C for 10 seconds for denaturation of template DNA, 50

334   °C for 10 seconds for primer annealing and 60°C for 4 minutes for extension. The

335   obtained product was further purified with the X-terminator kit (Applied Biosystems Inc.,

336   USA) according to the manufacturer's instructions. The purified product was subjected to

337   POP7 polymer electrophoresis using ABI model 3100 automatic sequencer (Applied

338   Biosystems, Inc., USA). Sequence reads were assembled, manually inspected and

339   trimmed using Geneious version 9.2 [73]. Sequences were submitted to Genbank and can

340   be found under the accession numbers MG765329-MG765371.

341

342   **Patient Medical Records.** Patient records were retrospectively collected with the kind

343   aid of the Guarujá Secretary of Health. From the 182 total serum samples collected, only

344   111 records that included basic patient information were available. Only 70 contained

345   diagnostic information and 56 contained a physician's medical records from all Guarujá

346   public healthcare clinics. Only 25 had usable hematology data. All adult subjects

347   included in this study provided an informed written consent, and a parent or guardian of

348   any child participant provided the written informed consent on their behalf.

349

350   **Phylogenic Analysis.**

351   To infer the phylogenetic relationships of our Parvovirus B19 sequences with

352   other partial VP1/VP2 genome sequences, we downloaded all available sequences from

353   Genbank that had country and date of the samples isolation. Sequences were aligned

354   using Clustal Omega version 1.2.1 (http://www.clustal.org/omega) and manually curated

355   by using JalView version 1.18-B8 (http://www.jalview.org/download). A maximum-

356   likelihood tree was exhaustively estimated in the nucleotide general time-reversible with

357   gamma-distribution rate variation and invariant sites model. Support for the tree was

358   accessed after 10,000 nonparametric boostrap replicates with FastTree version 2.1.8 [74].

359

360   **Ethics statement**

361   The Ethical Review Board of the Biomedical Science Institute at University of

362   São Paulo approved this study (Statement 933/CEP). All adult subjects provided an

363   informed written consent, and a parent or guardian of any child participant provided the

364   written informed consent on their behalf.

365

378

379   **References.**

380   1.   World Health Organization. National Passive Surveillance. 2017. Available at:
381   http://www.who.int/immunization/monitoring_surveillance/burden/vpd/surveillanc
382   e_type/passive/en/. Accessed 28 December 2017.
383   2.   Silva MMO, Rodrigues MS, Paploski IAD, et al. Accuracy of Dengue Reporting
384   by National Surveillance System, Brazil. Emerg. Infect. Dis. J. **2016**; 22:336.
385   Available at: http://wwwnc.cdc.gov/eid/article/22/2/15-0495.
386   3.   Nunes-Araújo FRF, Ferreira MS, Nishioka SDE a. Dengue fever in Brazilian
387   adults and children: assessment of clinical findings and their validity for diagnosis.
388   Ann. Trop. Med. Parasitol. **2003**; 97:415–9. Available at:
389   http://www.ncbi.nlm.nih.gov/pubmed/12831527. Accessed 7 May 2014.
390   4.   Teixeira MG, Siqueira JB, Ferreira GLC, Bricks L, Joint G. Epidemiological
391   trends of dengue disease in Brazil (2000-2010): a systematic literature search and
392   analysis. PLoS Negl. Trop. Dis. **2013**; 7:e2520. Available at:
393   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3871634&tool=pmcen
394   trez&rendertype=abstract. Accessed 5 May 2014.
395   5.   Yozwiak NL, Skewes-Cox P, Stenglein MD, Balmaseda A, Harris E, DeRisi JL.
396   Virus identification in unknown tropical febrile illness cases using deep
397   sequencing. PLoS Negl. Trop. Dis. **2012**; 6:e1485. Available at:
398   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3274504&tool=pmcen
399   trez&rendertype=abstract. Accessed 6 August 2013.
400   6.   Lysholm F, Wetterbom A, Lindau C, et al. Characterization of the viral
401   microbiome in patients with severe lower respiratory tract infections, using
402   metagenomic sequencing. PLoS One **2012**; 7:e30875.
403   7.   McMullan LK, Frace M, Sammons SA, et al. Using next generation sequencing to
404   identify yellow fever virus in Uganda. Virology **2012**; 422:1–5. Available at:
405   http://www.ncbi.nlm.nih.gov/pubmed/21962764. Accessed 3 September 2013.
406   8.   Briese T, Paweska JT, McMullan LK, et al. Genetic Detection and
407   Characterization of Lujo Virus, a New Hemorrhagic Fever–Associated Arenavirus
408   from Southern Africa. PLoS Pathog. **2009**; 5:8.
409   9.   Terzian ACB, Mondini A, Bronzoni RVDM, et al. Detection of Saint Louis
410   encephalitis virus in Dengue-suspected cases during a dengue 3 outbreak. Vector
411   Borne Zoonotic Dis. **2011**; 11:291–300. Available at:
412   http://www.ncbi.nlm.nih.gov/pubmed/20645866.
413   10.   da Silva Pessoa Vieira CJ, da Silva DJF, Barreto ES, et al. Detection of Mayaro
414   virus infections during a dengue outbreak in Mato Grosso, Brazil. Acta Trop.
415   **2015**; 147:12–16. Available at:
416   http://dx.doi.org/10.1016/j.actatropica.2015.03.020.
417   11.   Qiu J, Söderlund-Venermo M, Young NS. Human Parvoviruse. Clin Microbiol
418   Rev **2016**; 30:43–113.
419   12.   Waza K, Inoue K, Matsumura S. Symptoms Associated with Parvovirus B19
420   Infection in Adults: A Pilot Study. Intern. Med. **2007**; 46:1975–1978.
421   13.   Anderson MJ, Higgins PG, Davis LR, et al. Experimental parvoviral infection in
422   humans. J. Infect. Dis. **1985**; 152:257–265.
423   14.   Gubler DJ. Dengue and dengue hemorrhagic fever. Clin. Microbiol. Rev. **1998**;

424   11:480–496.

425   15.   Young NS, Brown KE. Parvovirus B19. N. Engl. J. Med. **2004**; 350:586–597.
426   Available at: http://www.nejm.org/doi/abs/10.1056/NEJMra030840.

427   16.   Lefrére JJ, Servant-Delmas A, Candotti D, et al. Persistent B19 infection in
428   immunocompetent individuals: Implications for transfusion safety. Blood **2005**;
429   106:2890–2895.

430   17.   Cohen BJ, Beard S, Knowles WA, et al. Chronic anemia due to parvovirus B19
431   infection in a bone marrow transplant patient after platelet transfusion. Transfusion
432   **1997**; 37:947–952.

433   18.   Erdman DD, Anderson BC, Török TJ, Finkel TH, Anderson LJ. Possible
434   transmission of parvovirus B19 from intravenous immune globulin. J. Med. Virol.
435   **1997**; 53:233–236.

436   19.   Von Kaisenberg CS, Jonat W. Fetal parvovirus B19 infection. Ultrasound Obstet.
437   Gynecol. **2001**; 18:280. Available at: http://www.blackwell-
438   synergy.com/rd.asp?code=UOG&vol=18&page=280&goto=abstract.

439   20.   Lui SL, Luk WK, Cheung CY, Chan TM, Lai KN, Peiris JS. Nosocomial outbreak
440   of parvovirus B19 infection in a renal transplant unit. Transplantation **2001**;
441   71:59–64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11211196.

442   21.   Heegaard ED, Brown KE. Human parvovirus B19. Clin. Microbiol. Rev. **2002**;
443   15:485–505. Available at: pm:12097253.

444   22.   Romano CM, Lauck M, Salvador FS, et al. Inter- and Intra-Host Viral Diversity in
445   a Large Seasonal DENV2 Outbreak. PLoS One **2013**; 8:e70318. Available at:
446   http://dx.plos.org/10.1371/journal.pone.0070318. Accessed 7 August 2013.

447   23.   Romano CM, de Matos AM, Araujo ESA, et al. Characterization of Dengue virus
448   type 2: new insights on the 2010 Brazilian epidemic. PLoS One **2010**; 5:e11811.

449   24.   Villabona-Arenas CJ, de Oliveira JL, de Sousa-Capra C, Balarini K, Pereira da
450   Fonseca CRT, Zanotto PM de A. Epidemiological dynamics of an urban Dengue 4
451   outbreak in São Paulo, Brazil. PeerJ **2016**; 4:e1892. Available at:
452   https://peerj.com/articles/1892.

453   25.   Kaufmann J, Buccola JM, Stead W, Rowley C, Wong M, Bates CK. Secondary
454   symptomatic parvovirus B19 infection in a healthy adult. J. Gen. Intern. Med.
455   **2007**; 22:877–878.

456   26.   Kurtzman GJ, Cohen B, Meyers P, Amunullah A, Young NS. Persistent B19
457   parvovirus infection as a cause of severe chronic anaemia in children with acute
458   lymphocytic leukaemia. Lancet **1988**; 2:1159–62. Available at:
459   http://www.ncbi.nlm.nih.gov/pubmed/2903376.

460   27.   Jacobson SK, Daly JS, Thorne GM, McIntosh K. Chronic Parvovirus B19
461   Infection Resulting in Chronic Fatigue Syndrome: Case History and Review. Clin.
462   Infect. Dis. **1997**; 24:1048–1051. Available at:
463   https://academic.oup.com/cid/article-lookup/doi/10.1086/513627.

464   28.   Musiani M, Manaresi E, Gallinella G, Cricca M, Zerbini M. Recurrent Erythema
465   in Patients with Long-Term Parvovirus B19 Infection. Clin. Infect. Dis. **2005**;
466   40:117–119.

467   29.   dos Passos Cunha M, dos Santos CA, de Lima Neto DF, et al. Outbreak of
468   chikungunya virus in a vulnerable population of Sergipe, Brazil—A molecular and
469   serological survey. J. Clin. Virol. **2017**; 97:44–49.

470   30.   Söderlund M, Von Essen R, Haapasaari J, Kiistala U, Kiviluoto O, Hedman K.
471   Persistence of parvovirus B19 DNA in synovial membranes of young patients with
472   and without chronic arthropathy. Lancet **1997**; 349:1063–1065.

473   31.   Cassinotti P, Burtonboy G, Fopp M, Siegl G. Evidence for persistence of human
474   parvovirus B19 DNA in bone marrow. J. Med. Virol. **1997**; 53:229–232.

475   32.   Foley B, O'Mahony J, Hill C, Morgan JG. Evidence for persistence of parvovirus
476   B19 DNA in livers of adults. J. Med. Virol. **2001**; 65:395–401.

477   33.   Enders M, Weidner A, Rosenthal T, et al. Improved diagnosis of gestational
478   parvovirus B19 infection at the time of nonimmune fetal hydrops. J. Infect. Dis.
479   **2008**; 197:58–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18171285.

480   34.   Gallinella G, Zuffi E, Gentilomi G, et al. Relevance of B19 markers in serum
481   samples for a diagnosis of parvovirus B19-correlated diseases. J. Med. Virol.
482   **2003**; 71:135–139.

483   35.   Hoebe CJPA, Claas ECJ, Steenbergen JE van, Kroes ACM. Confirmation of an
484   outbreak of parvovirus B19 in a primary school using IgM ELISA and PCR on
485   thumb prick blood samples. J. Clin. Virol. **2002**; 25:303–307.

486   36.   Erdman DD, Usher MJ, Tsou C, et al. Human parvovirus B19 specific IgG, IgA,
487   and IgM antibodies and DNA in serum specimens from persons with erythema
488   infectiosum. J. Med. Virol. **1991**; 35:110–115.

489   37.   Manaresi E, Gallinella G, Zuffi E, Bonvicini F, Zerbini M, Musiani M. Diagnosis
490   and quantitative evaluation of parvovirus B19 infections by real-time PCR in the
491   clinical laboratory. J. Med. Virol. **2002**; 67:275–281.

492   38.   Woolf AD, Campion G V, Chishick A, et al. Clinical manifestations of human
493   parvovirus B19 in adults. Arch. Intern. Med. **1989**; 149:1153–6. Available at:
494   http://www.ncbi.nlm.nih.gov/pubmed/2541666.

495   39.   Oliveira SA de, Camacho LAB, Pereira AC de M, Faillace TF, Setubal S,
496   Nascimento JP do. Clinical and epidemiological aspects of human parvovirus B19
497   infection in an urban area in Brazil (Niterói city area, State of Rio de Janeiro,
498   Brazil). Mem. Inst. Oswaldo Cruz **2002**; 97:965–70. Available at:
499   http://www.ncbi.nlm.nih.gov/pubmed/12471422.

500   40.   de Freitas RB, Durigon EL, Oliveira D de S, et al. The 'pressure pan' evolution of
501   human erythrovirus B19 in the Amazon, Brazil. Virology **2007**; 369:281–287.

502   41.   da Costa AC, Bendit I, de Oliveira ACS, Kallas EG, Sabino EC, Sanabani SS.
503   Investigation of human parvovirus B19 occurrence and genetic variability in
504   different leukaemia entities. Clin. Microbiol. Infect. **2013**; 19:E31–E43. Available
505   at: http://dx.doi.org/10.1111/1469-0691.12058.

506   42.   GIDEON Informatics I, Berger S. Parvovirus B19: Global Status: 2017 edition.
507   GIDEON Informatics, Incorporated, 2017. Available at:
508   https://books.google.com.br/books?id=KnIWDgAAQBAJ.

509   43.   de Freitas RB, Wong D, Boswell F, et al. Prevalence of human parvovirus (B19)
510   and rubellavirus infections in urban and remote rural areas in Northern Brazil. J.
511   Med. Virol. **1990**; 32:203–208.

512   44.   Cubel Garcia R de CN, Pereira RFA, Azevedo KML de, et al. Molecular diversity
513   of human parvovirus B19 during two outbreaks of erythema infectiosum in Brazil.
514   Brazilian J. Infect. Dis. **2017**; 21:102–106. Available at:
515   http://dx.doi.org/10.1016/j.bjid.2016.11.002.

516 45. Greenwald P, Bashe Jr WJ. An Epidemic of Erythema Infectiosum. Arch. Pediatr.
517   Adolesc. Med. **1963**; 107:68–72.
518 46. Oiwa H, Shimada T, Hashimoto M, et al. Clinical findings in parvovirus B19
519   infection in 30 adult patients in Kyoto. Mod. Rheumatol. **2011**; 21:24–31.
520   Available at: http://www.ncbi.nlm.nih.gov/pubmed/20680378.
521 47. Anderson MJ, Lewis E, Kidd IM, Hall SM, Cohen BJ. An outbreak of erythema
522   infectiosum associated with human parvovirus infection. J. Hyg. (Lond). **1984**;
523   93:85–93. Available at:
524   http://www.journals.cambridge.org/abstract_S0022172400060964.
525 48. Cohen BJ, Buckley MM, Clewley JP, Jones VE, Puttick AH, Jacoby RK. Human
526   parvovirus infection in early rheumatoid and inflammatory arthritis. Ann. Rheum.
527   Dis. **1986**; 45:832–838.
528 49. Anderson MJ, Jones SE, Minson AC. Diagnosis of human parvovirus infection by
529   dot-blot hybridization using cloned viral DNA. J. Med. Virol. **1985**; 15:163–172.
530 50. Anderson LJ, Tsou C, Parker RA, et al. Detection of antibodies and antigens of
531   human parvovirus B19 by enzyme-linked immunosorbent assay. J. Clin.
532   Microbiol. **1986**; 24:522–526.
533 51. Schwarz TF, Roggendorf M, Deinhardt F. Human parvovirus B19: ELISA and
534   immunoblot assays. J. Virol. Methods **1988**; 20:155–168.
535 52. Nobuo Y, Hiroyuki S, Kohtaro T, Akira Y, Kazuo S. Enzyme-linked
536   immunosorbent assay for IgG and IgM antibodies against human parvovirus B19:
537   use of monoclonal antibodies and viral antigen propagated in vitro. J. Virol.
538   Methods **1989**; 26:171–181.
539 53. Brown KE, Buckley MM, Cohen BJ, Samuel D. An amplified ELISA for the
540   detection of parvovirus B19 IgM using monoclonal antibody to FITC. J. Virol.
541   Methods **1989**; 26:189–198.
542 54. Clewley JP. Polymerase chain reaction assay of parvovirus B19 DNA in clinical
543   specimens. J. Clin. Microbiol. **1989**; 27:2647–2651.
544 55. Salimans MM, Van de Rijke FM, Raap AK, van Elsacker-Niele AM. Detection of
545   parvovirus B19 DNA in fetal tissues by in situ hybridisation and polymerase chain
546   reaction. J. Clin. Pathol. **1989**; 42:525–530.
547 56. Pattison JR, Jones SE, Hodgson J, et al. Parvovirus infections and hypoplastic
548   crisis in sickle-cell anaemia. Lancet **1981**; 317:664–665.
549 57. Rogers BB, Singer DB, Mak SK, Gary SW, Fikrig MF, McMILLAN PN.
550   Detection of human parvovirus B19 in early spontaneous abortuses using serology,
551   histology, electron microscopy, in situ hybridization, and the polymerase chain
552   reaction. Int. J. Gynecol. Obstet. **1993**; 43:232.
553 58. Söderlund-Venermo M, Hokynar K, Nieminen J, Rautakorpi H, Hedman K.
554   Persistence of human parvovirus B19 in human tissues. Pathol. Biol. (Paris). **2002**;
555   50:307–16. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12116849.
556 59. (CDC C for DC. Risks associated with human parvovirus B19 infection. MMWR.
557   Morb. Mortal. Wkly. Rep. **1989**; 38:81.
558 60. Weiner CP, Grose CF, Naides SJ. Diagnosis of fetal infection in the patient with
559   an ultrasonographically detected abnormality but a negative clinical history. Am. J.
560   Obstet. Gynecol. **1993**; 168:6–11. Available at: http://dx.doi.org/10.1016/S0002-
561   9378(12)90876-5.

562 61. Panchaud A, Stojanov M, Ammerdorffer A, Vouga M, Baud D. Emerging role of
563   Zika virus in adverse fetal and neonatal outcomes. Clin. Microbiol. Rev. **2016**;
564   29:659–694.
565 62. Calvet G, Aguiar RS, Melo ASO, et al. Detection and sequencing of Zika virus
566   from amniotic fluid of fetuses with microcephaly in Brazil: a case study. Lancet
567   Infect. Dis. **2016**; 16:653–660.
568 63. Sarno M, Sacramento GA, Khouri R, et al. Zika virus infection and stillbirths: a
569   case of hydrops fetalis, hydranencephaly and fetal demise. PLoS Negl. Trop. Dis.
570   **2016**; 10:e0004517.
571 64. Cugola FR, Fernandes IR, Russo FB, et al. The Brazilian Zika virus strain causes
572   birth defects in experimental models. Nature **2016**; 534:267–271. Available at:
573   http://dx.doi.org/10.1038/nature18296.
574 65. Sanabani S, Neto WK, Pereira J, Sabino EC. Sequence variability of human
575   erythroviruses present in bone marrow of Brazilian patients with various
576   parvovirus B19-related hematological symptoms. J. Clin. Microbiol. **2006**;
577   44:604–606.
578 66. Beer M, Conraths FJ, van der Poel WHM. 'Schmallenberg virus'--a novel
579   orthobunyavirus emerging in Europe. Epidemiol. Infect. **2013**; 141:1–8. Available
580   at: http://www.ncbi.nlm.nih.gov/pubmed/23046921. Accessed 7 April 2014.
581 67. Grard G, Fair JN, Lee D, et al. A novel rhabdovirus associated with acute
582   hemorrhagic fever in central Africa. PLoS Pathog. **2012**; 8:e1002924. Available at:
583   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3460624&tool=pmcen
584   trez&rendertype=abstract. Accessed 6 February 2014.
585 68. Yozwiak NL, Skewes-Cox P, Gordon A, et al. Human enterovirus 109: a novel
586   interspecies recombinant enterovirus isolated from a case of acute pediatric
587   respiratory illness in Nicaragua. J. Virol. **2010**; 84:9047–58. Available at:
588   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2937614&tool=pmcen
589   trez&rendertype=abstract. Accessed 22 August 2013.
590 69. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina
591   sequence data. Bioinformatics **2014**; 30:2114–2120.
592 70. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly
593   algorithm and its applications to single-cell sequencing. J. Comput. Biol. **2012**;
594   19:455–477.
595 71. Durigon EL, Erdman DD, Gary GW, Pallansch MA, Torok TJ, J.Anderson L.
596   Multiple primer pairs for polymerase chain reaction (PCR) amplification of human
597   parvovirus B19 DNA. J. Virol. Methods **1993**; 44:155–165.
598 72. Erdman DD, Durigon EL, Wang Q-Y, Anderson LJ. Genetic diversity of human
599   parvovirus B19: sequence analysis of the VP1/VP2 gene from multiple isolates. J.
600   Gen. Virol. **1996**; 77:2767–2774.
601 73. Kearse M, Moir R, Wilson A, et al. Geneious Basic: an integrated and extendable
602   desktop software platform for the organization and analysis of sequence data.
603   Bioinformatics **2012**; 28:1647–9. Available at:
604   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3371832&tool=pmcen
605   trez&rendertype=abstract. Accessed 13 August 2013.
606 74. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood
607   trees for large alignments. PLoS One **2010**; 5:e9490.

608 **Tables.**
609

610   Table 1. The demographics and diagnostics of Parvovirus B19 in Guarujá

Demographics of patients with B19V DNA^	Total
Female sex — no. (%)	66/111 (59)
Median age (range) — yr	27 (6−75)
5−17 yr (%)	19/106 (17.9)
18−39 yr (%)	61/106 (57.5)
≥40 yr (%)	26/106 (24.5)
Laboratory testing& — no./total (%)	
PCR-positive	115/182 (63.2)
PCR-positive and IgM-positive, IgG-negative	8/182 (4.4)
PCR-positive and IgG-positive, IgM-negative	47/182 (25.8)
PCR-positive alone	50/182 (27.5)
IgM-positive alone	4/182 (2.2)
IgM-positive and IgG-positive, PCR-negative	0/182 (0)
IgG-positive alone	25/182 (13.7)
PCR-negative, IgM-negative and IgG-negative	23/182 (12.6)
Clinical Diagnosis* — no./total (%)	
DENV NS1-specific ELISA test requested	46/70 (65.7)
Differential diagnosis of dengue	24/70 (34.2)

611   ^Four samples had no gender data, nine had no demographic data
612   &Borderline antibody reaction results were abstained from results
613   *Medical records for 70 patients were available. However, most records either were incomplete or illegible
614   for further analysis.
615

616

617   Table 2. Serology and clinical presentation in suspected Parvovirus B19 patients

Symptoms	No./total no. (%)	IgM+/IgG+ Total	IgM-/IgG+ M	F	Total	IgM-/IgG- M	F	Total
Fever	39/56 (69.6)	1 (50)	7 (70)	9 (75)	16 (73)	9 (69)	9 (82)	18 (75)
Arthralgia	6/56 (10.7)	0	2 (20)	0	2 (9)	1 (8)	2 (18)	3 (13)
Headache	24/56 (42.9)	0	4 (40)	5 (42)	9 (41)	5 (38)	7 (64)	12 (50)
Generalized Rash	6/56 (10.7)	1 (50)	2 (20)	0	2 (9)	0	1 (9)	1 (4)
Retro-orbital pain	8/56 (14.3)	0	1 (10)	3 (25)	4 (18)	3 (23)	0	3 (13)
Myalgia	34/56 (60.7)	1 (50)	8 (80)	7 (58)	15 (68)	6 (46)	8 (73)	14 (58)
Odynophagia	2/56 (3.6)	0	0	0	0	0	1 (9)	1 (4)
Complications								
Vomiting	6/56 (8.9)	0	1 (10)	0	1 (4)	2 (15)	2 (18)	4 (17)
Diarrhea	5/56 (9.3)	0	2 (20)	0	2 (9)	2 (15)	0	2 (8)
Thrombocytopenia	15/25 (60.0)	1 (50)	1/1 (100)	2/5 (40)	3/6 (50)	2/4 (50)	2/5 (40)	4/9 (44)
Total		2	10	12	22	13	11	24

618

619

620   Table 3. Age and sex of patients and their clinical presentations

Symptoms	No./total no. (%)	Age 5-17	>18	p-value^	Sex M	F	p-value^
Fever	39/56 (69.6)	11/13 (84)	28/43 (65)	0.3029	19/27 (68)	20/28 (71)	1
Arthralgia	6/56 (10.7)	0/13 (0)	6/43 (14)	0.3181	4/27 (14)	2/28 (7)	0.6687
Headache	24/56 (42.9)	5/13 (38)	19/43 (44)	0.7603	12/27 (43)	12/28 (43)	1
Generalized Rash	6/56 (10.7)	3/13 (23)	3/43 (7)	0.13	2/27 (7)	4/28 (14)	0.6695
Retro-orbital pain	8/56 (14.3)	1/13 (8)	7/43 (16)	0.6652	5/27 (18)	3/28 (11)	0.4688
Myalgia	34/56 (60.7)	4/13 (31)	30/43 (70)	0.02135	17/27 (71)	17/28 (61)	1
Odynophagia	2/56 (3.6)	0/13 (0)	2/43 (5)	1	1/27* (4)	1/28 (4)	1
Complications							
Vomiting	6/56 (8.9)	2/13 (15)	4/43 (9)	0.6286	4/27 (14)	2/28 (7)	0.4216
Diarrhea	5/56 (9.3)	0/13 (0)	5/43 (12)	0.58	5/27 (18)	0/28 (0)	0.02321
Thrombocytopenia	15/25 (60.0)	5/9 (55)	10/16 (63)	1	8/11 (73)	7/14 (50)	0.4139

621 *No serological data was available for this patient
622 ^Statistical inferences were performed using Fisher's exact test
623

624

625

Figure legends.

626   **Figure legends.**

627   Figure 1 – The 2013 to 2014 temporal and geographic spread of Parvovirus B19 cases in

628   the city of Guarujá. A) Weekly sampling of serum samples where Parvovirus B19 DNA

629   was detected (royal blue) and the proportion of samples that were first-time infections

630   (purple). B) The geographical location of the coastal city of Guarujá. C) The

631   geographical distribution of Parvovirus B19 cases in Guarujá.

632

633   Figure 2 – Phylogenetic estimation of the VP1/VP2 Parvovirus B19 gene fragment using

634   a maximum-likelihood tree. The Shimodaira-Hasegawa bootstrap values greater than

635   70% are shown at major nodes. Nodes were placed and colored at tree trips to indicate

636   the origin of sequences. Scale bar (s/s) is represented by substitutions per site. All

637   sequences from the Guarujá isolates can be found on Genbank under the accession

638   numbers MG765329-MG765371.

639

640   Supplementary Figure 1 - Phylogenetic estimation of the complete or nearly complete

641   Parvovirus B19 genomes using a radial maximum-likelihood tree. Red node tips indicate

642   assembled consensus sequences from each pool.
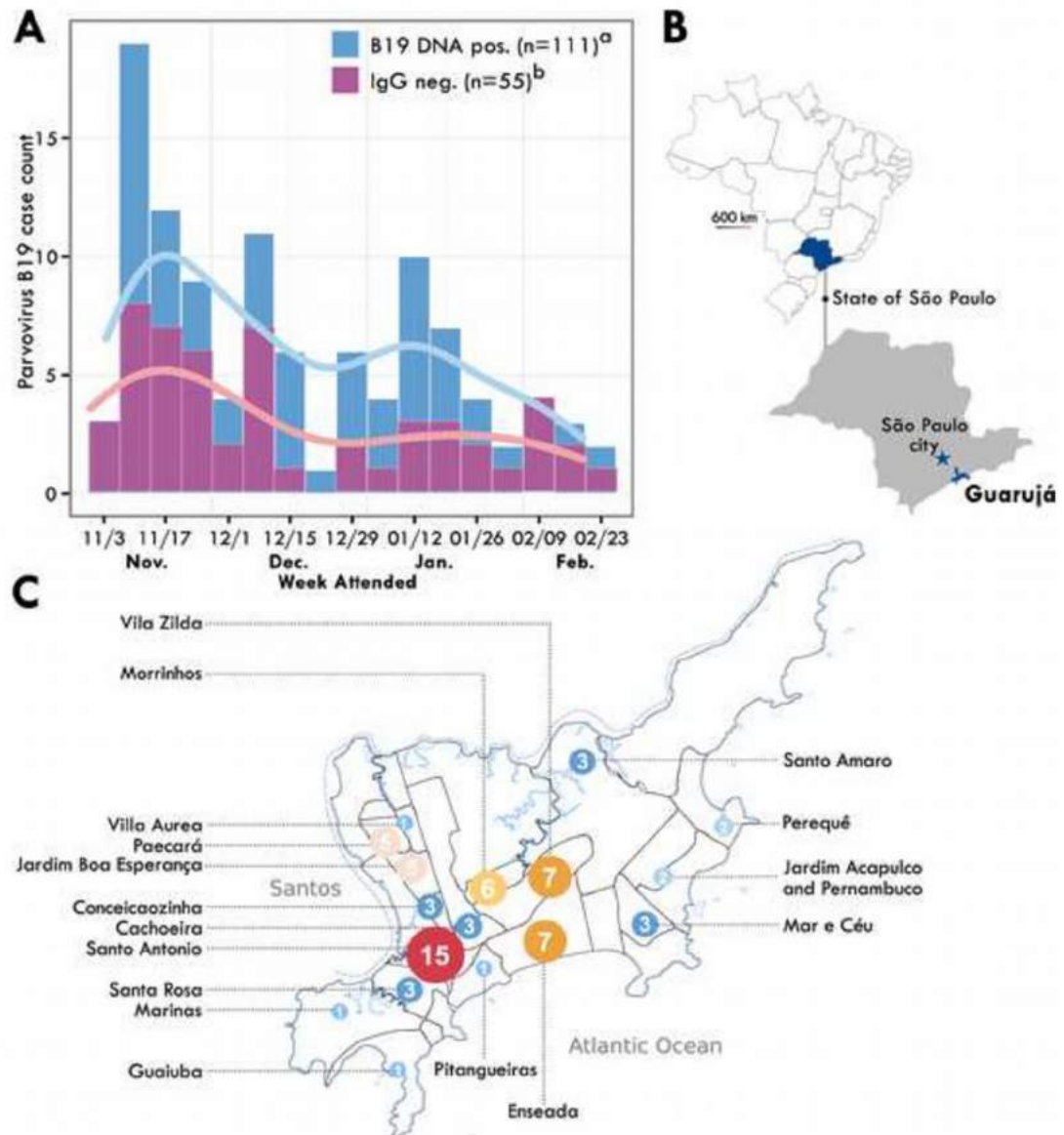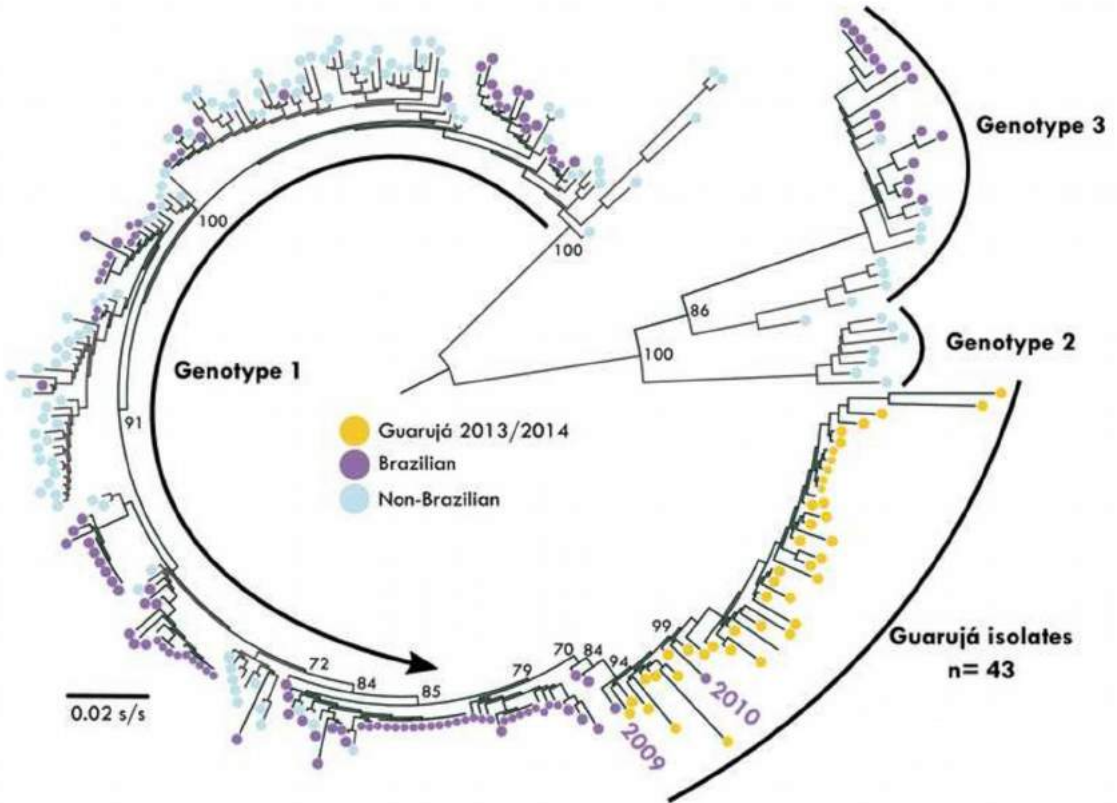
Figure 1

Genotype 1

● = complete genome assembled in pool

KC013330_BR_2008
KC013302_BR_2009
KC013338_BR_2009
KC013333_BR_2010
KC013314_BR_2009
KC013344_BR_2009
P5
P1
KC013331_BR_2009
P21
P24
P17
P12
P11
P22
P6
P20
KC013312_BR_2008
KC013351_BR_2010
KC013313_BR_2009
KC013303_BR_2007
KC013327_BR_2010

3A
2
3B

0.02