### **Ingra Morales Claro**

# Em tempo real, rápida detecção e sequenciamento de arbovírus no

### Brasil

Tese apresentada à Faculdade de Medicina da Universidade de São Paulo para obtenção do título de Doutora em Ciências

Programa de Doenças Infecciosas e Parasitárias

Orientadora: Profa. Dra. Ester Cerdeira Sabino

São Paulo 2021

### **Ingra Morales Claro**

# Em tempo real, rápida detecção e sequenciamento de arbovírus no

### Brasil

Tese apresentada à Faculdade de Medicina da Universidade de São Paulo para obtenção do título de Doutora em Ciências

Programa de Doenças Infecciosas e Parasitárias

Orientadora: Profa. Dra. Ester Cerdeira Sabino

São Paulo 2021

#### Dados Internacionais de Catalogação na Publicação (CIP)

Preparada pela Biblioteca da Faculdade de Medicina da Universidade de São Paulo

©reprodução autorizada pelo autor

Claro, Ingra Morales Em tempo real, rápida detecção e sequenciamento de arbovírus no Brasil / Ingra Morales Claro. -- São Paulo, 2021. Tese(doutorado)--Faculdade de Medicina da Universidade de São Paulo. Programa de Doenças Infecciosas e Parasitárias. Orientadora: Ester Cerdeira Sabino.
Descritores: 1.Arbovírus 2.Diagnóstico 3.Epidemiologia 4.Genômica 5.Metagenômica 6.Sequenciamento por nanoporos 7.RNA 8.Vírus 9.Vigilância em saúde pública 10.Reação em cadeia da polimerase multiplex
USP/FM/DBD-280/21

Responsável: Erinalva da Conceição Batista, CRB-8 6755

#### DEDICATÓRIA

A minha orientadora, a Professora Dra. Ester Cerdeira Sabino, pela confiança, inspiração, ensinamentos, conselhos, por todas as oportunidades que me levaram até aqui hoje, pelo carinho e amizade. Você foi a grande responsável por todo o meu crescimento profissional e pessoal nestes últimos anos.

Ao Professor Dr. Nicholas Loman e Dr. Joshua Quick, da University of Birmingham, pela oportunidade, por me receber com tanto carinho, cuidado e paciência, por todos os momentos especiais e felizes que me proporcionaram durante o projeto ZiBRA e durante a minha estadia em Birmingham, UK. E obrigada a todos do *Institute for Microbiology and Infection*, Emily Jane, Pablo Fuentes, Radoslaw, Jennie Law, Qing Yang, Alvaro Aguilar, Sam Nicholls pela recepção, parceria e amizade. Obrigada por todos os ensinamentos, contribuição, confiança, inspiração, que foram essenciais para o desenvolvimento e finalização desta tese.

Ao Dr. Nuno Faria, da Universidade de Oxford, pelos ensinamentos, conversas, amizade e por todas as oportunidades e confiança nos estudos. Obrigada também pela oportunidade de continuar minha carreira com você.

A todos do time ZiBRA, ZiBRA 2, em especial ao Prof. Dr. Luiz Alcantara, Prof. Dr. Marcio Nunes, Dra. Marta Giovanetti, Dr. Joilson Xavier, Dr. Felipe Iani, Dr. Vagner Fonseca, Dr. Allison Fabri, Prof. Dr. Tulio de Oliveira, Dra. Poliana Lemos, Dr. Luciano Franco, Dra. Patricia Vieira, Dra. Paola Paz, Dra. Fernanda de Bruicker-Nogueira. Obrigada pela experiência, conhecimentos compartilhados, por todos os momentos de trabalho duro, mas sempre com muita alegria, leveza e paixão pela ciência.

A todos do time CADDE, pelo trabalho incrível, incansável, sempre compartilhando experiências, amizades, momentos de comemoração, evolução, sempre torcendo um pelo outro e crescendo juntos. Toda a minha gratidão e amor por vocês - Dra. Ester Cerdeira Sabino, Dr. Nuno Faria, Dr. Nicholas Loman, Dr. Joshua Quick, Darlan Candido, Dra. Mariana Severo, Prof. Dr. Renato Santana, Dr. William Maciel, Dra. Sarah Hill, Dr. Julien Theze, Thais de Moura Coletti, Dr. Mariana Sequetim, Lucy Matkin, Erika Regina Manuli, Flavia Salles, Filipe Romero, Jaqueline Goes, Camila Alves, Pamela Andrade, Mariana Pinho, Bianca Costa, Esmenia Coelho, Midia Ferreira, Franciane Mendes, Lucas Franco,

Geovana Pereira, Leandro Menezes, Ian Nunes, Giulia Ferreira, Valquiria Reis, Raissa Heloisa Eliodoro, Mariane Amorim, Moritz Kraemer, Priscilla Paschoal, Beatriz Ribeiro, Joyce Vanessa, Mayara Bertanhe.

Aos colegas do Laboratório de Parasitologia LIM 46, Erika Regina Manuli, Dra. Léa Campos Oliveira, Dra. Cecilia Salete, Dr. Fabio Ghilardi, Dr. Carlos Henrique Valente, Camila Cruz, Marcia Hage, Natália Souza de Godoy, Dra. Regina Maia, Dra. Roberta Cristina, Gaspar Camilo, Gabriel Castanheira, Natalia Bueno e Felipe Silva. Pela amizade, colaboração, aprimoramento profissional, aprendizado e companhia diária.

Aos colegas do Laboratório de Bacteriologia LIM 54, a Professora Dra. Silvia Costa, Andréia Alcaia, Lauro Perdigão, Ana Paula Marchi, Débora Nagano e Bruna Carvalho, Victor Castro, Alessandra Nunes, Evelyn Sánchez.

Aos colegas do Laboratório de Virologia, LIM 52, Dra. Camila Romano, Dra. Lucy Santos Vilas Boas, Dr. José Eduardo Levi, Profa. Dra. Cassia Mendes, Profa. Dra. Tania Regina, Anderson de Paula, Noely Evangelista, Dra. Cynthia Motta do Canto, Dra. Maria Cristina Domingues Fink, Alvina Clara Felix, Ana Carolina Mamana, Giovana Caleiro, Caio Souza, Heuder Gustavo, Silvia Lima, Marli de Paula Estevam, Sônia Aparecida dos Anjos Ferraz, Wilton Santos Freire, Luciano Monteiro da Silva, Dra. Cristina Nunes, Natan Galvani, Nathalia Santiago, por tantos anos de amizade, cooperação, momentos felizes e risadas.

Ao Departamento de Moléstias Infecciosas e Parasitárias.

À Roseli Antonia Santo, à Vânia Regina Miguel, e à Luiza Maria Assis Vieira, pelo apoio administrativo.

Aos meus pais, Fábio Claro da Silva e Rosa Maria Morales da Silva, e familiares, Gabriel Amaral Ribeiro da Silva, Najela Morales Claro Trevizan, Hugo Albert Trevizan, por estarem do meu lado sempre, me apoiando, por todo o amor, paciência e incentivo. Aos meus avós Irineu Morales, Rosa Sanches Morales, Ismael Claro da Silva e Helena Leme da Silva.

A todos os meus parentes e amigos.

A Deus, pela minha existência e por iluminar o meu caminho.

#### AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Faculdade de Medicina da Universidade de São Paulo, Instituto de Medicina Tropical da Universidade de São Paulo, Universidade de Birmingham e Universidade de Oxford, sob financiamento das instituições: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) processo número: 201708012-9, e Fundação FAPESP-MRC/UKRI-NEWTON processo número: 2018/14389-0.

#### NORMALIZAÇÃO ADOTADA

Esta Tese está de acordo com as seguintes normas, em vigor no momento desta publicação:

Referências: adaptado de International Committee of Medical Journals Editors (Vancouver).

Universidade de São Paulo. Faculdade de Medicina. Divisão de Biblioteca e Documentação. *Guia de apresentação de dissertações, teses e monografias*. Elaborado por Anneliese Carneiro da Cunha, Maria Julia de A. L. Freddi, Maria F. Crestana, Marinalva de Souza Aragão, Suely Campos Cardoso, Valéria Vilhena. 3a ed. São Paulo: Divisão de Biblioteca e Documentação; 2011.

Abreviaturas dos títulos dos periódicos de acordo com *List of Journals Indexed in Index Medicus*.

# SUMÁRIO

Lista de tabelas, Lista de figuras, Lista de siglas, Lista de anexos

## Resumo

## Abstract

1 INTRODUÇÃO	1
2 OBJETIVOS	12
2.1 Objetivo geral	12
2.2 Objetivos específicos	12
3 CAPÍTULOS	13
3.1 Capítulo 1 - Desenvolvimento de um esquema de primers para detecção, ampl	ificação
e sequenciamento do genoma completo de múltiplos vírus em um único o	ensaio -
Superpool	13
3.1.1 Resumo	13
3.1.2 Material e métodos	18
3.1.2.1 Seleção de amostras	18
3.1.2.2 Esquema de <i>Primers</i> gerados durante a validação do protocolo <i>Superpool</i>	21
3.1.2.3 Extração de ácido nucléico viral	22
3.1.2.4 Síntese de cDNA e Multiplex PCR	22
3.1.2.5 Preparação e sequenciamento de biblioteca de nanopore para PCR multiplex	24
3.1.2.6 Análise bioinformática	24
3.1.3 Resultados	25
3.2 Capítulo 2 - Desenvolvimento de uma abordagem rápida utilizando a f	usão de
barcodes para amplificação do genoma completo de múltiplos vírus num único	o ensaio
denominado SMART-plex	32
3.2.1 Resumo	32
3.2.2 Material e métodos	35
3.2.2.1 Seleção de amostras	35
3.2.2.2 Primers gerados durante a validação do protocolo SMART-plex	37
3.2.2.3 Extração de ácido nucléico viral	

3.2.2.4 Síntese de cDNA SMART-plex e PCR rápido	39
3.2.2.5 Preparação e sequenciamento de biblioteca de nanopore ensaio SMART-plex	40
3.2.2.6 Análise bioinformática	41
3.2.3 Resultados	41
3.3 Capítulo 3 - Desenvolvimento de um protocolo de metagenômica viral SMART-	• <i>9N</i> e
um protocolo rápido de metagenômica <i>Rapid SMART-9N</i> para rastreio de vírus de l	RNA
e potencial descoberta de novos vírus	48
3.3.1 Resumo	48
3.3.2 Material e métodos	50
3.3.2.1 Seleção de amostras	50
3.3.2.2 Extração de ácido nucléico e teste RT-qPCR	53
3.3.2.3 Multiplex tiling PCR	54
3.3.2.4 Preparação e sequenciamento da biblioteca de sequenciamento nanopore para o e	nsaio
multiplex PCR	54
3.3.2.5 Síntese de cDNA SMART-9N	55
3.3.2.6 Preparação da biblioteca e sequenciamento nanopore para o ensaio SMART-9N	55
3.3.2.7 Síntese de cDNA Rapid SMART-9N	56
3.3.2.8 Preparação e sequenciamento da biblioteca de sequenciamento nanopore para o e	nsaio
Rapid SMART-9N	56
3.3.2.9 Análise bioinformática	57
3.3.3 Resultados	58
3.4 Capítulo 4 - Otimização do custo do sequenciamento MinION	78
4 DISCUSSÃO	82
5 CONCLUSÃO	89
REFERÊNCIAS	
ANEXOS	

#### LISTA DE TABELAS

**Tabela 1-** Exemplo do esquema de *primers* gerado pela ferramenta *Primal Scheme* paraamplificação do genoma completo do vírus zika (ZIKV)15

 Tabela 2 - Titulação dos vírus. Vírus, linhagem, gênero, FFU e células isoladas de cada vírus.

 19

**Tabela 3** - Descrição das amostras positivas para vírus da febre amarela (YFV) (n = 41) por reação em cadeia da polimerase quantitativo em tempo real com as amostras correspondentes, valores de Ct, cópias estimadas do genoma por mL e tamanho de referência do vírus ......20

**Tabela 5** - Resumo dos dados de sequenciamento nanopore do vírus da febre amarela (YFV) usando a abordagem de PCR multiplex (n = 21) com os valores Ct correspondentes ......27

 Tabela 12 - Descrição das amostras coletadas e protocolo realizado para cada amostra ......50

 **Tabela 16** - Resumo do sequenciamento e resultados das estatísticas de alinhamento para a cepa de amostra de referência Zika vírus BeH815744 (ZIKV) usando o método SMART-9N durante o desenvolvimento (n = 1 amostra) de acordo com a *input* de material (FFU/mL) ....62

**Tabela 17** - Resumo dos resultados de sequenciamento e estatísticas de alinhamento para a cepa referência Zika vírus BeH815744 (ZIKV) usando o método Rapid SMART-9N durante o desenvolvimento (n = 1 amostra) de acordo com o *input* de material (FFU / mL) ......63

**Tabela 23** - Realização do protocolo de digestão79

#### LISTA DE FIGURAS

**Figura 8** - Reproduzido por Quick et al. 2017. Fluxo de trabalho para *tilling amplicon* sequencing em plataformas MinION / Illumina, com números de etapas de procedimento associados indicados. Legenda: *Step 1 and 2*: Design/order primers = Passos 1 e 2: preparação e solicitação dos *primers* individuais; *Step 3A-i-ii*: RNA extraction = Passos 3*A-i-ii*: extração do RNA; *Step 3B*: DNA extraction = Passo 3B: extração do DNA; *Step 3A- iii-vi*: cDNA

synthesis = Passo 3A-iii-vi: síntese de cDNA; Steps 4-9: Multiplex PCR = Passos 4-9: PCR multiplex; Steps 10-11: Quantification and QC = Passos 10-11: Quantificação e controle de qualidade; Step 12: Library preparation and sequencing = Passo 12: preparo da biblioteca e sequenciamento; Steps 13-16: Analysis pipeline = Passos 13-16 Pipeline análises; Steps 17 Quality control = Passos 17 e 18: Controle and 18: de qualidade..... 

**Figura 14** - Gráficos de cobertura da amostra de referência - ZIKV para o ensaio SMART-plex utilizando *primers* contendo as sequências da ONT - gerado por Tablet ......42

**Figura 15** - Gráficos de cobertura da amostra de referência - ZIKV para o ensaio SMART-plex utilizando *primers* contendo as sequências da NEB - gerado por Tablet ......42

**Figura 16 -** Gráficos de cobertura da amostra de referência - ZIKV para o ensaio SMART-plex utilizando diferentes temperaturas para o teste da transcrição reversa, 42°C, 47°C e 52°C, utilizando *primers* contendo as sequências da ONT - gerado por Tablet .......44

**Figura 17 -** Gráficos de cobertura da amostra de referência - ZIKV para o ensaio SMART-plex utilizando temperatura de transcrição reversa de 52°C. *Primers* utilizados: 9

nucleotídeos gene específicos + *tag* com sequências da NEB - gerado por Tablet ......45

#### LISTA DE SIGLAS

- ZIKV Zika vírus
- DENV Dengue vírus
- EBOV Ebola vírus
- HIV Imunodeficiência humana
- YFV Yellow fever virus
- CHIKV Chikungunya vírus
- ZiBRA Zika in Brazil Real-Time Analysis
- ECSA East-Central South African
- US\$ Dollar americano
- RNA Ácido Ribonucleico
- km2 Quilômetros quadrados
- NCBI National Center for Biotechnology Information
- NHP Primatas não humanos
- ACRV Área com recomendação de vacinação
- OMS Organização Mundial de Saúde
- SE Semana epidemiológica
- COVID-19 Doença de coronavírus 2019
- OROV Vírus Oropouche
- MAYV Vírus Mayaro
- SLV Vírus Saint Louis
- WNV Vírus do nilo ocidental
- RT-PCR Reação em cadeia da polimerase em tempo real
- **ONT Oxford Nanopore Technologies**
- UK United Kingdom
- USB Universal Serial Bus
- DNA Ácido Desoxirribonucleico
- Gb Giga pares de bases
- Kb Kilo pares de bases
- PCR Reação em cadeia da polimerase
- ZiBRA 2 Zika in Brazil Real-Time Analysis 2

RT-qPCR - Reação em cadeia da polimerase quantitativo em tempo real

- cDNA DNA complementar
- LACEN Laboratório Central
- Dra Doutora
- Dr Doutor
- CADDE Centre for Arbovirus Discovery, Diagnosis, Genomics and Epidemiology
- ROCV Vírus Rocio
- ILHV Vírus Ilhéus
- TCV Vírus Tacaiuma
- STLV Vírus Saint Louis
- nts Nucleotídeos
- BED Browser Extensible Data
- QC Controle de qualidade
- SMART Switching mechanism at 5 end of RNA template
- MEM Minimum Essential Medium
- DMEM Dulbecco's Modified Eagle Medium
- FFU focus forming units
- SSP Strand Switching primer
- Ct Cycle threshold
- HC-FMUSP Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo
- bp Pares de bases
- TE Tampão de eluição
- EB *Elution buffer*
- μl microlitros
- ng nanogramas
- µM micromolar
- dNTP deoxyribonucleotide triphosphate
- Cat. No. Catálogo número
- dsDNA double stranded DNA viruses
- EUA Estados Unidos da América
- mM milimolar
- NFW Nuclease free-water

### LISTA DE ANEXOS

- Anexo I Ficha do aluno com disciplinas realizadas e concluídas.
- Anexo II Artigos publicados.
- Anexo III Parecer Consubstanciado do CEP
- Anexo IV Comprovante submissão do artigo

#### **RESUMO**

Claro IM. *Em tempo real, rápida detecção e sequenciamento de arbovírus no Brasil* [tese]. São Paulo: Faculdade de Medicina, Universidade de São Paulo; 2021.

Os vírus emergentes e reemergentes transmitidos por artrópodes (arbovírus) sempre foram uma preocupação global para a saúde humana. O Brasil é um grande país tropical que fornece condições ideais para a existência de muitos arbovírus que são mantidos em uma ampla variedade de ciclos zoonóticos. Nos últimos anos, o Brasil foi afetado por uma onda de graves epidemias, vírus sobrepostos, principalmente causados pelos vírus Zika (ZIKV), dengue (DENV), imunodeficiência humana (HIV), Ebola (EBOV), febre amarela (YFV), e Chikungunya (CHIKV). As epidemias resultantes causaram alta morbidade, mortalidade e custos econômicos. Além disso, há no Brasil a co-circulação de outras arboviroses negligenciadas ou que são pouco discutidas na literatura médica que vêm sendo responsáveis por causar pequenos surtos ou casos esporádicos em diferentes regiões do país. No entanto, a nossa compreensão destes surtos é dificultada pelo desafio do diagnóstico laboratorial (muitas vezes possui baixa especificidade e/ou requer o conhecimento a priori dos vírus a serem analisados) e diagnóstico clínico (geralmente febre, dor de cabeça, dores nas articulações, erupções cutâneas), que por si só, se sobrepõem àquelas causadas pela co-circulação dos vírus transmitidos por artrópodes, resultando em uma disponibilidade limitada de dados de vigilância epidemiológica. Portanto, há uma necessidade urgente de melhorar nossa capacidade de realizar a vigilância de arbovírus a partir de amostras clínicas através de melhores diagnósticos moleculares. Ferramentas de sequenciamento são abordagens que podem, por sua vez, serem usadas para detectar e monitorar a diversidade de arbovírus, identificar linhagens emergentes e controlar o potencial de introdução de novos vírus, sendo capaz de preparar ou até mesmo prevenir novos surtos. Em 2016, o projeto ZiBRA (Zika no Brasil Real Time Analysis) foi pioneiro de uma nova abordagem para a vigilância de arbovírus usando uma tecnologia de sequenciamento portátil baseada em nanopores para obtenção de genomas completos do vírus Zika em tempo real - chamada "epidemiologia genômica". Sendo uma abordagem ainda complexa e limitada pelo número de vírus que podem ser sequenciados simultaneamente (uma reação para cada vírus), o que aumenta muito o custo e não permite a identificação de outros vírus que não estão sendo testados na reação, esta opção ainda é de difícil adoção como protocolo e tecnologia de instrumento de vigilância

de rotina em laboratórios de pesquisa e saúde pública. Neste trabalho buscamos gerar novas metodologias que possam ser utilizadas para a vigilância completa de arbovírus, incluindo um painel de detecção e sequenciamento do genoma completo dos principais arbovírus de importância pública e duas abordagens de sequenciamento por metagenômica para rastreio e potencial descoberta de novos vírus. Além disso, buscamos realizar a redução do custo de sequenciamento de arbovírus a US\$ 10 por amostra e padronizar uma melhoria no fluxo de trabalho o que acarretará na adoção desta tecnologia por pesquisadores, laboratórios de diagnóstico e laboratórios de vigilância pública resultando em uma maior precisão na vigilância de arboviroses com potencial de introdução e disseminação no Brasil e em outros países e acelerar o ritmo das pesquisas sobre genética e epidemiologia de infecções por arbovírus.

**Descritores:** Arbovírus; Diagnóstico; Epidemiologia; Genômica; Metagenômica; Sequenciamento por nanoporos; RNA; Vírus; Vigilância em saúde pública; Reação em cadeia da polimerase multiplex.

#### ABSTRACT

Claro IM. *Real-time, rapid detection and sequencing of arboviruses in Brazil* [thesis]. São Paulo: "Faculdade de Medicina, Universidade de São Paulo"; 2021.

Emerging and re-emerging viruses transmitted by arthropods (arboviruses) have always been a global concern for human health. Brazil is a largely tropical country that provides ideal conditions for the existence of many arboviruses that are maintained in a wide variety of zoonotic cycles. In recent years, Brazil has been affected by a wave of serious epidemics, overlapping viruses, mainly caused by the Zika virus (ZIKV), dengue virus (DENV), human immunodeficiency virus (HIV), Ebola virus (EBOV), yellow fever virus (YFV), and Chikungunya virus (CHIKV). The resulting epidemics caused high morbidity, mortality, and economic costs. In addition, there is in Brazil the co-circulation of other arboviruses that are neglected or that are little discussed in the medical literature, which have been responsible for causing small outbreaks or sporadic cases in different regions of the country. However, our understanding of these outbreaks is hampered by the challenge of laboratory diagnosis (often has low specificity and/or requires a priori knowledge of the viruses to be analyzed) and clinical diagnosis (generally fever, headache, joint pain, rashes), which by themselves overlap with those caused by the co-circulation of arthropod-borne viruses, resulting in limited availability of epidemiological surveillance data. Therefore, there is an urgent need to improve our ability to conduct arbovirus surveillance from clinical specimens through better molecular diagnostics. Sequencing tools are approaches that can, in turn, be used to detect and monitor arbovirus diversity, identify emerging strains and control the potential for the introduction of new viruses, being able to prepare or even prevent new outbreaks. In 2016, the ZiBRA (Zika in Brazil Real-Time Analysis) project pioneered a new approach to arbovirus surveillance using portable nanopore-based sequencing technology to obtain complete genomes of the Zika virus in real time - called "genomic epidemiology". Being an approach that is still complex and limited by the number of viruses that can be sequenced simultaneously (one reaction for each virus), which greatly increases the cost and does not allow the identification of other viruses that are not being tested in the reaction, this option is still difficult to be adopted as protocol and technology as a routine surveillance tool in research and public health laboratories. In this work, we seek to generate new methodologies that can be used for complete arbovirus surveillance, including a full genome detection and

sequencing panel of the main arboviruses of public importance and two metagenomic sequencing approaches for screening and potential discovery of new viruses. In addition, we seek to reduce the cost of arbovirus sequencing to US\$ 10 per sample and standardize an improvement in the workflow which will lead to the adoption of this technology by researchers, diagnostic laboratories, and public surveillance laboratories resulting in greater accuracy in the surveillance of arboviruses with the potential for introduction and spread in Brazil and other countries, and accelerate the pace of research on the genetics and epidemiology of arbovirus infections.

**Descriptors:** Arboviruses; Diagnosis; Epidemiology; Genomics; Metagenomics; Nanopore sequencing; RNA; Viruses; Public health surveillance; Multiplex polymerase chain reaction.

# 1 INTRODUÇÃO

Os vírus emergentes e reemergentes transmitidos por artrópodes (arbovírus) sempre foram uma preocupação para a saúde humana, especialmente em regiões tropicais e subtropicais, estando entre os problemas de saúde pública mais importantes dentro das doenças infecciosas. [1]

O Brasil é um grande país tropical (8.510.345,538km<sup>2</sup>) [2] que possui um ecossistema que fornece condições ideais para a existência de muitos arbovírus, mantidos em uma grande variedade de ciclos zoonóticos. O desmatamento, migração populacional, ocupação desordenada das áreas urbanas e condições sanitárias precárias favorecem a amplificação, transmissão viral, além da ampla distribuição e suscetibilidade dos insetos vetores em todas as regiões do país [3]. De fato, mais de 200 diferentes espécies de arbovírus foram isoladas no Brasil, e cerca de 40 delas causam doenças em humanos [4].

Inicialmente, os arbovírus foram classificados de acordo com critérios sorológicos (classificação antigênica). Todavia, atualmente, uma base molecular é utilizada para classificação taxonômica e, de acordo com a proposta mais recente, os arbovírus foram classificados em quatro famílias principais: Togaviridae (gênero Alphavirus), Flaviviridae (gênero Flavivirus), Bunyaviridae (gênero Orthobunyavirus e Phlebovirus) e Reoviridae [5-9]. Dentre as famílias com maior relevância no contexto nacional, podemos citar as famílias Flaviviridae e Togaviridae, em que fazem parte os vírus da Dengue (DENV), Zika (ZIKV), Febre Amarela (YFV) e Chikungunya (CHIKV), que nas últimas décadas, vêm sendo as principais causas de infecções emergentes e reemergentes, resultantes em alta morbidade, mortalidade e custos econômicos [10].

A transmissão do vírus Zika (ZIKV) nas Américas foi confirmada pela primeira vez em abril de 2015 no nordeste do Brasil [11,12]. Até o ano de 2018, o Brasil teve o maior número de casos reportados no mundo (> 300,000) [13] e também o maior número de casos associados com microcefalia e outras deficiências em recém-nascidos (2,366 casos) [14,15]. As cepas isoladas durante o surto no Brasil foram identificadas como linhagem asiática [16]. No entanto, em estudo recente, as sequências disponíveis na plataforma *National Center for Biotechnology Information* (NCBI) *Nucleotide database* (https://www.ncbi.nlm.nih.gov/) entre os anos 2015-2019 foram caracterizadas em sublinhagens: Micronesia, Cambodia, Malaysia (Linhagem Asiática) e Senegal, Nigéria e Uganda (Linhagem Africana) [17]. Além disso foi relatado uma inversão no perfil dos subtipos a partir do ano de 2019, com o subtipo da Micronésia representando 89,2% das sequências disponíveis, enquanto que durante os anos 2016, 2017 e 2018, a proporção de sequências submetidas com o subtipo Camboja era maior que 90% dos casos [17]. O estudo também identificou o surgimento de 5,4% de todas as sequências de 2019 classificadas como linhagem africana (em janeiro e maio), isoladas em dois estados do Brasil: Rio Grande do Sul em primatas não humanos (NHP) da espécie *Alouatta guaribe* [18] e no Sudeste, do Rio de Janeiro em uma espécie de mosquito *Aedes albopictus* [19].

Essas descobertas alertam para a circulação do ZIKV de linhagem africana nas Américas em NHP e em mosquitos que não tinham sido relatados anteriormente no Brasil. Levanta também a possibilidade de que o ZIKV tenha sido introduzido nas Américas em mais de uma ocasião [19], apontando uma falha na vigilância epidemiológica e no mapeamento das linhagens e sub linhagens disseminadas no país nos últimos anos.

Em 2020 foram reportados 3.692 casos prováveis de ZIKV, taxa de incidência 1,8 casos por 100 mil habitantes. Já em 2021, foram notificados 1.431 casos prováveis, correspondendo a uma taxa de incidência de 0,7 casos por 100 mil habitantes no país (Figura 1), com 15 genomas completos disponíveis. Em relação a 2020, os dados representam uma diminuição de 38,9% no número de casos [20].



**Figura 1** - Curva epidêmica dos casos prováveis do vírus Zika, por semanas epidemiológicas de início de sintomas, Brasil, 2020 e 2021\*. Legenda: n= número; SE=Semana epidemiológica. Fonte: Sinan NET.

O vírus Chikungunya (CHIKV) foi introduzido pela primeira vez em setembro de 2014, onde o Ministério da Saúde do Brasil confirmou casos autóctones transmitidos por CHIKV no Estado Federal do Amapá, em Feira de Santana [21]. Estudos genômicos revelaram o genótipo Asiático como circulante no Norte do país. Entretanto, três genomas de amostras de Feira de Santana revelaram uma linhagem distinta, o genótipo East-Central South African (ECSA), que teria entrado nas Américas primeiro [22,23]. Esse vírus foi importado por um viajante de Angola, mas a sua posterior distribuição e a distinção dos genótipos responsáveis pelos casos CHIKV no Brasil não tem sido devidamente realizada. A rápida expansão do CHIKV, junto à severa morbidade e ao fardo econômico, faz do CHIKV um dos arbovírus de maior importância, e uma grande ameaça à saúde global. [24]

O vírus da febre amarela é uma ameaça re-emergente que resultou em centenas de mortes em 2017, sendo o maior observado no país por décadas, com impactos sem precedentes na história recente da doença no país [25, 26]. O recente surto levou à ampliação da Área Com Recomendação de Vacinação (ACRV) para todo o território nacional [27]. Após ampla vacinação, nenhum caso foi confirmado entre julho de 2020 e janeiro de 2021 entre 125 casos humanos suspeitos notificados. No entanto, 37 casos em PNH foram confirmados laboratorialmente no mesmo período, entre 574 notificados, concentrados nas regiões sul, centro-oeste e sudeste do Brasil [27] (Figura 2), alertando sobre possível circulação do vírus em artrópodes e PNH das regiões e potencial infecção em humanos não vacinados.



**Figura 2** - Distribuição das epizootias em primatas não humanos (PNH) e dos casos humanos confirmados para febre amarela no Brasil, por município de ocorrência ou do local provável de infecção, entre julho de 2020 e janeiro de 2021. Legenda: PNH = Primatas não humanos; Km = quilômetros. Fonte: Sinan NET.

A doença ocasionada pelo vírus da Dengue é endêmica em mais de 100 países. A Organização Mundial da Saúde (OMS) estimou que mais de 40% da população mundial corre risco de ser infectada pelo vírus da Dengue, sendo observado cerca de 50 a 100 milhões de casos relatados mundialmente todos os anos [28]. No Brasil, no ano de 2016, foram registrados 1,500.535 casos prováveis de dengue, com mais de 500 óbitos [29]. Até a semana epidemiológica (SE) 8 de 2021 foram notificados 72.093 casos prováveis (taxa de incidência de 34,0 casos por 100 mil habitantes) de DENV no Brasil. Em comparação com o ano de 2020, houve uma redução de 75% de casos registrados para o mesmo período analisado [30].



**Figura 3** - Curva epidêmica dos casos prováveis de dengue, por semanas epidemiológicas de início de sintomas, Brasil, 2020 e 2021\*. Legenda: SE = Semana epidemiológica. Fonte: Sinan NET

No entanto, atualmente o Brasil enfrenta uma pandemia da doença de coronavírus 2019 (COVID-19) desde fevereiro de 2020. Sendo assim, desde a confirmação dos primeiros casos, observou-se uma diminuição dos registros de casos prováveis de DENV, CHIKV, e ZIKV no Brasil [30] (Figura 4). Esta diminuição pode ser consequência do receio da população em procurar atendimento em uma unidade de saúde, bem como uma possível subnotificação ou atraso nas notificações das arboviroses, associadas a mobilização das equipes de vigilância e assistência para o enfrentamento da pandemia.



**Figura 4** -Distribuição da taxa de incidência de dengue, chikungunya e zika, por município, Brasil, SE 1 a 16/2021. Legenda: Km = quilômetros; SE = Semana epidemiológica. Fonte: Sinan NET.

Além disso, co-circulam no Brasil junto com ZIKV, CHIKV, YFV e DENV outras arboviroses negligenciadas ou que são menos discutidas na literatura médica, em que podemos incluir os vírus Oropouche (OROV), Mayaro (MAYV), Saint Louis (SLEV), West Nile (WNV) [31-34], entre outros, que vêm sendo responsáveis por causar pequenos surtos ou casos esporádicos e pontuais em diferentes regiões do país, dificultando a compreensão das interações entre vírus geneticamente semelhantes dentro de populações de mosquitos, PNH e humanos. Portanto, há uma necessidade premente e urgente de entender melhor a epidemiologia e a biologia dessas infecções por arbovirus.

Até o momento, nossa capacidade de gerenciar esses surtos é prejudicada pelo desafio de fazer um diagnóstico clínico definitivo, pois muitos desses vírus são frequentemente indistinguíveis daqueles causados por vírus co-circulantes e algumas doenças bacterianas [35, 36]. Os testes diagnósticos podem ser limitados pela baixa especificidade, no caso dos testes sorológicos, ou exigir o conhecimento *a priori* dos vírus alvos no caso de

RT-PCR (reação em cadeia da polimerase em tempo real). Na rotina diagnóstica, normalmente os casos sintomáticos são testados apenas para os vírus de maior importância para a saúde pública no momento. Para os casos negativos, no entanto, não há uma metodologia implementada que possa detectar e/ou rastrear o agente patógeno responsável pelos casos. Por essas razões, a doença febril aguda frequentemente permanece sem diagnóstico, levando à falha da vigilância epidemiológica, como demonstrado anteriormente pelo ZIKV no Brasil, onde a detecção foi retardada por pelo menos um ano a partir do início real do surto [16] (Figura 5).



**Figura 5** - Reproduzido por Faria et al. 2017. Dados filogenéticos do vírus Zika de amostras coletadas durante o projeto ZiBRA, demonstrando a introdução do Zika no Nordeste do Brasil no final de 2013 e a disseminação subsequente no Brasil e nas Américas. Legenda: Jan = Janeiro; Jul = Julho; RJ = Rio de Janeiro; SP = São Paulo; TO = Tocantins; Earliest detection = Detecção precoce; Polynesia = Polinésia; N-Brazil = Norte-Brazil; NE-Brazil = Nordeste-Brazil; Caribbean = Caribe; Central America = America central; South America = America do Sul; This study = este estudo; Microcephaly = microcefalia.

As abordagens de sequenciamento são uma solução para estes problemas, pois podem detectar e caracterizar dezenas ou centenas de diferentes *amplicons* em um único ensaio. Criticamente, resultados ambíguos são menos prováveis, porque a sequência é obtida fornecendo informações sobre um alvo viral específico, e também se o resultado é genuíno ou o resultado de contaminação laboratorial. O sequenciamento é uma fonte mais rica de informação epidemiológica, uma vez que as sequências podem inferir as relações epidemiológicas entre os vírus. Esta informação é crucial no fornecimento de parâmetros importantes sobre o surto, incluindo o tamanho e relação dos casos. Isso pode, por sua vez, ser usado para monitorar a diversidade de arboviroses, identificar linhagens emergentes, entender as rotas de transmissão e disseminação em todo o Brasil e controlar o potencial de introdução de novos vírus [37]. Essas informações podem ter implicações significativas do ponto de vista de rastreio, testes de diagnóstico, monitorização de pacientes e desenvolvimento de vacinas. No entanto, o sequenciamento ainda é utilizado com pouca frequência para a vigilância de rotina dos arbovírus, devido à complexidade associados a estas abordagens quando comparado aos testes moleculares já padronizados e implementados.

Nos últimos anos, novas ferramentas foram impulsionados por avanços tecnológicos em sequenciamento. A recente disponibilidade da tecnologia da Oxford Nanopore Technologies (ONT, UK) - MinION vem democratizando o uso do sequenciamento. O MinION é um sequenciador portátil, que tem um custo de US \$1,000 e pode ser alimentado e controlado a partir da porta USB de um laptop padrão. Isso, por sua vez, significa que os laboratórios podem instalar a capacidade de sequenciamento de genomas a um custo total de US \$5,000, significativamente menor do que as tecnologias de sequenciamento anteriores (geralmente custando US \$200,000). Utilizando uma tecnologia a base de nanopores, as bases de DNA são identificadas a partir de mudanças na condutividade elétrica gerada à medida que os filamentos de DNA passam por um poro biológico, podendo gerar até 30 gigabases (Gb) de dados de sequências ultra longas (> 100 kilobases (Kb)) [38], uma das principais vantagems em relação às tecnologias de segunda geração. Leituras longas facilitam a montagem de genomas maiores, mais complexos e a resolução de sequências repetitivas no genoma.

Nosso grupo aplicou em 2016 uma técnica de **epidemiologia genômica** em tempo real ao vírus Zika durante o projeto ZiBRA - *Zika in Brazil Real Time Analysis* <u>http://www.zibraproject.org</u> em 2016 [16], um projeto itinerante de colaboração multicêntrica entre a Universidade de Oxford, a Universidade de Birmingham, Instituto Evandro Chagas, Universidade de São Paulo e Fundação Oswaldo Cruz. Após grandes resultados demonstrados durante a epidemia do vírus Ebola na África Ocidental, utilizando a tecnologia de sequenciamento por nanoporos da ONT e exploramos sua portabilidade para obter um tempo de resposta mais rápido, sequenciando as amostras perto de onde foram coletadas [39]. Um

protocolo intitulado "*Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples*" [40] foi conduzido usando sequenciamento de *amplicons* direcionados no MinION (ONT, UK) para obtenção do genoma completo do ZIKV diretamente de material clínico, sem necessário isolamento e cultura. Abrangendo uma ampla região geográfica, amostras históricas e de pacientes com uma variedade de apresentações clínicas, o projeto testou 1349 amostras clínicas de RNA para o ZIKV em um laboratório móvel nos estados do Rio Grande do Norte, Paraíba, Recife, Maceió e Bahia, capturou e testou 850 mosquitos dos campos urbanos e periurbanos e gerou 254 seqüências do genoma do ZIKV [16].



**Figura 6** - Distribuição do risco de transmissão local do vírus Zika nos estados das regiões norte e nordeste do Brasil, mapa à esquerda. Distribuição da probabilidade de ocorrência de focos do mosquito *Aedes aegypti* nos estados das regiões norte e nordeste do Brasil, mapa à direita. Fonte: <u>http://www.zibraproject.org</u>. Legenda: Risk of local Zika transmission = Risco de transmissão local do vírus zika; None = nenhum; Seasonal = Sazonal; Year round = Todo ano; PA = Para; MA = Maranhão; CE = Ceará; RN = Rio Grande do Norte; PB = Paraíba; PE = Pernambuco; AL = Alagoas; SE = Sergipe; SA = Salvador; Probability of *Ae. aegypti* occurrence = Probabilidade de ocorrência de *Aedes aegypti*.

O projeto ZiBRA teve continuação com o projeto ZiBRA 2, e hoje engloba outras regiões do Brasil (Norte, Sudeste e Centro-oeste) e da América do Sul (Paraguai e Costa Rica) e realizou não apenas a vigilância genômica do ZIKV, como também CHIKV, YFV, DENV 1-4, OROV e MAYV (https://github.com/zibraproject) [41-43].

Diversos estudos de vigilância genômica utilizando o MinION foram realizados desde então [44 - 47], no entanto, mesmo sendo a tecnologia de sequenciamento mais barata e acessível para a vigilância genômica de arbovírus nos dias de hoje, o custo por amostra do

sequenciamento ainda é alto, em torno de US \$ 50 a US \$ 100, devido aos reagentes empregados e número de amostras sequenciadas. Também, há uma complexidade significativa associada à operação do instrumento, criação de bibliotecas de sequenciamento e análise de dados e o tempo para realização do protocolo (8 horas e 30 minutos) ainda é muito maior quando comparado a outros métodos moleculares (em torno de 3 horas no total - extração do RNA e RT-qPCR), o que ainda torna improvável a implementação e metodologia de escolha nos centros de diagnóstico e vigilância epidemiológica do Brasil. Ainda, esta abordagem é melhor quando a cepa do surto é conhecida, e é menos adequada para diversos grupos virais ou descoberta de vírus.

A metagenômica viral, o processo de sequenciamento do conteúdo total de ácido nucleico viral em uma amostra (normalmente cDNA ou DNA), permite a caracterização genômica de vírus conhecidos e novos de maneira não direcionada. Essa técnica é particularmente útil para diagnósticos, laboratórios clínicos e vigilância em saúde pública [48-51]. No entanto, o sequenciamento metagenômico viral diretamente de amostras clínicas pode resultar em baixa sensibilidade, especialmente em amostras com uma baixa abundância de material genômico viral em relação ao ácido nucleico derivado do hospedeiro [52-54].

Um novo projeto intitulado CADDE - Brazil-UK *Centre for Arbovirus Discovery, Diagnosis, Genomics and Epidemiology* (http://caddecentre.org) teve início em Julho de 2019 coordenado pela Professora Dra. Ester Sabino e Professor Dr. Nuno Faria, reunindo centros no Brasil e no Reino Unido - Universidade de São Paulo e Universidade de Oxford, acompanhados pela Secretaria de Saúde de São Paulo, da Faculdade de Medicina de São José do Rio Preto, da Universidade de Birmingham e Universidade London School, com o objetivo de desenvolver novas tecnologias genômicas portáteis a fim de realizar a caracterização de vírus conhecidos e descoberta de novos de uma maneira **simples, rápida e de baixo custo**, além de aplicar novas ferramentas de análises epidemiológicas, matemáticas e espaciais para que pesquisadores locais e centros de diagnóstico tenham melhor acesso e possam utilizar essas metodologias para rastrear a dinâmica da infecção em alta resolução à medida que as epidemias se desdobram com o objetivo de melhorar a prontidão do sistema de saúde brasileiro para futuros surtos arbovirais.

Esta tese faz parte do projeto CADDE, portanto, neste trabalho, buscamos solucionar os problemas encontrados no ZiBRA e ZiBRA2, seguindo os objetivos do CADDE de desenvolvimento de novas tecnologias genômicas para ajudar a implementação de uma nova ferramenta de epidemiologia genômica nos centros de saúde do Brasil.

Neste trabalho, foi realizada a padronização de um painel para detecção e sequenciamento do genoma completo dos arbovírus de maior importância pública, denominado Superpool acrescentando inovações técnicas desenvolvendo um esquema de primers que poderá amplificar múltiplos vírus num único ensaio, incluindo os vírus, Zika, Chikungunya, Febre Amarela, Mayaro, Oropouche, vírus do Nilo Ocidental (WNV), Dengue 1-4 (DENV1, DENV2, DENV3, DENV4), Rocio (ROCV), Ilhéus (ILHV), Tacaiuma (TCV) e Saint Louis (STLV) - Capítulo 1; Desenvolvimento de uma abordagem rápida utilizando a fusão de barcodes para amplificação do genoma completo de múltiplos vírus num único ensaio denominado SMART-plex - Capítulo 2; Desenvolvimento de protocolos de metagenômica viral SMART-9N e Rapid SMART-9N para rastreio e potencial descoberta de novos vírus - Capítulo 3; Otimização do custo do sequenciamento MinION - Capítulo 4. Essas inovações são altamente manejáveis e realizáveis e propõe reduzir o custo do sequenciamento de arbovírus por amostra para os painéis e simplificar o fluxo de trabalho de forma que possa ser mais prontamente adotado por pesquisadores e laboratórios de diagnóstico que desejam explorar estas novas tecnologias. Assim pretendemos disponibilizar uma metodologia completa de vigilância dos arbovírus a fim de trazer beneficios diretos à continuação e evolução dos projetos de vigilância genômica e outros projetos que abordam pesquisas de vigilância em tempo real de arbovírus, facilitando a capacidade de identificar com precisão vírus e variantes ligadas para quasispecies individuais, melhorando a detecção de patógenos para aplicações de diagnóstico e vigilância e trazendo impacto imediato no ritmo de pesquisa sobre a genética e epidemiologia de infecções por arbovírus e nossa compreensão desse problema urgente de saúde pública.

### **2 OBJETIVOS**

### 2.1 **Objetivo geral:**

O objetivo geral deste projeto é desenvolver uma abordagem completa de vigilância genômica de arbovírus no Brasil a partir de metodologias de detecção por sequenciamento MinION, com o objetivo principal de reduzir o custo, o tempo e a complexidade dos ensaios. Isso fará com que o custo do ensaio esteja de acordo com os testes de diagnóstico moleculares existentes e seja utilizável em ambientes de diagnóstico, de pesquisa e de saúde pública.

### 2.2 **Objetivos específicos:**

- Obter uma amplificação bem-sucedida do genoma completo de amostras testes de Zika vírus, Chikungunya vírus, vírus da Febre Amarela, Mayaro vírus, Oropouche vírus, vírus do Nilo Ocidental, Dengue 1-4 vírus, Rocio virus, Ilhéus vírus, Tacaiuma vírus e Saint Louis vírus utilizando um painel de arbovírus e ensaio Multiplex PCR - *Superpool*;

- Obter uma amplificação bem-sucedida do genoma completo de amostras testes de Zika vírus, Chikungunya vírus, vírus da Febre Amarela, Mayaro vírus, Oropouche vírus, vírus do Nilo Ocidental, Dengue 1-4 vírus, Rocio virus, Ilhéus vírus, Tacaiuma vírus e Saint Louis vírus utilizando um painel rápido de arbovírus e ensaio *SMART-plex*;

- Obter uma identificação e amplificação bem-sucedida do genoma completo de arbovírus presentes em amostras clínicas utilizando uma metodologia metagenômica - ensaio *SMART-9N*;

- Obter uma identificação e amplificação bem-sucedida do genoma completo de arbovírus presentes em amostras clínicas utilizando uma metodologia rápida de metagenômica - ensaio *Rapid SMART-9N*;

- Padronizar um protocolo de sequenciamento bem sucedido a base de tecnologia nanopore rápido, simples e de baixo custo.

# **3 CAPÍTULOS**

Descrevemos aqui três protocolos de ponta a ponta totalmente integrados para sequenciamento rápido e vigilância completa de genomas virais diretamente de amostras clínicas. Todos os planos de trabalho aqui apresentados foram realizados na Universidade de Birmingham e no Instituto de Medicina Tropical da Universidade de São Paulo, sob auxílio da Professora Dra. Ester Cerdeira Sabino, Professor Dr. Nicholas Loman, e Dr. Joshua Quick.

**3.1 Capítulo 1** - Desenvolvimento de um esquema de *primers* para detecção, amplificação e sequenciamento do genoma completo de múltiplos vírus em um único ensaio - *Superpool*.

#### 3.1.1 Resumo

O PCR Multiplex é o método mais econômico no momento para amplificar e sequenciar vírus presentes em amostras clínicas em baixa abundância e tem sido utilizado com sucesso em trabalhos anteriores como com os vírus Ebola e Zika [39, 40]. Este método pode fornecer tanto enriquecimento como amplificação em um único passo que consiste em uma amplificação do genoma completo do vírus alvo utilizando um conjunto de *primers* multiplexados desenvolvidos pela ferramenta online *Primal Scheme* (http://primal.zibraproject.org) (Figura 7).



**Figura 7** - Reproduzido por Quick et al. 2017. *Overview* do esquema de *primers* do vírus Zika e Chikungunya gerado pela ferramenta *Primal Scheme*. (a) Caixa de envio para ferramenta de design de *primer* online. Scheme name = nome do esquema; amplicon length = tamanho do amplicon; Overlap = sobreposição. (b) Tabela de resultados de primer. (c) Esquema mostrando produtos de amplicon esperados para cada pool no contexto genômico para os esquemas ZikaAsian e ChikAsianECSA.

Em resumo, para a geração dos esquemas de *primers*, o website *Primal Scheme* pede o upload de um arquivo no formato FASTA (Figura 6 - a) que deve ser do genoma mais representativo do vírus em interesse, o comprimento do *amplicon* em nucleotídeos (nts) e o tamanho do *overlap* em nts, que é o comprimento da sobreposição entre os amplicons vizinhos. *Amplicons* mais longos são preferíveis, pois significam um número menor de pares de *primers* necessários por reação. Eles também aumentam a quantidade de informações de
ligação que podem ser recuperadas como haplótipos, o que é importante para a investigação da diversidade dentro do hospedeiro. No entanto, *amplicons* mais longos são menos propensos a uma amplificação bem sucedida quando o número de cópias virais é baixo ou há degradação da amostra (por exemplo, devido ao armazenamento, transporte inadequado ou prolongado). Para os vírus pesquisados neste trabalho os comprimentos dos amplicons escolhidos foram entre 400-500 nts; *Overlap*: padrão = 75 a 100 nts.

Os arquivos de saída incluem uma tabela contendo as sequências de *primers* a serem solicitados, um arquivo BED com os locais dos *primers* que podem ser usados posteriormente para a padronização e um diagrama do esquema de *primers* para cada vírus (*pools* 1 e 2). (Tabela 1) (Figura 7 - b,c)

Primer	Sequência (5'-3')	Pool
400_1_out_L	GACAGTTCGAGTTTGAAGCGAAAG	1
400_1_out_R	AGTATGCACTCCCACGTCTAGT	1
400_2_out_L	AAGAAAGATCTGGCTGCCATGC	2
400_2_out_R	TGATTCCAACCAGGTTTGCGAC	2
400_3_out_L*	AGATGACGTCGATTGTTGGTGC	1
400_3_out_R*	TACGGTGACACAACCTCCATGT	1
400_4_out_L	TCAGGTGCATAGGAGTCAGCAA	2
400_4_out_R	GGAGCCATGAACTGACAGCATT	2
400_5_out_L	AGAACGTTAGTGGACAGAGGCT	1
400_5_out_R	TGTGCGTCCTTGAACTCTACCA	1
400_6_out_L	TTGATTGTGAACCGAGGACAGG	2
400_6_out_R	CCATCTGTCCCTGCGTACTGTA	2
400_7_out_L	TGAAGGGCGTGTCATACTCCTT	1
400_7_out_R	CGCCTCCAACTGATCCAAAGTC	1
400_8_out_L*	GGGAGAAGAAGATCACCCACCA	2
400_8_out_R*	TTGACTGCTGCTGCCAATCTAC	2
400_9_out_L*	GCCTTAGGGGGAGTGTTGATCT	1
400_9_out_R*	GAGTGGGCATTCCTTCAGTGTG	1
400_10_out_L	ACGGTCGTTGTGGGATCTGTAA	2
400_10_out_R	GTGGGACTTTGGCCATTCACAT	2
400_11_out_L	CAGCCGTTATTGGAACAGCTGT	1

 Tabela 1- Exemplo do esquema de primers gerado pela ferramenta Primal Scheme para amplificação do genoma completo do vírus zika (ZIKV).

400_11_out_R	CCTGGGCCTTATCTCCATTCCA	1
400_12_out_L	CACTAAGGTCCACGTGGAGGAA	2
400_12_out_R	TATCAGCGCCAGATGAGCTACA	2
400_13_out_L	TGGCAGTGCTGGTAGCTATGAT	1
400_13_out_R	AGAGAGAGGAGCATAAACCCCC	1
400_14_out_L*	CAATGGTTTTGCTTTGGCCTGG	2
400_14_out_R*	TTTCCCATGTGATGTCACCTGC	2
400_15_out_L	CCCTAGCGAAGTACTCACAGCT	1
400_15_out_R	TACACTCCATCTGTGGTCTCCC	1
400_16_out_L*	GTGGCATGAACCCAATAGCCAT	2
400_16_out_R*	GCTCCAATGTCCCCATCCTTTG	2
400_17_out_L**	GTGGTCCATGGAAGCTAGATGC	1
400_17_out_R**	CCTCTAAGGGCCTCCTCCATTT	1
400_18_out_L**	CTGTTGAGTGCTTCGAGCCTTC	2
400_18_out_R**	TGGTGAGTTGGAGTCCGGAAAT	2
400_19_out_L	TATGGATGAGGCCCACTTCACA	1
400_19_out_R	GCCATCAAGTATGACCGGCTTT	1
400_20_out_L*	GGCTGGAAAACGGGTCATACAG	2
400_20_out_R*	CCTTTGCTCCGTCCTAAGCTTG	2
400_21_out_L*	AGAGACTGACGAAGACCATGCA	1
400_21_out_R*	CTCCAAAAGCCGCTCCTCTTT	1
400_22_out_L**	TGGACCAGACACGGAGAGAAAA	2
400_22_out_R**	ATTCTGGCTGGCTCAATTTCCG	2
400_23_out_L	CGTCTTGATGAGGAACAAGGGC	1
400_23_out_R	AAGTGGTCACTGCATGTTGGAC	1
400_24_out_L**	TAATGGGAAGGAGAGAGGAGGG	2
400_24_out_R**	TCTCCACTTGGGGGTCAATTGT	2
400_25_out_L**	CCCTGACCCTAATAGTGGCCAT	1
400_25_out_R**	CCTTCCATTTCTCTCCCAGGGT	1
400_26_out_L	ACTGGAACTCCTCTACAGCCAC	2
400_26_out_R	ACCAGGGCCTCCTTTTGTGTAT	2
400_27_out_L	AGTGCAAAGCTGAGATGGTTGG	1
400_27_out_R	ATGTGTAGAGTTGCGGGAGAGT	1
400_28_out_L	GGTGGGGGATTGGCTTGAAAAA	2
400_28_out_R	GGGCCTCATAGCTTCCATGGTA	2
400_29_out_L	AGGATGTGAATCTCGGCTCTGG	1
400_29_out_R	ATGCTGCATTGCTACGAACCTT	1

400_30_out_L	AAAAGTGGACACTAGGGTGCCA	2
400_30_out_R	TAATCCCAGCCCTTCAACACCA	2
400_31_out_L	ACAAGGGGAATTTGGAAAGGCC	1
400_31_out_R	CGTAAGTGACAACTTGTCCGCT	1
400_32_out_L	AAATGGAAAAAGGGCACAGGGC	2
400_32_out_R	TGTCCCATCCAGTTGAGGGTTT	2
400_33_out_L*	CAAACGAATGGCAGTCAGTGGA	1
400_33_out_R*	ATCCACACTCTGTTCCACACCA	1
400_34_out_L*	ATTTCCACAGAAGGGACCTCCG	2
400_34_out_R*	TGACTAGCAGGCCTGACAACAT	2
400_35_out_L	ACCACCTGGGCTGAGAACATTA	1
400_35_out_R	ACCACTAGTCCCTCTTCTGGAG	1

Legenda: R = Right; L = Left.

Contudo, este método ainda é limitado ao número de vírus que podem ser detectados e sequenciados simultaneamente.

A concepção de ensaios de PCR multiplex com um grande número de pares de *primers* é difícil devido à maior probabilidade de interações entre estes, o que reduz a eficiência da amplificação e conduz a falhas na cobertura do genoma alvo. Neste trabalho, combinamos todos os esquemas de *primers* já desenvolvidos e validados individualmente em uma só reação, em que denominamos de *Superpool*, incluindo os vírus:

a) Zika;

b) Dengue (os quatro sorotipos);

- c) Chikungunya;
- d) Febre amarela;
- e) Mayaro;
- f) Oropouche;
- g) West Nile;
- h) Saint Louis;
- i) Rocio;
- j) Ilhéus;
- k)Tacaiuma.

O protocolo para validação do painel prosseguiu em seis estágios: (*Step 1 and 2*) preparação e solicitação dos *primers* individuais, (*Step 3A-i-ii*) extração do RNA viral, (*Step 3A- iii-vi*) síntese de cDNA, (*Steps 4-9*) PCR multiplex, (*Steps 10-11*) Quantificação e controle de qualidade, (*Step 12*) preparo da biblioteca de sequenciamento e sequenciamento MinION e (*Steps 13-18*) análise bioinformática e controle de qualidade (QC) (Figura 8).



**Figura 8** - Reproduzido por Quick et al. 2017. Fluxo de trabalho para *tilling amplicon sequencing* em plataformas MinION / Illumina, com números de etapas de procedimento associados indicados. Legenda: *Step 1 and 2*: Design/order primers = Passos 1 e 2: preparação e solicitação dos *primers* individuais; *Step 3A-i-ii*: RNA extraction = Passos 3*A-i-ii*: extração do RNA; *Step 3B*: DNA extraction = Passo 3B: extração do DNA; *Step 3A-ii-ii*: vi: cDNA synthesis = Passo 3*A- iii-vi*: síntese de cDNA; *Steps 4-9*: Multiplex PCR = Passos *4-9*: PCR multiplex; *Steps 10-11*: Quantification and QC = Passos *10-11*: Quantificação e controle de qualidade; *Step 12*: Library preparation and sequencing = Passo *12*: preparo da biblioteca e sequenciamento; *Steps 13-16*: Analysis pipeline = Passos 13-16 Pipeline análises; *Steps 17 and 18*: Quality control = Passos 17 e 18: Controle de qualidade.

#### 3.1.2 Material e métodos

### 3.1.2.1 Seleção de amostras

Utilizamos para desenvolvimento inicial dos métodos amostras de isolados do vírus Zika - Referência: OMS - (11474/16), disponível em *School of Biosciences*, *University of*  *Birmingham*, Reino Unido; isolados virais previamente identificados por RT-qPCR e sequenciamento viral dos vírus Zika (ZIKV), Chikungunya (CHIKV), Febre Amarela (YFV), Mayaro (MAYV), Oropouche (OROV), *West Nile* (WNV), Dengue 1-4 (DENV1, DENV2, DENV3, DENV4), Rocio (ROCV), Ilhéus (ILHV), Tacaiuma (TCV) e Saint Louis (STLV) fornecidas pelo Centro de Pesquisa em Virologia da Faculdade de Medicina de Ribeirão Preto, Ribeirão Preto (SP), Brasil. As cepas foram propagadas em células Vero (CCL-81; ATTC, Manssas, EUA) com *Minimum Essential Medium* (MEM) ou *Dulbecco's Modified Eagle Medium* (DMEM), ou em células C6/36 com meio Leibovitz (L-15) (Cat No. L1518, Sigma-Aldrich, USA) em diferentes tempos de incubação, 2 a 3 dias para Alphavirus, 4 a 8 dias para Flavivirus e Orthobunyavirus. Em seguida, os estoques virais foram separados em alíquotas de 0.2 mL, e quantificados pelo ensaio *focus forming units* (FFU) usando células Vero (Tabela 2) [56].

Vírus	Сера	Gênero	FFU	Célula
Dengue 1	Mochizuki	Flavivirus	8,1x104	Vero
Dengue 2	-	Flavivirus	8,1x104	Vero
Dengue 3	-	Flavivirus	2,1x104	Vero
Dengue 4	-	Flavivirus	1,4x106	Vero
Zika	SpH2015	Flavivirus	2,8x106	Vero
Yellow Fever	17DD	Flavivirus	3,3x106	C6/36
West Nile	NY99	Flavivirus	7,5x108	C6/36
Saint Louis	Porton TVP-564	Flavivirus	2,8x107	C6/36
Rocio	SPH 34675	Flavivirus	2,1x106	C6/36
Chikungunya	S27	Alphavirus	3,6x107	Vero
Mayaro	BeAr 20290	Alphavirus	2,4x106	Vero
Oropouche	BeAn 19991	Orthobunyavirus	4,9x104	C6/36
Tacaiuma	BeAn 73	Orthobunyavirus	2,1x106	C6/36
llhéus	HBe 7445	Flavivirus	2,5x107	C6/36

Tabela 2 - Titulação dos vírus. Vírus, linhagem, gênero, FFU e células isoladas de cada vírus.

Legenda: FFU = focus forming units

Para validação metodológica, as amostras clínicas humanas incluíram:

- 41 amostras de plasma previamente positivas para YFV coletadas entre 11 de janeiro e 10 de maio de 2018, com valores de Cts  $\leq$  37 (Tabela 3) obtidas no Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (HC-FMUSP), São Paulo, Brasil (Tabela 3). As amostras foram amplificadas por PCR multiplex e apenas aquelas com amplificação visível em gel de agarose foram sequenciadas [57].

Os participantes ou seus representantes legais forneceram consentimento informado assinado. O panorama ético foi fornecido pelos conselhos de revisão institucional do HC-FMUSP e do Instituto de Doenças Infecciosas "Emílio Ribas ".

**Tabela 3** - Descrição das amostras positivas para vírus da febre amarela (YFV) (n = 41) por reação em cadeia da polimerase quantitativo em tempo real com as amostras correspondentes, valores de Ct, cópias estimadas do genoma por mL e tamanho de referência do vírus.

Amostra	Vírus	Tipo de Amostr a	Valor do Ct	Número estimado de cópias do genoma na amostra (/mL)	Tamanho da referência (nts)
1002	YFV	Plasma	33.4	2.50E+00	10,859
1003	YFV	Plasma	37.0	1.00E+00	10,859
1004	YFV	Plasma	36.0	1.20E+00	10,859
1005	YFV	Plasma	20.1	1.80E+04	10,859
1007	YFV	Plasma	36.8	1.20E+00	10,859
1009	YFV	Plasma	19.1	3.90E+04	10,859
1010	YFV	Plasma	33.7	2.50E+00	10,859
1019	YFV	Plasma	17.2	2.10E+05	10,859
1020	YFV	Plasma	29.5	5.80E+01	10,859
1023	YFV	Plasma	28.0	2.00E+02	10,859
1024	YFV	Plasma	33.0	2.50E+00	10,859
1026	YFV	Plasma	28.4	1.10E+02	10,859
1028	YFV	Plasma	28.2	1.10E+02	10,859
1029	YFV	Plasma	17.4	2.10E+05	10,859
1030	YFV	Plasma	4.6	1.50E+10	10,859
1032	YFV	Plasma	22.9	5.50E+03	10,859
1036	YFV	Plasma	23.3	2.75E+03	10,859
1037	YFV	Plasma	31.9	1.00E+01	10,859
1042	YFV	Plasma	23.3	2.75E+03	10,859
1043	YFV	Plasma	25.8	8.00E+02	10,859
1045	YFV	Plasma	23.7	2.75E+03	10,859

-					
1053	YFV	Plasma	26.6	2.80E+02	10,859
1055	YFV	Plasma	26.9	2.80E+02	10,859
1059	YFV	Plasma	31.0	2.80E+01	10,859
1063	YFV	Plasma	32.0	7.00E+00	10,859
1065	YFV	Plasma	33.0	2.80E+00	10,859
1066	YFV	Plasma	29.3	3.00E+01	10,859
1068	YFV	Plasma	27.3	1.50E+02	10,859
1069	YFV	Plasma	30.1	2.80E+01	10,859
1070	YFV	Plasma	28.0	1.10E+02	10,859
1072	YFV	Plasma	27.6	1.50E+02	10,859
1073	YFV	Plasma	32.5	7.00E+00	10,859
1074	YFV	Plasma	27.9	1.50E+02	10,859
1075	YFV	Plasma	24.1	1.70E+03	10,859
1076	YFV	Plasma	34.8	2.30E+00	10,859
1078	YFV	Plasma	28.6	1.10E+02	10,859
1079	YFV	Plasma	25.5	6.50E+02	10,859
1080	YFV	Plasma	31.6	1.10E+02	10,859
1081	YFV	Plasma	25.2	6.50E+02	10,859
1082	YFV	Plasma	33.8	2.80E+00	10,859
1086	YFV	Plasma	28.2	1.10E+02	10,859

Legenda: YFV = vírus da febre amarela; Ct = limite de ciclo; nts = nucleotídeos;

PCR = reação em cadeia da polimerase.

3.1.2.2 Esquema de Primers gerados durante a validação do protocolo Superpool.

Os esquemas de *primers* para os vírus ZIKV, CHIKV, YFV, MAYV, OROV, WNV, DENV1, DENV2, DENV3, DENV4, ROCV, ILHV, TCV e STLV foram gerados pela ferramenta online *Primal Scheme* [40], com *amplicons* escolhidos entre 400-1000 bp e o *Overlap* entre 75 a 100 nts. Os esquemas gerados e referências usadas estão disponíveis em: https://www.dropbox.com/sh/27jp5s4ynko9i0m/AAC5VTNdM2nhqimXWRbYV2cJa?dl=0. Os *primers* gerados para cada vírus foram pré-diluídos em tampão TE 1X (Cat No. 93302, Sigma-Aldrich, USA) a uma concentração final de 100  $\mu$ M. Para a preparação dos pools dos *primers*, os *primers* a 100  $\mu$ M foram combinados, de acordo com os esquemas gerados, adicionando um volume igual de cada um (5  $\mu$ L), gerando um pool estoque de 100  $\mu$ M (pool A e pool B).

3.1.2.3 Extração de ácido nucléico viral

Para as amostras de referência, ZIKV, YFV, CHIKV, MAYV, OROV, WNV, DENV 1-4, ROCV, ILHV, TCV e STLV, o RNA viral foi isolado de a partir de 200 µl do material titulado usando o QIAamp Viral RNA Mini Kit (Cat No. 52906, Qiagen, Alemanha) de acordo com as instruções do fabricante e eluído em 50 µl de tampão de eluição (EB). Para as amostras clínicas de YFV, o RNA viral foi extraído a partir de 200 µl de material usando a plataforma de extração automática de DNA/RNA do sistema NucliSENS EasyMag (BioMerieux, UK) e eluído em 50 µl.

3.1.2.4 Síntese de cDNA e Multiplex PCR

O cDNA foi sintetizado a partir do RNA de cada amostra previamente extraído utilizando *primers* randômicos hexâmeros (Cat No. N8080127, Thermo Fisher Scientific, EUA) e ProtoScript II Reverse Transcriptase (Cat No. E6560, New England BioLabs, EUA) de acordo com as instruções do fabricante. O cDNA foi então amplificado usando o ensaio multiplex PCR (40) com os esquemas de primers gerados na etapa (3.1.2.2) a 10  $\mu$ M para cada vírus individualmente e combinando todos os esquemas em uma única reação - *Superpool.* (Tabela 4)

**Tabela 4** - Realização do ensaio Multiplex PCR. PCR realizado individualmente e combinando todos os *primers* - *Superpool*. Superpool - *Primers* ZIKV, MAYV, OROV, CHIKV, DENV1-4, YFV, WNV, STLV, TCM, ILHV, ROCV.

Individual Multiplex PCF	र	s
Componentes	Volumes - 25 µL reação final	C
5x Q5 reaction buffer	5 µL	5
10 mM dNTPs	0.5 µL	1
Q5 DNA Polymerase	0.25 μL	G
NFW	15.75 µL para A e B	N
Primer pool (A ou B) 10 μΜ	1 µL pool A e B	P M D
cDNA	2.5 µL	
		c

Superpool	
Componentes	Volumes - 25 µL reação final
5x Q5 reaction buffer	5 µL
10 mM dNTPs	0.5 µL
Q5 DNA Polymerase	0.25 µL
NFW	6.75 µL para A e B
Primer pool (A ou B) ZIKV, MAYV,OROV,CHIKV,YFV, DENV1-4, WNV, STLV, ILHV, ROCV, TACV; 10 µM	10 μL de cada virus pool A 10 μL de cada vírus pool B
cDNA	2.5 µL

Ciclag	em	
98°C	30sec	1x
98°C	1min	
65°C	5min	- 35x
4°C	Hold	

Legenda: YFV = vírus da febre amarela; ZIKV = Zika vírus; MAYV = Mayaro vírus; OROV = Oropouche vírus; CHIKV = Chikungunya vírus; DENV = Dengue vírus; WNV = West nile vírus; STLV = Saint Louis vírus; ILHV = Ilheus vírus; ROCV = Rocio vírus; TACV = Tacaiuma vírus; Ct = limite de ciclo;nts = nucleotídeos;  $\mu$ L = microlitro;  $\mu M = micromolar;$ dNTP = desoxirribonucleotídeos trifosfato; NFW = *Nuclease free water*; PCR = reação em cadeia da polimerase; cDNA = DNA complementar; sec = segundos; min = minutos.

Os produtos de PCR foram então purificados usando uma proporção de 1:1 de *beads* magnéticas AMPure XP (Cat No. A63881, Beckman Coulter, UK) e quantificados usando fluorimetria com o kit Qubit dsDNA High Sensitivity Assay (Cat No. Q32854, Life

Technologies, EUA) e instrumento Qubit 3.0 (Life Technologies, EUA), ambos de acordo com as instruções do fabricante.

Um gel foi preparado com os produtos de PCR gerados utilizando o E-gel Agarose 2% (Cat No. G402002, Thermo Fisher Scientific, EUA) em equipamento E-gel (Thermo Fisher Scientific, EUA) e a corrida foi realizada até que as bandas fossem distinguíveis por transiluminação.

3.1.2.5 Preparação e sequenciamento de biblioteca de nanopore para PCR multiplex

As bibliotecas para sequenciamento MinION, Oxford Nanopore Technologies (ONT, Reino Unido, UK) foram preparadas usando um *input* total de 200 ng combinando as amostras de maneira equimolar usando os kits EXP-NBD104 (1-12) e EXP-NBD114 (13-24) (ONT, Reino Unido, UK). Bibliotecas de sequenciamento foram geradas usando o kit de ligação SQK-LSK109 (ONT, Reino Unido, UK). 20 ng das bibliotecas finais foram carregados em uma *flow cell* FLO-MIN106 no dispositivo MinION (ONT, Reino Unido, UK) e sequenciados usando MinKNOW 1.15.1 (ONT, Reino Unido, UK) com o script de execução padrão de 48 horas.

### 3.1.2.6 Análise bioinformática

Os arquivos brutos FAST5 foram convertidos em FASTQ usando o software Guppy basecaller versão 2.2.7 GPU (ONT, Reino Unido, UK), depois trimados e demultiplexados usando Porechop v.0.3.2pre (https://github.com/rrwick/Porechop) que identifica sequências de DNA diretamente dos dados brutos e encontra e remove os adaptadores ligados durante o sequenciamento. Os arquivos FASTQ das amostras foram alinhados e mapeados contra um painel gerado contendo o genoma referência de cada vírus estudado no painel (número de acesso do GenBank: JF912190 (YFV), KX893855.1 (ZIKV), KP164568.1 (CHIKV), KP691623.1, KP691622.1 KP691621.1 (OROV - segmentos S, M e L), NC\_043694.1, NC\_043693.1, NC\_043692.1 (TACV - segmentos S, M e L), MH643887.1 (WNV), NC\_003417.1 (MAYV), NC\_007580.2 (STLV), NC\_009028.2 (ILHV), NC\_040776.1 (ROCV), KF672760.1 (DENV1), NC\_001474.2 (DENV2), NC\_001475.2 (DENV3) e JN559741.2 (DENV4)), usando minimap2 versão 2.28.0 e convertido para um arquivo BAM

usando samtools sort [58] (painel disponível: https://www.dropbox.com/sh/27jp5s4ynko9i0m/AAC5VTNdM2nhqimXWRbYV2cJa?dl=0.). NanoStat software versão 1.1.2 foi usado para calcular o número total de *reads* (*raw files*) e comprimento mínimo do contig para cobrir 50 por cento do genoma (N50) das *reads* mapeadas (*BAM file*). Tablet 1.19.05.28 [59] foi usado para visualização do genoma e para calcular o número de *reads* mapeadas, porcentagem de cobertura do genoma e profundidade de cobertura.

### 3.1.3 Resultados

Neste estudo, geramos a metodologia *Superpool*, um painel para amplificação de genomas virais, que utiliza uma abordagem de PCR multiplex previamente padronizada [40], combinando em uma só reação 14 esquemas de primers de arbovírus de importância pública para o Brasil.

Os esquemas de *primers* foram gerados para os vírus Zika, Chikungunya, Febre Amarela, Mayaro, Oropouche, *West nile* (WNV), Dengue 1-4 (DENV1, DENV2, DENV3, DENV4), Rocio (ROCV), Ilhéus (ILHV), Tacaiuma (TCV) e Saint Louis (STLV) pela ferramenta online *Primal Scheme*. Os comprimentos dos *amplicons* dos esquemas de *primers* escolhidos foram entre 400-1000 pares de base (bp) e o *Overlap* padrão (comprimento da sobreposição entre os *amplicons* vizinhos) (padrão = 75 a 100 nucleotídeos (nt)) (disponíveis em

https://www.dropbox.com/sh/27jp5s4ynko9i0m/AAC5VTNdM2nhqimXWRbYV2cJa?dl=0). Os esquemas foram preparados para cada vírus individualmente e misturando todos os pools de *primers* gerados, **Superpool**, totalizando aproximadamente 1020 *primers* no total combinados em dois pools (pool A e pool B) (cerca de 35-40 pares de *primers* para cada vírus).

O teste inicial foi realizado utilizando a amostra controle de ZIKV (11474/16) disponível na Universidade de Birmingham, Reino Unido, que foi submetida aos ensaios multiplex PCR [40] e Superpool (apenas com os esquemas de primers dos vírus YFV, MAYV, OROV, CHIKV e ZIKV), seguido de visualização dos produtos amplificados em gel de agarose E-gel. Esta amostra tinha valor de Ct = 25 e concentração de RNA de  $60ng/\mu$ L. Em

ambos os testes, o RNA foi testado em diferentes diluições de 10x em série, compreendendo de 1x até 1: 10,000, equivalente a 60 ng/ $\mu$ L a 6 pg/ $\mu$ L de *input* de material.

As bandas específicas do tamanho correto do esquema foram observadas (em torno de 400 pb) para o vírus testado individualmente e com os *primers* do Superpool. As bandas foram detectadas até a diluição 1:1,000 em ambos os ensaios. Não foram detectadas bandas nos controles negativos testados, indicando que não houve contaminação (Figura 9).



**Figura 9** - Eletroforese em gel de agarose (2% de agarose) de produtos amplificados por multiplex PCR usando conjuntos de primers de PCR específicos para ZIKV e pelo ensaio *Superpool - All* (vírus ZIKV, OROV, MAYV, YFV e CHIKV) utilizando isolado viral de ZIKV (11474/16) em diferentes diluições - 1, 1:10, 1:100, 1:1,000, 1:10,000.

O ensaio multiplex PCR foi então realizado com o esquema de primers de YFV em 41 amostras clínicas humanas positivas para YFV por RT-qPCR coletadas durante a epidemia de YFV de 2018 em São Paulo, Brasil. O valor Ct mediano foi 27,74, variando de 4,6 a 37 correspondendo de 1,50E10 até 1,00E00 cópias do genoma por mL de plasma [57]. Após a quantificação do produto de PCR e a visualização em gel de agarose E-gel, 21 amostras apresentaram bandas específicas distinguíveis por transiluminação e foram selecionadas para continuar a preparação da biblioteca e sequenciamento por nanoporos (Tabela 5). As amostras sequenciadas de YFV (n = 21) tiveram valores médios de Ct = 25,57, entre 5 e 37. A porcentagem de leituras mapeadas ao genoma referência variou de 1,71% a 97,47%, com uma profundidade de cobertura média entre 72,50x a 3370x, e a maioria das amostras com cobertura do genoma em torno de 99,82% sendo a menor 78,11% (Tabela 5).

					%			
			%	Profundidad	Bases			%20x
	Valor do		reads	e média de	com	%		cobertur
Amostra	Ct	n° <i>reads</i>	mapeadas	cobertura	cobertura	Identidade	N50	a
1003	37.0	31,07	51.85	76.65	99.84	88.9	659	86.36
1005	20.1	68,63	24.72	782.25	99.89	88.5	599	91.64
1009	19.1	20,15	91.08	852.99	99.89	88.9	593	94.09
1019	17.2	54,51	83.20	2,11	99.95	89.1	595	91.88
1020	29.5	124,52	1.71	100.93	99.82	88.7	656	89.11
1023	28.0	64,27	69.62	1,95	92.77	91.5	563	61.13
1024	33.0	179,93	15.14	1,22	87.63	91.3	560	55.70
1029	17.4	35,83	91.41	1,51	99.89	89.1	585	91.68
1030	4.6	65,57	97.47	3,10	99.89	88.9	595	91.85
1032	22.9	61,57	42.48	1,20	99.82	89.0	583	91.51
1042	23.3	27,53	6.25	82.43	99.82	88.8	643	87.95
1045	23.7	51,04	90.14	2,11	99.83	88.9	585	91.73
1053	26.6	78,44	91.50	3,37	99.85	89.0	594	93.92
1055	26.9	103,99	4.83	238.17	78.43	91.9	587	31.66
1059	31.0	25,86	5.87	72.50	99.82	88.8	660	88.77
1063	32.0	38,97	4.17	77.05	99.82	88.8	637	91.02
1065	33.0	39,73	5.51	104.05	99.82	88.9	625	91.35
1068	27.3	245,11	2.40	221.91	86.40	91.6	525	44.53
1073	32.5	49,49	11.24	247.57	78.11	90.7	567	35.01
1074	27.9	80,31	3.09	117.94	99.82	88.7	634	91.55
1075	24.1	171,54	21.26	1,57	87.22	91.5	551	65.64

 $\label{eq:table_$ 

Legenda: YFV = Vírus da febre amarela;

Ct = limite do ciclo;

reads = leituras;

N50: Metade da sequência do genoma é coberta por contigs maiores ou iguais ao tamanho do contig N50.

Identidade: Identidade comparada à referência do genoma.

A partir destes dados, foram escolhidas 3 amostras com Cts variando entre 5 e 23 correspondendo de 1,50E10 até 5.50E+03 cópias do genoma por mL de plasma [57]. Essas

amostras foram submetidas ao ensaio Superpool (YFV, MAYV, OROV, CHIKV e ZIKV). Após a quantificação do produto de PCR foi realizada a preparação da biblioteca de sequenciamento por nanoporos. As amostras sequenciadas de YFV (n = 3) tiveram porcentagem de leituras mapeadas ao genoma referência variou de 91.61% a 94,12%, com uma profundidade média de cobertura entre 469.174x a 5,151,06x para os Cts 4.6 e 23 respectivamente (Tabela 6). O teste Superpool foi realizado para avaliar se o ensaio com um grande número de pares de *primers* na mesma reação afetaria a cobertura do genoma alvo devido às interações entre os *primers* em amostras clínicas. Quando analisado o gráfico de cobertura das *reads* alinhadas (BAM) mapeadas ao genoma referência (Figura 10), os resultados demonstraram que o ensaio Superpool pode gerar uma cobertura genômica semelhante ao teste de *primers* de arbovírus presentes individualmente conforme o genoma.

Tabela 6 - Informações detalhadas do sequenciamento do genoma do vírus da Febre Amarela, utilizando ométodo PCR Multiplex - Superpool.

Amostras	Valor do Ct	n⁰ <i>reads</i>	% Bases com cobertura	Profundidade máxima de cobertura	Profundidade média de cobertura
1029 Superpool	17.4	119,016	94.12%	13,333	3,392.106
1030 Superpool	4.6	182,054	94.11%	19,820	5,151.06
1032 Superpool	23.0	15,929	91.61%	2,023	469.174

Legenda: Ct = limite do ciclo; n<sup>o</sup> = número; *reads* = leituras; % = porcentagem;



**Figura 10** - Gráficos de cobertura da amostra 1029 amplificadas pelo esquema de PCR Multiplex de YFV (painel superior) e para o esquema Superpool (painel inferior) gerado pela ferramenta Tablet.

Por último, o ensaio Superpool foi investigado aumentando para 14 o número esquemas de *primers* na mesma reação (ZIKV, CHIKV, YFV, MAYV, OROV, DENV1-4, WNV, SLV, ROCV, TCMV e ILHV), somando aproximadamente 1020 *primers* no total. Este ensaio foi testado em isolados virais de ZIKV, CHIKV, YFV, MAYV, OROV, DENV1-4, WNV, SLV, ROCV, TCMV e ILHV fornecidos pelo Centro de Pesquisa em Virologia da Faculdade de Medicina de Ribeirão Preto, Ribeirão Preto (SP), Brasil. Os isolados tinham entre 2,1E04, a 7,5E08 focus forming units FFUs (Tabela 2). [56]

Começando com o ensaio multiplex PCR, os isolados virais de DENV-1, DENV-2, ZIKV, CHIKV, WNV, STLV, ROCV, MAYV, OROV, TCMV e ILHV foram sequenciados com os primers individuais de cada vírus respectivamente (Tabela 7). A porcentagem de cobertura variou entre 88.83% para o segmento M do vírus TCMV até 100% para os vírus WNV, ROCV, MAYV, ILHV e OROV segmentos L e M. A profundidade média apresentada foi de 640.787 com mínimo de 400.389 e máximo de 15,557.461x. As mesmas amostras de cDNA dos vírus ZIKV, DENV-1, DENV-2, MAYV e OROV foram selecionadas e submetidas ao método Superpool com os 14 esquemas de primers descritos acima (Tabela 7). As maiores porcentagens de leitura de bases mapeadas observadas foram 100% e 98.7% para os vírus MAYV e OROV segmento S, respectivamente, com uma média encontrada de 94,8%.

Comparamos a porcentagem de bases com cobertura com as amostras de ZIKV, DENV-1, DENV-2, MAYV e OROV para cada método - multiplex PCR e Superpool (Figura 11). A comparação revelou um padrão de cobertura semelhante entre os dois métodos e melhor cobertura com o método Superpool para o vírus DENV-2 quando comparado com o método multiplex PCR.

Amostras	Primer	n⁰ <i>reads</i>	n⁰ <i>reads</i> mapeadas	Profundidade máxima de cobertura	% Bases com cobertura	N50
DENV-1	DENV-1	31,481	18,038	400.389	97.402	828
DENV-2	DENV-2	35,868	6,798	598.544	99.736	1,008
ZIKV	ZIKV	20,821	15,349	901.015	98.566	666
CHIKV	CHIKV	15,714	10,833	657.346	95,064	920
WNV	WNV	19,024	14,347	662.495	100	587
STLV	STLV	39,535	32,047	1,405.016	96.106	586

**Tabela** 7 - Informações detalhadas sobre o sequenciamento ONT, usando os esquemas de Multiplex PCR individualmente ou contendo todos os *primers* - *Superpool*.

ROCV	ROCV	32,463	27,551	1,289.4	100	592
MAYV	MAYV	28,668	16,939	1,063.07	100	591
OROV	OROV	104,702	S - 38,895 L - 14,218 M - 33,739	S - 15,557.461 L - 989.94 M - 3,669.255	S - 98.707 L - 100 M - 100	565
TCMV	TCMV	63,665	S - 11,919 L - 36,504 M - 9,052	S - 1,299.424 L - 11,440.746 M - 624.934	S - 99.954 L - 99.179 M - 88.83	446
ILHV	ILHV	21,101	4,441	253,314	100	560
ZIKV	Superpool	15,188	8,596	381,768	98,575	561
DENV-1	Superpool	24,811	17,482	337.603	97.189	569
DENV-2	Superpool	23,105	13,678	497.863	98.578	448
MAYV	Superpool	16,499	7,347	483.394	100	594
OROV	Superpool	58,163	S - 17,848 L - 364 M - 3,545	S - 7,276.109 L - 25.273 M - 385.11	S - 98.707 L - 73.428 M - 97.044	388

Legenda: YFV = vírus da febre amarela;

ZIKV = Zika vírus;

MAYV = Mayaro vírus;

OROV = Oropouche vírus;

CHIKV = Chikungunya vírus;

DENV = Dengue vírus;

WNV = West nile vírus;

STLV = Saint Louis vírus;

ILHV = Ilheus vírus; ROCV = Rocio vírus;

TACV = Tacaiuma vírus;

Superpool = Combinação de todos os primers;

 $n^{\circ} = número;$ 

*reads* = leituras;

% = porcentagem;

N50 = Metade da sequência do genoma é coberta por contigs maiores ou iguais ao tamanho do contig N50.



**Figura 11**: Porcentagem de cobertura de cada vírus amplificado pelos métodos Superpool e com o método multiplex PCR, esquemas individuais. Legenda: ZIKV = Zika vírus; OROV = Oropouche vírus; DENV = Dengue vírus; MAYV = Mayaro vírus; Percentage Coverage = porcentagem da cobertura; % = porcentagem.

**3.2 Capítulo 2 -** Desenvolvimento de uma abordagem rápida utilizando a fusão de barcodes para amplificação do genoma completo de múltiplos vírus num único ensaio denominado SMART-plex.

## 3.2.1 Resumo

Descrita originalmente em 2001, a estratégia frequentemente chamada de SMART (Switching mechanism at 5'end of RNA transcript) (Takara Bio USA, Inc) mostrou-se promissora na geração de bibliotecas de cDNA, mesmo a partir de amostras de RNA derivadas de células [55]. A maioria dos métodos já descritos baseiam-se na iniciação a partir de oligo-dT, seguido da amplificação alvo de moléculas de mRNA poliadeniladas. Neste protocolo, adaptamos a tecnologia SMART como um novo método universal de rápida detecção, amplificação e preparação de bibliotecas denominado SMART-plex. Neste método, usamos apenas os primers reversos (right) gerados pelo software primal scheme para cada vírus estudado (podemos aqui adaptar o método para todos os vírus validados no superpool) e adicionamos uma tag na região 5' de cada primer, para a síntese da primeira fita de cDNA (Tabela 8).

Tabela 8- Exemplo	do esquema dos pri	mers right gerad	o pela ferramenta	a Primal Scheme	com a adição d	la <i>tag</i>
(em negrito) com a	sequência da ONT	para amplificaçã	ão do genoma co	ompleto do vírus	Zika (ZIKV) o	com o
ensaio SMART-plex						

Primer	Sequência (5'-3')					
ZIKA_1_RLB_9	TTTTCGTGCGCCGCTTCAACAGTATGCACTC					
ZIKA_2_RLB_9	TTTTCGTGCGCCGCTTCAACTGATTCCAACC					
ZIKA_3_RLB_9	TTTTTCGTGCGCCGCTTCAACTACGGTGACAC					
ZIKA_4_RLB_9	TTTTTCGTGCGCCGCTTCAACGGAGCCATG					
ZIKA_5_RLB_9	TTTTCGTGCGCCGCTTCAACTGTGCGTCC					
ZIKA_6_RLB_9	TTTTCGTGCGCCGCTTCAACCCATCTGTC					
ZIKA_7_RLB_9	TTTTCGTGCGCCGCTTCAACCGCCTCC					
ZIKA_8_RLB_9	TTTTTCGTGCGCCGCTTCAACTTGACTGCTGC					
ZIKA_9_RLB_9	TTTTTCGTGCGCCGCTTCAACGAGTGGGCA					
ZIKA_10_RLB_9	TTTTTCGTGCGCCGCTTCAACGTGGGACTTT					
ZIKA_11_RLB_9	TTTTCGTGCGCCGCTTCAACCCTGGGC					
ZIKA_12_RLB_9	TTTTTCGTGCGCCGCTTCAACTATCAGCGCC					
ZIKA_13_RLB_9	TTTTTCGTGCGCCGCTTCAACAGAGAGAGAGGAGC					

ZIKA_14_RLB_9	TTTTTCGTGCGCCGCTTCAACTTTCCCATGT
ZIKA_15_RLB_9	TTTTCGTGCGCCGCTTCAACTACACTCCATC
ZIKA_16_RLB_9	TTTTTCGTGCGCCGCTTCAACGCTCCAATGT
ZIKA_17_RLB_9	TTTTTCGTGCGCCGCTTCAACCCTCTAAGGG
ZIKA_18_RLB_9	TTTTTCGTGCGCCGCTTCAACTGGTGAGTTGG
ZIKA_19_RLB_9	TTTTTCGTGCGCCGCTTCAACGCCATCAAGT
ZIKA_20_RLB_9	TTTTCGTGCGCCGCTTCAACCCTTTGCTC
ZIKA_21_RLB_9	TTTTTCGTGCGCCGCTTCAACCTCCAAAAGCC
ZIKA_22_RLB_9	TTTTTCGTGCGCCGCTTCAACATTCTGGCTG
ZIKA_23_RLB_9	TTTTTCGTGCGCCGCTTCAACAAGTGGTCACT
ZIKA_24_RLB_9	TTTTTCGTGCGCCGCTTCAACTCTCCACTTG
ZIKA_25_RLB_9	TTTTCGTGCGCCGCTTCAACCCTTCCATT
ZIKA_26_RLB_9	TTTTTCGTGCGCCGCTTCAACACCAGGGC
	TTTTCCTCCCCCCTTCAACATCTCTACACTT
ZIKA_27_RLB_9	TITICGIGCGCCGCTICAACATGIGIAGAGIT
ZIKA_27_RLB_9 ZIKA_28_RLB_9	TTTTCGTGCGCCGCTTCAACAGGGCCTC
ZIKA_27_RLB_9 ZIKA_28_RLB_9 ZIKA_29_RLB_9	TTTTCGTGCGCCGCTTCAACATGTGTAGAGTT TTTTCGTGCGCCGCTTCAACGGGCCTC TTTTTCGTGCGCCGCTTCAACATGCTGCATTG
ZIKA_27_RLB_9 ZIKA_28_RLB_9 ZIKA_29_RLB_9 ZIKA_30_RLB_9	TTTTCGTGCGCCGCTTCAACATGTGTAGAGTT         TTTTTCGTGCGCCGCTTCAACGGGCCTC         TTTTTCGTGCGCCGCTTCAACATGCTGCATTG         TTTTTCGTGCGCCGCTTCAACTAATCCCAGC
ZIKA_27_RLB_9 ZIKA_28_RLB_9 ZIKA_29_RLB_9 ZIKA_30_RLB_9 ZIKA_31_RLB_9	TTTTCGTGCGCCGCTTCAACATGTGTAGAGTT         TTTTTCGTGCGCCGCTTCAACATGCTGCATTG         TTTTTCGTGCGCCGCTTCAACTAATCCCAGC         TTTTTCGTGCGCCGCTTCAACCGTAAGTGA CAAC
ZIKA_27_RLB_9 ZIKA_28_RLB_9 ZIKA_29_RLB_9 ZIKA_30_RLB_9 ZIKA_31_RLB_9 ZIKA_32_RLB_9	TTTTCGTGCGCCGCTTCAACATGTGTAGAGTT         TTTTTCGTGCGCCGCTTCAACGGGCCTC         TTTTTCGTGCGCCGCTTCAACATGCTGCATTG         TTTTTCGTGCGCCGCTTCAACCAACTAATCCCAGC         TTTTTCGTGCGCCGCTTCAACCGTAAGTGA CAAC         TTTTTCGTGCGCCGCTTCAACTGTCCCATC
ZIKA_27_RLB_9 ZIKA_28_RLB_9 ZIKA_29_RLB_9 ZIKA_30_RLB_9 ZIKA_31_RLB_9 ZIKA_32_RLB_9 ZIKA_33_RLB_9	TTTTTCGTGCGCCGCTTCAACATGTGTAGAGTT         TTTTTCGTGCGCCGCTTCAACGGGCCTC         TTTTTCGTGCGCCGCTTCAACATGCTGCATTG         TTTTTCGTGCGCCGCTTCAACCAACTAATCCCAGC         TTTTTCGTGCGCCGCTTCAACCGTAAGTGA CAAC         TTTTTCGTGCGCCGCTTCAACTGTCCCATC         TTTTTCGTGCGCCGCTTCAACATCCACACTC
ZIKA_27_RLB_9 ZIKA_28_RLB_9 ZIKA_29_RLB_9 ZIKA_30_RLB_9 ZIKA_31_RLB_9 ZIKA_32_RLB_9 ZIKA_33_RLB_9 ZIKA_34_RLB_9	TTTTTCGTGCGCCGCTTCAACATGTGTAGAGTT         TTTTTCGTGCGCCGCTTCAACATGCTGCATTG         TTTTTCGTGCGCCGCTTCAACATGCTGCATCG         TTTTTCGTGCGCCGCTTCAACCGTAAGTGA CAAC         TTTTTCGTGCGCCGCTTCAACCGTAAGTGA CAAC         TTTTTCGTGCGCCGCTTCAACCGTAAGTGA CAAC         TTTTTCGTGCGCCGCTTCAACATGTCCCATC         TTTTTCGTGCGCCGCTTCAACATCCACACTC         TTTTTCGTGCGCCGCTTCAACTGACTAGCAGG

Em seguida, a partir do mecanismo derivado das enzimas transcriptases reversas (RTases) do tipo MMLV (*Moloney Murine Leukemia Vírus*), nucleotídeos não modelados (principalmente desoxicitidinas) são adicionados à extremidade 3' da fita de cDNA recém-sintetizada, podemos assim utilizar o mecanismo de *template switching* através dos *primers* SSP (*Strand Switching Primers*) que possuem sequências rGrGrG complementares às desoxicitidinas adicionadas durante a transcrição. Estes produtos de cDNA marcados podem ser então amplificados utilizando barcodes já disponíveis da Oxford Nanopore Technologies (ONT, UK) (*Rapid PCR barcoding kit*) que contém a mesma sequência das *tags* adicionadas aos *primers* gene específicos e SSPs, além de uma molécula química, por via de uma reação de PCR convencional. Os produtos em seguida são purificados e os adaptadores de sequenciamento podem ser adicionados rapidamente via *click chemistry*, uma reação química

que torna não necessária a utilização de enzimas de ligação e preparação de biblioteca de sequenciamento (Figura 12).

\_ \_ \_ \_ \_ \_

Reverse transcription with multiplex primers



Template switching with SSP

Amplification with barcoded primers

**Figura 12** - Esquema do protocolo **SMART-plex**. Legenda: *Reverse transcription with multiplex primers* = transcrição reversa com primers multiplex; *Template switching with SSP* = *template switching* com primers (*Strand switching primers*) SSP; *Amplification with barcoded primers* = amplificação com primers com barcodes.

O PCR Multiplex embora seja o método de escolha no momento, é um método demorado (>10 horas) com grande demanda de trabalho manual. Neste protocolo também investigamos o resultado da utilização de diferentes *primers* de diferentes tamanhos na mesma reação de PCR para avaliar a sua capacidade de amplificar e gerar o genoma completo de arbovírus presentes individualmente ou/e em misturas a fim de gerar um completo painel rápido com os principais arboviruses circulantes no Brasil.

O protocolo prossegue em seis estágios: (i) preparação dos pools dos *primers SMART-plex*, (ii) extração do RNA, (iii) síntese de cDNA, (iv) PCR, (v) preparo da biblioteca de sequenciamento e sequenciamento MinION e (vi) análise bioinformática e controle de qualidade (QC).

3.2.2 Material e métodos

3.2.2.1 Seleção de amostras

Utilizamos para desenvolvimento dos métodos amostras de isolados do vírus Zika -Referência: OMS - (11474/16), disponível em *School of Biosciences*, *University of Birmingham*, Reino Unido.

Para validação metodológica, as amostras clínicas humanas incluíram:

- 41 amostras de plasma previamente positivas para YFV coletadas entre 11 de janeiro e 10 de maio de 2018, com valores de Cts  $\leq$  37 (Tabela 3) obtidas no Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (HC-FMUSP), São Paulo, Brasil (Tabela 9) [57].

Os participantes ou seus representantes legais forneceram consentimento informado assinado. O panorama ético foi fornecido pelos conselhos de revisão institucional do HC-FMUSP e do Instituto de Doenças Infecciosas "Emílio Ribas ".

Amostra	Vírus	Tipo de Amostr a	Valor Ct	Número estimado de cópias do genoma na amostra (/mL)	Tamanho da referência (nts)
1002	YFV	Plasma	33.4	2.50E+00	10,859
1003	YFV	Plasma	37.0	1.00E+00	10,859
1004	YFV	Plasma	36.0	1.20E+00	10,859
1005	YFV	Plasma	20.1	1.80E+04	10,859
1007	YFV	Plasma	36.8	1.20E+00	10,859
1009	YFV	Plasma	19.1	3.90E+04	10,859
1010	YFV	Plasma	33.7	2.50E+00	10,859
1019	YFV	Plasma	17.2	2.10E+05	10,859
1020	YFV	Plasma	29.5	5.80E+01	10,859
1023	YFV	Plasma	28.0	2.00E+02	10,859

**Tabela 9** - Descrição das amostras positivas para vírus da febre amarela (YFV) (n = 41) por reação em cadeia da polimerase quantitativo em tempo real (RT-qPCR) com as amostras correspondentes, valores de Ct, cópias estimadas do genoma por mL e tamanho de referência do vírus.

1024	YFV	Plasma	33.0	2.50E+00	10,859
1026	YFV	Plasma	28.4	1.10E+02	10,859
1028	YFV	Plasma	28.2	1.10E+02	10,859
1029	YFV	Plasma	17.4	2.10E+05	10,859
1030	YFV	Plasma	4.6	1.50E+10	10,859
1032	YFV	Plasma	22.9	5.50E+03	10,859
1036	YFV	Plasma	23.3	2.75E+03	10,859
1037	YFV	Plasma	31.9	1.00E+01	10,859
1042	YFV	Plasma	23.3	2.75E+03	10,859
1043	YFV	Plasma	25.8	8.00E+02	10,859
1045	YFV	Plasma	23.7	2.75E+03	10,859
1053	YFV	Plasma	26.6	2.80E+02	10,859
1055	YFV	Plasma	26.9	2.80E+02	10,859
1059	YFV	Plasma	31.0	2.80E+01	10,859
1063	YFV	Plasma	32.0	7.00E+00	10,859
1065	YFV	Plasma	33.0	2.80E+00	10,859
1066	YFV	Plasma	29.3	3.00E+01	10,859
1068	YFV	Plasma	27.3	1.50E+02	10,859
1069	YFV	Plasma	30.1	2.80E+01	10,859
1070	YFV	Plasma	28.0	1.10E+02	10,859
1072	YFV	Plasma	27.6	1.50E+02	10,859
1073	YFV	Plasma	32.5	7.00E+00	10,859
1074	YFV	Plasma	27.9	1.50E+02	10,859
1075	YFV	Plasma	24.1	1.70E+03	10,859
1076	YFV	Plasma	34.8	2.30E+00	10,859
1078	YFV	Plasma	28.6	1.10E+02	10,859
1079	YFV	Plasma	25.5	6.50E+02	10,859
1080	YFV	Plasma	31.6	1.10E+02	10,859
1081	YFV	Plasma	25.2	6.50E+02	10,859
1082	YFV	Plasma	33.8	2.80E+00	10,859
1086	YFV	Plasma	28.2	1.10E+02	10,859

Legenda: YFV = vírus da febre amarela; Ct = limite de ciclo;

nts = nucleotídeos; PCR = reação em cadeia da polimerase.

3.2.2.2 Primers gerados durante a validação do protocolo SMART-plex.

Para o ensaio *SMART-plex* os *primers* foram gerados pela mesma ferramenta *Primal Scheme* (https://primalscheme.com) utilizada para o desenvolvimento do método Superpool, seguindo as mesmas instruções anteriores para cada vírus. No ensaio *SMART-plex* sintetizamos apenas os *primers Right* (reversos), de diferentes tamanhos, com diferentes *tags* na região 5' de cada *primer*. Além disso, é necessário a síntese do *primer* SSP (*Strand Switching primer*) contendo a mesma *tag* dos *primers* genes específicos seguido da sequência rGrGrG.

Para a padronização deste ensaio foram testados diferentes sequências das *tags* de diferentes kits (ONT - PCR-cDNA Sequencing kit, NEB Single Cell/Low Input RNA Library Prep Kit for Illumina<sup>®</sup> é ONT - *Rapid barcoding kit - RLB*). Também foram testados diferentes comprimentos das sequências contendo os nucleotídeos gene específicos do vírus Zika gerados no *Primal Scheme* (18 nt, 15 nt, 12 nt e 8 nt).

Primers testados durante a validação do protocolo SMART-plex:

Oligos ONT - PCR-cDNA Sequencing kit

Tag + Zika\_400\_35\_RIGHT\_\*

/5phos/ACT TGC CTG TCG CTC TAT CTT CAC CAC TAG TCC CTC TTC TG - 3' (tag + 18nt)

/5phos/ACT TGC CTG TCG CTC TAT CTT CAC CAC TAG TCC CTC TT- 3' (tag + 15nt)

/5phos/ACT TGC CTG TCG CTC TAT CTT CAC CAC TAG TCC CT- 3' (tag + 12nt) /5phos/ACT TGC CTG TCG CTC TAT CTT CAC CAC TAG TC- 3' (tag + 9nt)

SSP\_3mG

## 5' - TTT CTG TTG GTG CTG ATA TTG C TrGrGrG - 3'

PR2 (e.g. in PRM of cPRM)

# TTTCTGTTGGTGCTGATATTGC

Oligos NEB

**Tag** + Zika\_400\_35\_NEB\_\*

AAG CAG TGG TAT CAA CGC AGA GTA CAC CAC TAG TCC CTC TTC T- 3' (tag + 18nt)

AAG CAG TGG TAT CAA CGC AGA GTA CAC CAC TAG TCC CTC TT - 3' (tag + 15nt)

AAG CAG TGG TAT CAA CGC AGA GT A CAC CAC TAG TCC CT- 3' (tag + 12nt) AAG CAG TGG TAT CAA CGC AGA GT A CAC CAC TAG TC - 3' (tag + 9nt)

NEBNext Template Switching Oligo (SSP)

5' - GCT AAT CAT TGC AAG CAG TGG TAT CAA CGC AGA GT A CAT rGr GrG - 3

NEBNext Single Cell cDNA PCR *Primer* (PCR) 5<sup>'</sup>- AAG CAG TGG TAT CAA CGC AGA GT - 3<sup>'</sup>

Oligos RLB - SMART-plex

SMART-plex YFV (all YFV\_Right\_500) YFV\_500\_Right\_1\_RLB - YFV\_500\_Right\_1\_27\_RLB

RLB SSP GCT AAT CAT TGC **TTT TTC GTG CGC CGC TTC A**AC ATrG rGrG

RLB PCR TTT TTC GTG CGC CGC TTC A

Os esquemas de primers para os vírus ZIKV, CHIKV, YFV, MAYV, OROV, WNV, DENV1, DENV2, DENV3, DENV4, ROCV, ILHV, TCV e STLV foram gerados durante validação do ensaio Superpool com amplicons escolhidos entre 400-1000 bp e o Overlap entre 75 a 100 nt, disponíveis em: https://www.dropbox.com/sh/27jp5s4ynko9i0m/AAC5VTNdM2nhqimXWRbYV2cJa?dl=0. Após validação do método SMART-plex, os primers foram adaptados adicionando a tag e os primers right foram gerados para cada vírus, disponíveis em: https://www.dropbox.com/sh/27jp5s4ynko9i0m/AAC5VTNdM2nhqimXWRbYV2cJa?dl=0. Após síntese, os primers foram pré-diluídos em tampão TE 1X (Cat No. 93302, Sigma-Aldrich, USA) a uma concentração final de 100 µM. Para a preparação dos pools dos primers, os primers a 100 µM foram combinados, de acordo com os esquemas gerados, adicionando um volume igual de cada um (5  $\mu$ L), gerando um pool estoque de 100  $\mu$ M.

### 3.2.2.3 Extração de ácido nucléico viral

Para a amostra referência de ZIKV o RNA viral foi isolado de a partir de 200 μl do material titulado usando o QIAamp Viral RNA Mini Kit (Cat No. 52906, Qiagen, Alemanha) de acordo com as instruções do fabricante e eluído em 50 μl de tampão de eluição (EB). Para as amostras clínicas de YFV, o RNA viral foi extraído a partir de 200 μl de material usando a plataforma de extração automática de DNA/RNA do sistema NucliSENS EasyMag (BioMerieux, UK) e eluído em 50 μl.

### 3.2.2.4 Síntese de cDNA SMART-plex e PCR rápido

Para a síntese do cDNA do ensaio SMART-plex, 10  $\mu$ l do RNA extraído, 1  $\mu$ l do *pool* de *primers* SMART-plex *primers right*, 2  $\mu$ M e 1  $\mu$ l de dNTPs 10 mM (Cat No. N0447L, New England BioLabs, EUA) foram misturados e incubados por 5 min a 65 ° C, depois colocados em gelo. 4  $\mu$ l do buffer SuperScript IV First-Strand Synthesis System, 1  $\mu$ l de DTT 0,1 M, 1  $\mu$ l de RNase OUT, 1  $\mu$ l de RLB TSO (oligo de RNA) (GCTAATCATTGCTTTTTTCGTGCGCCGCTTCAACATrGrGrG, 2  $\mu$ M), e 1  $\mu$ L de SuperScript IV (Cat No. 18091200, Thermo Fisher Scientific, USA) foram misturados com 12  $\mu$ l do RNA previamente anelado e incubados por 90 min a 42 ° C seguido por 10 min a 70

 $^{\circ}$  C. Esses produtos de cDNA foram então amplificados usando o kit RLB (1-12) SQK-RPB004 kit (ONT, Reino Unido) ou RLB PCR (TTTTTCGTGCGCCGCTTCA, 20  $\mu$ M) como controle da PCR (Tabela 10).

Tabela 10 - Realização do ensaio SMART-plex PCR.

LongAmp Tog 2X mostor mix	25	Ciclagem			
	25 u		95°C	45sec	1x
RLB (01-12) ou RLB-PCR (20uM)	0.5 ul ou 1 ul		95°C 15sec		
			62°C	15sec	26x
SMART cDNA	5 ul		65°C	5min	
NEW	10 5 vil ov 10 vil		65°C	10min	1x
INT" VV	19.5 ul 60 19 ul		4°C	Hold	

Legenda: NFW = *Nuclease free water*; PCR = reação em cadeia da polimerase; cDNA = DNA complementar; sec = segundos; min = minutos.

Os produtos de PCR foram então purificados usando uma proporção de 1:1 de *beads* magnéticas AMPure XP (Cat No. A63881, Beckman Coulter, UK) e quantificados usando fluorimetria com o kit Qubit dsDNA High Sensitivity Assay (Cat No. Q32854, Life Technologies, EUA) e instrumento Qubit 3.0 (Life Technologies, EUA), ambos de acordo com as instruções do fabricante.

Um gel foi preparado com os produtos de PCR gerados utilizando o E-gel Agarose 2% (Cat No. G402002, Thermo Fisher Scientific, EUA) em equipamento E-gel (Thermo Fisher Scientific, EUA) e a corrida foi realizada até que as bandas fossem distinguíveis por transiluminação.

3.2.2.5 Preparação e sequenciamento de biblioteca de nanopore ensaio SMART-plex

O preparo das bibliotecas foi realizado seguindo as instruções do fabricante. As bibliotecas MinION foram preparadas usando 50 ng de cada cDNA amplificado, agrupados de maneira equimolar - 400 ng em 10  $\mu$ l de Tris-HCl 10 mM pH 8,0 com NaCl 50 mM. 1  $\mu$ l de adaptador RAP foi adicionado à biblioteca e incubado à temperatura ambiente durante 5 min.

50 ng das bibliotecas finais foram carregados em *flow cells* FLO-MIN106 no dispositivo MinION (ONT, Reino Unido) e sequenciados usando MinKNOW 1.15.1 com o script de execução padrão de 48 horas.

### 3.2.2.6 Análise bioinformática

Os arquivos brutos FAST5 foram convertidos em FASTQ usando o software Guppy basecaller versão 2.2.7 GPU (ONT, Reino Unido, UK), depois trimados e demultiplexados usando Porechop v.0.3.2pre (https://github.com/rrwick/Porechop) que identifica sequências de DNA diretamente dos dados brutos e encontra e remove os adaptadores ligados durante o sequenciamento. Os arquivos FASTQ das amostras foram alinhados e mapeados contra um painel gerado contendo o genoma referência de cada vírus estudado no painel (número de acesso do GenBank: JF912190 (YFV), KX893855.1 (ZIKV), KP164568.1 (CHIKV), KP691623.1, KP691622.1 KP691621.1 (OROV - segmentos S, M e L), NC 043694.1, NC 043693.1, NC 043692.1 (TACV - segmentos S, M e L), MH643887.1 (WNV), NC 003417.1 (MAYV), NC 007580.2 (STLV), NC 009028.2 (ILHV), NC 040776.1 (ROCV), KF672760.1 (DENV1), NC 001474.2 (DENV2), NC 001475.2 (DENV3) e JN559741.2 (DENV4)), usando minimap2 versão 2.28.0 e convertido para um arquivo BAM usando samtools sort [58] (painel disponível: https://www.dropbox.com/sh/27jp5s4ynko9i0m/AAC5VTNdM2nhqimXWRbYV2cJa?dl=0.). NanoStat software versão 1.1.2 foi usado para calcular o número total de reads (raw files) e comprimento mínimo do contig para cobrir 50 por cento do genoma (N50) das reads mapeadas (BAM file). Tablet 1.19.05.28 [59] foi usado para visualização do genoma e para calcular o número de *reads* mapeadas, porcentagem de cobertura do genoma e profundidade de cobertura.

## 3.2.3 Resultados

Neste estudo, desenvolvemos a tecnologia **SMART-plex**, uma abordagem baseada no kit já disponível e validado NEBNext Single-cell/low-input RNA (cat No. E6420, New England BioLabs, USA). Modificamos esse método para utilização de *primers* gene específicos marcados com uma *tag* adicional para anelamento inicial, seguido pela síntese de cDNA com *primers Strand Switching* (SSP). A posterior amplificação é realizada usando uma versão de kit de sequenciamento compatível da Oxford Nanopore Technologies (ONT, UK), que possui *primers* complementares às *tags* já contendo barcodes ligados. A ligação dos *barcodes* durante o ensaio PCR permite que o preparo da biblioteca de sequenciamento que antes era de 6 horas, passe a ser 10 minutos, sendo necessário apenas a ligação dos adaptadores, via *click chemistry*, permitindo uma rápida detecção, caracterização e sequenciamento do genoma completo dos vírus alvos (Figura 13). Comparando a tecnologia SMART-plex com a tecnologia padrão-ouro já padronizada Multiplex PCR (40), o tempo do fluxo de trabalho laboratorial caiu de 14 horas para 5 horas e 40 minutos (62%).



**Figura 13** - Comparação entre as etapas e o tempo dos fluxos de trabalho - sequenciamento *tilling amplicon* - Multiplex PCR e SMART-plex. O método Rapid SMART-9N contém o esquema *strand switching* com primers gene específicos marcados como parte da síntese de cDNA SMART e PCR com primers marcados contendo barcodes. Legendas: hrs = horas; cDNA synthesis and Multiplex PCR = síntese do DNA completar e PCR multiplex; Clean up and quantification = purificação e quantificação; Library preparation = preparo da biblioteca; Ligation of barcodes = ligação dos barcodes; Ligation of adapters = ligação dos adaptadores; SMART cDNA synthesis and barcoding PCR = Síntese de cDNA SMART e PCR com *barcoding*; Ligation of adapter via click chemistry = Ligação dos adaptadores via *click* químico; min = minutos.

Os esquemas de primers foram gerados para os vírus Zika, Chikungunya, Febre Amarela, Mayaro, Oropouche, West nile (WNV), Dengue 1-4 (DENV1, DENV2, DENV3, DENV4), Rocio (ROCV), Ilhéus (ILHV), Tacaiuma (TCV) e Saint Louis (STLV) pela ferramenta online Primal Scheme durante a validação do teste Superpool. Os comprimentos dos *amplicons* dos esquemas de *primers* escolhidos foram entre 400-1000 pares de base (bp) e o Overlap padrão (comprimento da sobreposição entre os amplicons vizinhos) (padrão = 75 a 100 nucleotídeos (nt)) (disponíveis em https://www.dropbox.com/sh/27jp5s4ynko9i0m/AAC5VTNdM2nhqimXWRbYV2cJa?dl=0). Para o método SMART-plex, adaptamos os primers gerados acima e utilizamos apenas os primers reversos (right) em diferentes tamanhos para cada vírus individualmente e primer. adicionando regiões 5' de cada tags nas (disponíveis em https://www.dropbox.com/sh/27jp5s4ynko9i0m/AAC5VTNdM2nhqimXWRbYV2cJa?dl=0).

O teste inicial foi realizado utilizando a amostra controle de ZIKV (11474/16) disponível na Universidade de Birmingham, Reino Unido, que foi submetida ao ensaio SMART-plex. Esta amostra tinha valor de Ct = 25 e concentração de RNA de 60 ng/μL. Primeiro foi realizada a comparação entre os *primers* contendo a tag dos kits ONT - PCR-cDNA Sequencing (/5phos/ACT TGC CTG TCG CTC TAT CTT C) e NEB - Single Cell/Low Input RNA Library Prep Kit for Illumina<sup>®</sup>, (AAG CAG TGG TAT CAA CGC AGA GT) adicionadas aos *primers* gene específicos de ZIKV e primers *Strand Switching* (SSP). A única diferença entre os dos kits são as sequências das *tags* e os comprimentos das sequências dos *primers* SSP - 23 nt para o kit da ONT e 39 nt para o kit da NEB e a adição da sequência /5phos/ nos primers polyA do kit da ONT. Para esse teste, foi utilizado apenas um *primer* - Tag + Zika\_400\_35\_RIGHT (18nt) para início da transcrição reversa. A síntese de cDNA foi realizada a uma temperatura de 42°C que é a temperatura ideal para o mecanismo de *Strand Switching* descrita nos protocolos dos kits testados para o protocolo SMART.



**Figura 14** - Gráficos de cobertura da amostra de referência - ZIKV para o ensaio SMART-plex utilizando *primers* contendo as sequências da ONT - gerado por Tablet.



**Figura 15** - Gráficos de cobertura da amostra de referência - ZIKV para o ensaio SMART-plex utilizando *primers* contendo as sequências da NEB - gerado por Tablet.

Esse primer deveria amplificar apenas região alvo do а primer Zika 400 35 RIGHT, a região circulada em vermelho nas figuras 14 e 15 de aproximadamente 400 bp. No entanto, o teste resultou em apenas amplificações inespecíficas. Isso ocorreu porque a temperatura de anelamento dos primers gene específicos são maiores > 65°C, devido ao tamanho dos primers gerados + tag. No entanto, a temperatura ideal para o mecanismo Strand Switching é de 42 °C resultando em anelamento inespecífico dos primers sobre o genoma do ZIKV.

Com esse resultado do teste realizado, no entanto, pudemos notar uma grande diferença no tamanho das *reads* geradas quando utilizados os *primers* contendo as sequências

do kit NEB Single Cell/Low Input RNA Library Prep Kit for Illumina<sup>®</sup>. Além disso, notamos que para conseguir uma temperatura de anelamento não tão alta, que fosse compatível com a temperatura de anelamento dos *primers* gene específicos e de transcrição reversa dos *primers* SSP, teríamos que usar os *primers* mais curtos, contendo apenas 9 nucleotídeos com as sequências dos *primers* gene específicos + *tag*.

Portanto, no teste seguinte foram testadas diferentes temperaturas de anelamento para a síntese de cDNA - 42, 47 e 52°C, utilizando *primers* contendo 9 nts + a *tag* do kit da ONT, buscando amplificação específica dos *primers* contendo as sequências do ZIKV (Figura 16).



**Figura 16** - Gráficos de cobertura da amostra de referência - ZIKV para o ensaio SMART-plex utilizando diferentes temperaturas para o teste da transcrição reversa, 42°C, 47°C e 52°C, utilizando *primers* contendo as sequências da ONT - gerado por Tablet.

Com esse teste notamos que a temperatura ideal para a reação é a de 52°C. Com essa temperatura, obtivemos uma melhor amplificação na região alvo e menor amplificação inespecífica observada pelo genoma do ZIKV quando comparado com os testes anteriores.

Por último, testamos essa mesma condição com os primers da NEB. Por apresentar tamanhos de reads maiores quando comparado com os primers da ONT (Figuras 14 e 15), o

que seria importante para o teste, testamos os *primers* com 9 nucleotídeos + *tag* NEB com a temperatura de transcrição reversa a 52°C.



**Figura 17** - Gráficos de cobertura da amostra de referência - ZIKV para o ensaio SMART-plex utilizando temperatura de transcrição reversa de 52°C. *Primers* utilizados: 9 nucleotídeos gene específicos + *tag* com sequências da NEB - gerado por Tablet.

A partir dos resultados apresentados com os *primers* da NEB com a temperatura ideal da reação, adaptamos os primers da ONT, para que possuísse a sequência complementar do kit rápido de sequenciamento da ONT e também tivesse como produto de sequenciamento *reads* mais longas como apresentado na figura 17. Adicionamos então nucleotídeos complementares da NEB à *tag* da ONT (SSP: GCT AAT CAT TGC **TTT TTC GTG CGC CGC TTC A**AC ATrG rGrG). O próximo passo foi testar as condições padronizadas descritas anteriormente com amostras clínicas de YFV já sequenciadas pelo ensaio padronizado multiplex PCR (25), com o ensaio SMART-plex, e também multiplexando todos os primers gerados pelo *Primal Scheme* SMART-plex-YFV. Foram selecionadas amostras com valores de Ct entre 17 e 36. (Tabela 11) (Figura 18).

Amostras	Valor do Ct	n⁰ <i>reads</i>	Cobertura	%Identidade	N50
1055_YFV_plex	26.9	233824	100%	91.6	403.0
1004_YFV_plex	36.0	39636	100%	91.9	569.0
1020_YFV_plex	29.5	278728	100%	90.9	439.0
1045_YFV_plex	23.7	105912	100%	91.6	461.0
1005_YFV_plex	20.1	106090	95.2%	91.5	445.0
1019_YFV_plex	17.2	21305	100%	91.5	933.0

**Tabela 11** - Resumo dos dados de sequenciamento de nanoporos de vírus usando a abordagem SMART-plex com amostras positivas para o vírus da febre amarela (YFV) (n = 6) com os valores Ct correspondentes.



Figura 18: Gráfico de cobertura da amostra ARB 1019 para o ensaio SMART-plex gerado por Tablet.

O teste obteve cobertura genômica de 100% em 5 das 6 amostras testadas, com a menor cobertura de 95.2%. Para a amostra 1019, a profundidade média de cobertura foi de 879 e profundidade máxima de 1,334, demonstrando que o teste SMART-plex é uma metodologia que poderá ser utilizada para sequenciamento de arbovírus após finalização dos testes e validação. O resultado apresentado aqui revelou melhor cobertura genômica que os outros protocolos testados anteriormente - PCR Multiplex, com cobertura genômica média de 94%. Ainda esse protocolo é de menor complexidade e compatível com o kit *Rapid Barcoding kit* da ONT, que após validação, será possível realizar todo o procedimento em até 4 horas e adicionando um maior número de barcodes para diminuir o custo da reação.

**3.3 Capítulo 3** - Desenvolvimento de um protocolo de metagenômica viral *SMART-9N* e um protocolo rápido de metagenômica *Rapid SMART-9N* para rastreio de vírus de RNA e potencial descoberta de novos vírus

### 3.3.1 Resumo

Os vírus de RNA são responsáveis por causar uma ampla gama de doenças humanas e veterinárias. Nas últimas décadas, os vírus de RNA têm sido uma das principais causas de infecções emergentes e reemergentes, incluindo o vírus Zika (ZIKV), o vírus da dengue (DENV), o vírus da imunodeficiência humana (HIV), o vírus Ebola (EBOV), o vírus da febre amarela (YFV), e recentemente, síndrome respiratória aguda grave coronavírus 2 (SARS-CoV-2). As epidemias e pandemias resultantes causaram alta morbidade, mortalidade e custos econômicos. [10]

Até o momento, nossa capacidade de gerenciar esses surtos é prejudicada pelo desafio de fazer um diagnóstico clínico definitivo, pois muitos desses vírus são frequentemente indistinguíveis daqueles causados por vírus co-circulantes e alguns patógenos bacterianos [35, 36]. Os testes diagnósticos podem ser limitados por baixa especificidade, no caso dos testes sorológicos, ou exigir o conhecimento *a priori* dos vírus a serem alvo no caso de reação em cadeia da polimerase quantitativo em tempo real (RT-qPCR). Por essas razões, a doença febril aguda frequentemente permanece sem diagnóstico, levando à falha da vigilância epidemiológica. Os sistemas de vigilância genômica rápida são essenciais para identificar vírus emergentes, detectar e monitorar a diversidade viral e serem ferramentas capazes de preparar ou até mesmo prevenir novos surtos. [60]

Novas ferramentas foram impulsionadas por avanços tecnológicos em sequenciamento. Os primeiros exemplos de vigilância genômica em tempo real [16, 39] foram conduzidos a partir de sequenciamento de amplicons direcionados no equipamento MinION (Oxford Nanopore Technologies (ONT)). Esses estudos exploraram a portabilidade do sequenciamento de nanoporos para obter um tempo de resposta mais rápido, sequenciando as amostras perto de onde foram coletadas. Embora bem-sucedida para a epidemia de EBOV na África Ocidental e surtos de ZIKV, CHIKV, DENV e YFV no Brasil [23, 44, 61], esta abordagem é melhor quando a cepa do surto é conhecida, mas é menos adequada para diversos grupos virais ou para descoberta de vírus.

A metagenômica viral, o processo de sequenciamento do conteúdo total de ácido nucleico em uma amostra (normalmente cDNA ou DNA), permite a caracterização genômica de vírus conhecidos e novos vírus de maneira não direcionada. Essa técnica é particularmente útil para diagnósticos, laboratórios clínicos e vigilância genômica em saúde pública [48-50]. No entanto, o sequenciamento metagenômico viral diretamente de amostras clínicas pode resultar em baixa sensibilidade, especialmente em amostras com uma baixa abundância de material genômico viral em relação ao ácido nucleico derivado do hospedeiro [52-54]. Durante o projeto ZiBRA em 2016, não foi possível gerar sequências genômicas completas diretamente de amostras clínicas usando abordagens metagenômicas [40]. Além da baixa carga viral apresentada nas amostras (com valores de Cts entre 33,9 e 35,9 - equivalente a 10-48 cópias do genoma por microlitro), as *flow cells* R6 (chips para sequenciamento ONT) disponíveis geravam uma média de 500 Mb de dados por sequenciamento, sendo considerado na época, uma tecnologia de baixo rendimento. Neste caso, as sequências do vírus Zika continham <0,01% do conjunto de dados, resultando em cobertura incompleta. Maior cobertura e profundidade são críticas para a reconstrução precisa do genoma e subsequente inferência filogenética.

Recentemente, melhorias no sequenciamento por nanoporos, (lançamento da *flow cell* R9.5.1) e novos kits de sequenciamento (SQK-LSK109) estão permitindo gerar >50 Gb de dados por sequenciamento, indicando que sequenciamento metagenômico diretamente de amostras clínicas poderia ser agora viável, o que potencialmente significa que novos vírus e linhagens podem ser detectados a partir de um teste rápido e universal. O sequenciamento metagenômico usando tecnologia nanopore já se mostrou promissor por Kafetzopoulou et al. (2018) que relatou o sequenciamento metagenômico de amostras de LASV, DENV e CHIKV [62], e por Lewandowski et al. (2019) que sequenciou o vírus Influenza de amostras respiratórias [63].

Neste estudo, descrevemos uma abordagem *SMART (Switching mechanism at 5' end of RNA Template)* de alta sensibilidade e baixo *input* para metagenômica utilizando a tecnologia nanopore, para detecção e sequenciamento de vírus de RNA de amostras de isolados ou diretamente de amostras clínicas. A abordagem *SMART* foi originalmente descrita em 2001 [55], em um protocolo em que foi utilizado oligo-dT *priming* para direcionar moléculas de mRNA poliadeniladas. Adaptamos este método para *priming* randômico para a síntese de cDNA seguido de amplificação por PCR, e em seguida adaptamos com primers de

PCR ligados aos *barcodes* - *Rapid SMART-9N*, permitindo a amplificação viral e adição de *barcodes* em uma única etapa.

O método SMART-9N recuperou uma alta proporção de *reads* virais do isolado de ZIKV em até 6e00 FFU/mL de *input* de material, incluindo 94,4% do genoma em uma única *read*. Os métodos foram validados em YFV e SARS-CoV-2 diretamente de amostras clínicas de plasma e de nasofaringe residual, respectivamente. O desempenho deste ensaio foi comparado a um método de PCR multiplex padrão-ouro [40], demonstrando melhorias na sensibilidade do sequenciamento, cobertura, profundidade, custo e complexidade de ambos *SMART-9N* e *Rapid SMART-9N*, permitindo uma detecção aprimorada de patógenos para aplicação de diagnóstico e vigilância de vírus de RNA.

3.3.2 Material e métodos

3.3.2.1 Seleção de amostras

Utilizamos para testes iniciais dos métodos amostras de isolados do vírus Zika -Referência: OMS - (11474/16), disponível na *School of Biosciences*, *University of Birmingham*, Reino Unido.

Para desenvolvimento do método usamos uma cepa isolada de ZIKV BeH815744 (GenBank No. de Acesso KU365780) propagada em células Vero (CCL-81; ATTC, Manssas, EUA) com meio essencial mínimo (MEM) por 2 horas a 37 ° C e 5% de CO2. O sobrenadante foi removido e o MEM suplementado com 2% de soro fetal bovino, 1% de penicilina e estreptomicina foram adicionados. As células foram incubadas por 4 dias até 70% do efeito citopático. Posteriormente, o sobrenadante da cultura celular foi coletado e a replicação viral foi confirmada por meio da reação de RT-qPCR [64] e quantificada por um ensaio FFU em células Vero [56]. Esta amostra foi usada para avaliar o desempenho de todos os três métodos: PCR multiplex, *SMART-9N* e *Rapid SMART-9N*.

As abordagens metagenômicas, *SMART-9N* e *Rapid SMART-9N* também foram testadas em diferentes diluições de 10x em MEM até 1-1,000.000 para avaliação do limite de detecção dos métodos (Tabela 12, Tabela 13).
Amostra	Multiplex PCR	SMART-9N	Rapid SMART-9N
ZIKV referência	Х	Х	Х
YFV amostra clínica	X	Х	Х
SARS-CoV-2 amostra clínica	X		Х

Tabela 12 - Descrição das amostras coletadas e protocolo realizado para cada amostra.

Legenda: ZIKV = vírus Zika;

YFV = vírus da febre amarela;

SARS-CoV-2 = síndrome respiratória aguda grave coronavírus 2.

**Tabela 13** - Descrição das amostras positivas para a amostra de referência do zika vírus cepa BeH815744 (ZIKV) (n = 1), vírus da febre amarela (YFV) (n = 41) e síndrome respiratória aguda grave coronavírus 2 (SARS-CoV-2) (n = 10) por reação em cadeia da polimerase em tempo real (RT-qPCR) com os tipos de amostra correspondentes, valores Ct, unidades formadoras de foco estimadas (FFU) por mililitro ou cópias estimadas do genoma por mL e o tamanho da referência de cada vírus (nts).

Amostra	Vírus	Tipo de amostra	Valor do Ct	Estimados valores de unidades formadoras de foco (FFU) por mililitro ou cópias do genoma por mL	Tamanho de referência (nts)
ZIKV Ref	ZIKV	Isolate	15.1	6E+07	10,808
1002	YFV	Plasma	a 33.4 2.50E+00		10,859
1003	YFV	Plasma	asma 37.0 1.00E+00		10,859
1004	YFV	Plasma	asma 36.0 1.20E+00		10,859
1005	YFV	Plasma	20.1 1.80E+04		10,859
1007	YFV	Plasma	36.8	1.20E+00	10,859
1009	YFV	Plasma	19.1	3.90E+04	10,859
1010	YFV	Plasma	33.7	2.50E+00	10,859
1019	YFV	Plasma	17.2	2.10E+05	10,859
1020	YFV	Plasma	29.5 5.80E+01		10,859
1023	YFV	Plasma	28.0	2.00E+02	10,859
1024	YFV	Plasma	33.0	2.50E+00	10,859
1026	YFV	Plasma	28.4	1.10E+02	10,859
1028	YFV	Plasma	28.2	1.10E+02	10,859
1029	YFV	Plasma	17.4	2.10E+05	10,859
1030	YFV	Plasma	4.6	1.50E+10	10,859
1032	YFV	Plasma	22.9	5.50E+03	10,859
1036	YFV	Plasma	23.3	2.75E+03	10,859
1037	YFV	Plasma	31.9	1.00E+01	10,859
1042	YFV	Plasma	23.3	2.75E+03	10,859
1043	YFV	Plasma	25.8	8.00E+02	10,859

1045	YFV	Plasma	23.7	2.75E+03	10,859
1053	YFV	Plasma	26.6	2.80E+02	10,859
1055	YFV	Plasma	26.9	2.80E+02	10,859
1059	YFV	Plasma	31.0	2.80E+01	10,859
1063	YFV	Plasma	32.0	7.00E+00	10,859
1065	YFV	Plasma	33.0	2.80E+00	10,859
1066	YFV	Plasma	29.3	3.00E+01	10,859
1068	YFV	Plasma	27.3	1.50E+02	10,859
1069	YFV	Plasma	30.1	2.80E+01	10,859
1070	YFV	Plasma	28.0	1.10E+02	10,859
1072	YFV	Plasma	27.6	1.50E+02	10,859
1073	YFV	Plasma	32.5	7.00E+00	10,859
1074	YFV	Plasma	27.9	1.50E+02	10,859
1075	YFV	Plasma	24.1	1.70E+03	10,859
1076	YFV	Plasma	34.8	2.30E+00	10,859
1078	YFV	Plasma	28.6	1.10E+02	10,859
1079	YFV	Plasma	25.5	6.50E+02	10,859
1080	YFV	Plasma	31.6	1.10E+02	10,859
1081	YFV	Plasma	25.2	6.50E+02	10,859
1082	YFV	Plasma	33.8	2.80E+00	10,859
1086	YFV	Plasma	28.2	1.10E+02	10,859
2159	SARS-CoV-2	NPS	32.2	2.80E+02	29,903
2163	SARS-CoV-2	NPS	27.6	5.40E+03	29,903
2168	SARS-CoV-2	NPS	33.3	1.30E+02	29,903
2173	SARS-CoV-2	NPS	25.5	2.10E+04	29,903
2182	SARS-CoV-2	NPS	28.7	2.80E+03	29,903
2187	SARS-CoV-2	NPS	31.1	5.80E+02	29,903
2192	SARS-CoV-2	NPS	27.5	5.80E+03	29,903
2207	SARS-CoV-2	NPS	28.9	2.40E+03	29,903
5148	SARS-CoV-2	NPS	22.6	1.50E+05	29,903
5157	SARS-CoV-2	NPS	21.8	2.40E+05	29,903

Legenda: ZIKV = vírus Zika; YFV = vírus da febre amarela; SARS-CoV-2 = síndrome respiratória aguda grave coronavírus 2;

Ct = limite de ciclo

FFU = unidades formadoras de foco

nts = nucleotídeos;

NPS = esfregaço nasofaríngeo

Para validação metodológica, as amostras clínicas humanas incluíram:

- 41 amostras de plasma previamente positivas para YFV por RT-qPCR [65] coletadas entre 11 de janeiro e 10 de maio de 2018, com valores de Ct  $\leq$  37 (Tabela 13) obtidos no Hospital das Clínicas, Faculdade de Medicina da Universidade de São Paulo (HC-FMUSP), São Paulo, Brasil. As amostras foram amplificadas por multiplex PCR e apenas aquelas amplificadas e visíveis em gel de agarose foram sequenciadas. Destas, as amostras com uma gama de valores de Ct foram selecionadas para sequenciamento *SMART-9N* e *Rapid SMART-9N* (Tabela 12);

- Dez amostras residuais de nasofaringe previamente positivas para SARS-CoV-2 por RT-qPCR [66], coletadas entre 17 de novembro de 2020 e 5 de janeiro de 2021, com valores de Ct  $\leq$  34, obtidas no Instituto de Medicina Tropical, Faculdade de Medicina da Universidade de São Paulo, Brasil (Tabela 13). Essas amostras foram selecionadas para os métodos multiplex PCR e *Rapid SMART-9N* (Tabela 12).

Os participantes ou seus representantes legais forneceram consentimento informado assinado. O panorama ético foi fornecido pelos conselhos de revisão institucional do HC-FMUSP e do Instituto de Doenças Infecciosas "Emílio Ribas " para o estudo YFV. Para o estudo SARS-CoV-2, a aprovação ética foi feita pelo comitê nacional de ética em pesquisa (Comissão Nacional de Ética em Pesquisa), protocolo número CAAE 30127020.0.0000.0068.

## 3.3.2.2 Extração de ácido nucléico e teste RT-qPCR

As amostras foram centrifugadas por 5 minutos a 10.000 xg. Para os isolados de ZIKV, o RNA viral foi isolado de 200 µl do material recuperado usando o kit QIAamp Viral RNA Mini (Cat No. 52906, Qiagen, Alemanha) de acordo com as instruções do fabricante e eluído em 50 µl de tampão de eluição (EB) . Para as amostras nasofaríngeas de SARS-CoV-2 e plasma de YFV, o RNA viral foi extraído de 200 µl de material usando a plataforma de extração automática de DNA / RNA do sistema NucliSENS EasyMag (BioMerieux, Reino Unido) e eluído em 50 µl. Os valores de Ct foram previamente determinados para todas as amostras por RT-qPCR para ZIKV [64], YFV [65] e SARS-CoV-2 [66].

O RNA extraído foi tratado com TURBO DNase (Cat No. AM2239, Thermo Fisher Scientific, EUA) a 37 ° C por 30 min para remoção de DNA residual e concentrado em 10  $\mu$ l usando o kit Zymo RNA clean & concentrator-5 (Cat No. R1016, Zymo Research, EUA), seguindo as instruções dos fabricantes.

#### 3.3.2.3 Multiplex tiling PCR

As amostras foram submetidas ao sequenciamento do genoma completo usando uma abordagem padrão-ouro de sequenciamento de amplicon denominada multiplex PCR [40]. Resumidamente, o cDNA foi sintetizado a partir de amostras de RNA positivas para os vírus testados usando primers random hexamers (Cat No. N8080127, Thermo Fisher Scientific, EUA) e ProtoScript II reverse transcriptase (Cat No. E6560, New England BioLabs, EUA) de acordo com as instruções do fabricante. O cDNA foi então amplificado com o ensaio multiplex PCR previamente padronizado para ZIKV, YFV [40, 44], e com o esquema ARTIC V3 para SARS-CoV-2 [67] (https://artic.network/ncov- 2019). Os produtos de PCR foram purificados usando uma proporção de 1:1 de beads magneticas AMPure XP (Cat No. A63881, Beckman Coulter, UK) e quantificados usando fluorimetria com o kit Qubit dsDNA High Sensitivity Assay (Cat No. Q32854, Life Technologies, EUA) no instrumento Qubit 3.0 (Life Technologies, EUA), ambos de acordo com as instruções do fabricante. Um gel foi preparado com os produtos de PCR gerados utilizando E-gel Agarose 2% (Cat No. G402002, Thermo Fisher Scientific, EUA) e o equipamento E-gel (Thermo Fisher Scientific, EUA) e a corrida foi realizada até que as bandas fossem distinguíveis por transiluminação. As amostras que apresentaram bandas específicas foram selecionadas para o sequenciamento MinION (ONT, UK).

3.3.2.4 Preparação e sequenciamento da biblioteca de sequenciamento nanopore para o ensaio multiplex PCR

Bibliotecas MinION foram preparadas usando um *input* total de 100 ng, os *barcodes* foram ligadas e agrupados de forma equimolar usando os kits EXP-NBD104 (1-12) e EXP-NBD114 (13-24) (ONT, UK). Bibliotecas de sequenciamento foram geradas usando o kit SQK-LSK109 (ONT, UK). 20 ng das bibliotecas finais foram carregados em novas *flow* 

*cells* FLO-MIN106 no dispositivo MinION (ONT, UK) e sequenciados usando MinKNOW 1.15.1 com o script de execução padrão de 48 horas.

3.3.2.5 Síntese de cDNA SMART-9N

Para a síntese de cDNA SMART-9N, 10 µl do RNA concentrado, 1 µl NEB-RT 9N (AAGCAGTGGTATCAACGCAGAGTACNNNNNNNN, 2µM) e 1 µl 10 mM de deoxyribonucleotide triphosphate (dNTPs) mix (Cat No. N0447L, New England Biolabs, USA) foram misturados e incubados por 5 min a 65 ° C, em seguida, colocados em gelo. 4 µl da SuperScript IV first-strand buffer, 1 µl de 0,1 M DTT, 1 µl de RNase OUT, 1 µl de NEB-SSP (RNA oligo) (GCTAATCATTGCAAGCAGTGGTATCAACGCAGAGTACATrGrGrG, 2 µM) e 1 µL de SuperScript IV (Cat No. 18091200, Thermo Fisher Scientific, USA) foram misturados com 12 µl do RNA previamente anelado antes da incubação por 90 min a 42 ° C seguido por 10 min a 70 ° C. Esses produtos de cDNA double-tagged foram amplificados usando 5 µl de Q5 reaction buffer (Cat No. M0491, New England BioLabs, EUA), 0,5 µl 10 µM dNTP, 1 µl NEB-PCR (AAGCAGTGGTATCAACGCAGAGT, 20 µM), 15,75 µl de nuclease free-water (NFW), 0,25 µl de Q5 DNA polymerase e 2,5 µl de cDNA. As condições dos ciclos da PCR foram: 98 ° C por 45 s, seguido por 30 ciclos de 98 ° C por 15 s, 62 ° C por 15 s e 65 ° C por 5 min e uma etapa final de 65 ° C por 10 min. Os produtos foram purificados usando uma proporção de 1:1 de *beads* magnéticas AMPure XP (Cat No. A63881, Beckman Coulter, UK) e quantificados usando fluorimetria com o ensaio Qubit dsDNA High Sensitivity (Cat No. Q32854, Life Technologies, EUA) e instrumento Qubit 3.0 (Life Technologies, EUA), ambos de acordo com as instruções do fabricante.

3.3.2.6 Preparação da biblioteca e sequenciamento nanopore para o ensaio SMART-9N

Bibliotecas MinION foram preparadas usando um *input* total de 50 ng para cada amostra, os *barcodes* foram ligados e agrupados de forma equimolar usando os kits EXP-NBD104 (1-12) e EXP-NBD114 (13-24) (ONT, UK). Bibliotecas de sequenciamento foram geradas usando o kit SQK-LSK109 (ONT, UK). 50 ng das bibliotecas finais foram

carregados em novas *flow cells* FLO-MIN106 no dispositivo MinION (ONT, UK) e sequenciados usando MinKNOW 1.15.1 com o script de execução padrão de 48 horas.

#### 3.3.2.7 Síntese de cDNA *Rapid SMART-9N*

Para a síntese de cDNA do ensaio Rapid SMART-9N, 10 µl do RNA concentrado, 1 μl RLB RT tagged 9N (TTTTTCGTGCGCCGCTTCAACNNNNNNNN, 2 μM) e 1 μl de dNTPs 10 mM (Cat No. N0447L, New England BioLabs, EUA) foram misturados e incubados por 5 min a 65 ° C, depois colocados em gelo. 4 µl da SuperScript IV first-strand *buffer*, 1 µl de 0,1 M DTT, 1 µl de RNase OUT, 1 µl de RLB TSO (oligo de RNA) (GCTAATCATTGCTTTTTCGTGCGCCGCTTCAACATrGrGrG, 2 µM) e 1 µL SuperScript IV (Cat No. 18091200, Thermo Fisher Scientific, USA) foram misturados com 12 µl do RNA previamente anelado antes da incubação por 90 min a 42 ° C seguido por 10 min a 70 ° C. Esses produtos de cDNA double-tagged foram amplificados usando 25 µl de LongAmp Taq 2X master mix (Cat No. M0287, New England BioLabs, USA), 19,5 µl de NFW, 0,5 µl de RLB (1-12) do kit SQK-RPB004 (ONT, Reino Unido) e 5 µl de cDNA. As condições dos ciclos de PCR foram: 95 ° C por 45 s, seguido por 25-30 ciclos de 95 ° C por 15 s, 56 ° C por 15 s e 65 ° C por 5 min e uma etapa final de 65 ° C por 10 min. Para a etapa de PCR, o RLB PCR (TTTTTCGTGCGCCGCTTCA, 20 µM) pode ser usado como controle de PCR, alterando as condições dos ciclos para 98 ° C por 45 s, seguido por 25-30 ciclos de 98 ° C por 15 s, 62 ° C por 15 s e 65 ° C por 5 min e uma etapa final de 65 ° C por 10 min. Os produtos foram purificados usando uma proporção de 1:1 de beads magnéticas AMPure XP (Cat No. A63881, Beckman Coulter, UK), quantificados usando fluorimetria com o kit Qubit dsDNA High Sensitivity Assay (Cat No. Q32854, Life Technologies, EUA) e instrumento Qubit 3.0 (Life Technologies, EUA), ambos de acordo com as instruções do fabricante.

3.3.2.8 Preparação e sequenciamento da biblioteca de sequenciamento nanopore para o ensaio *Rapid SMART-9N* 

As bibliotecas MinION foram preparadas usando 50 ng de cada cDNA amplificado, agrupados de modo equimolar em um total de 400 ng em 10 µl de Tris-HCl 10 mM pH 8,0 com NaCl 50 mM. 1 µl do adaptador RAP foi adicionado à biblioteca e incubado à

temperatura ambiente durante 5 min. 50 ng das bibliotecas finais foram carregados em novas *flow cells* FLO-MIN106 no dispositivo MinION (ONT, Reino Unido) e sequenciados usando MinKNOW 1.15.1 com o script de execução padrão de 48 horas.

### 3.3.2.9 Análise bioinformática

Os arquivos FAST5 gerados (raw files) foram transformados em FASTQ files (basecalling) usando o software Guppy versão 2.2.7 GPU basecaller (Oxford Nanopore depois demultiplexados e trimados usando Porechop v.0.3.2pre Technologies), (https://github.com/rrwick/Porechop). Os arquivos FASTQ de cada barcode foram alinhados e mapeados contra o genoma referência de cada vírus estudado (número de acesso do GenBank JF912190 (YFV), KX893855.1 (ZIKV) e MN908947 (SARS-CoV-2)) usando minimap2 versão 2.28.0 e convertido em arquivos BAM usando SAMtools [58]. NanoStat versão 1.1.2 foi usado para calcular o número de reads totais (raw files) e o comprimento mínimo do contig para cobrir 50 por cento do genoma (N50) das reads alinhadas. Tablet 1.19.05.28 [59] foi usado para visualização do genoma e para calcular o número de reads mapeadas, porcentagem de cobertura do genoma e profundidade de cobertura. samtools stats e samtools depth [58] foram usados para calcular o tamanho das reads mais longas (long *reads*) e a cobertura do genoma em 20x, respectivamente. Para a análise do ensaio multiplex PCR, length filtering, quality test, e primmer trimming foram realizados para cada barcode usando artic guppyplex e variant calling e geração das sequências consensos usando a minion Nanopolish versão artic do pipeline de bioinformática ARTIC (https://github.com/artic-network / fieldbioinformatics). Para os ensaios SMART-9N e o Rapid SMART-9N, variant calling foi realizado com medaka variants e a sequência consenso foi construída usando medaka consensus (ONT, UK).

Para a busca de outro RNA viral nas amostras clínicas, a classificação taxonômica foi realizada usando Kraken versão 2.0.8-beta [68], MiniKraken2\_v1\_8GB Kraken 2 Database [https://ccb.jhu.edu/software/kraken2/downloads. shtml].

### 3.3.3 Resultados

Neste estudo, desenvolvemos duas metodologias, **SMART-9N** e **Rapid SMART-9N**. A abordagem SMART-9N é baseada no kit já disponível e validado NEBNext Single-cell/low-input RNA (cat No. E6420, New England BioLabs, USA), modificado para o uso de *primers* randômicos (9N) para anelamento do material genético disponível e preparação de bibliotecas de sequenciamento utilizando *barcodes* nativos (ONT, UK). O método NEB usa amplificação por *single-primer* PCR, que descobrimos que poderíamos realizar usando primers já contendo os *barcodes* do kit Rapid PCR Barcoding (ONT, UK) se modificarmos a sequência dos oligos RT e SSP. Esta abordagem permite a amplificação do RNA com *input* de material na faixa de picogramas, tornando-a ideal para aplicações de baixo *input*. Comparamos a complexidade, os custos e o tempo necessário para o trabalho laboratorial com uma abordagem de multiplex PCR previamente padronizada [40]. Em comparação com o multiplex PCR, o tempo total de trabalho laboratorial caiu 15% e 61% para o SMART-9N e Rapid SMART-9N, respectivamente, e os custos com reagentes foram reduzidos em 40% e 45% (Figura 19).



Abreviaturas: SSP: Strand Switching Primer. US \$, dólar americano.

**Figura 19** - Comparação entre as etapas e o custo dos fluxos de trabalho - Multiplex PCR, SMART-9N e Rapid SMART-9N. Legendas: hrs = horas; cDNA synthesis and Multiplex PCR = síntese do DNA completar e PCR multiplex; Clean up and quantification = purificação e quantificação; Library preparation = preparo da

biblioteca; Ligation of barcodes = ligação dos barcodes; Ligation of adapters = ligação dos adaptadores; SMART cDNA synthesis and PCR = Síntese de cDNA SMART e PCR; SMART cDNA synthesis and barcoding PCR = Síntese de cDNA SMART e PCR com *barcoding*; Ligation of adapter via click chemistry = Ligação dos adaptadores via *click* químico; min = minutos.

O teste inicial foi realizado utilizando a amostra controle de ZIKV (11474/16) disponível na Universidade de Birmingham, Reino Unido, que foi submetida aos ensaios *SMART-9N* nas diluições 1:100, 1:1000, 1:10000 (Tabela 14). Esta amostra tinha valor de Ct = 25 e concentração de RNA de 60 ng/ $\mu$ L. O teste resultou em uma cobertura genômica de 99.7% para todas as diluições testadas com profundidade de cobertura de até 3,000 com um *input* de apenas 5 pg (Figura 20), compatível com ensaios de uma única célula - *single cell* [69]. Ainda, o N50 apresentado foi de 2,9kb para a diluição 1:100, e quando analisadas as leituras individualmente, o teste obteve coberturas do genoma completo do ZIKV em uma única *read* (aproximadamente 11Kb) (Figura 21).

**Tabela 14** - Informações detalhadas do sequenciamento do genoma do vírus Zika referência, utilizando o método SMART-metagenomics com diferentes concentrações de input.

Amostra	n⁰ <i>reads</i>	% <i>reads</i> mapeadas	Profundidade de cobertura	% Cobertura do genoma	%identidad e	N50
ZIKV 1:100	5698	51	779.258	99.7%	92.8	2,967
ZIKV 1:1000	33433	33	1,932.953	99.7%	92.8	1,971
ZIKV 1:10000	38943	47	3,447.144	99.7%	92.6	2,095

KJ776791.2   consens	sus length: 10,807 (0) Advanced	reads: 18,143   features: 22	,374 Memory usage: 1431.99 M	B (19)					
Open Assembly Features Data	Read Packing	Q Zoom:	Page Left Prev Feature Prev View	<ul> <li>Page Right</li> <li>Next Feature</li> <li>Next View</li> </ul>	👆 Jump to Base	Read Info	RS Off RS Centre RS Custom	Show Cigar	
	visual	Aujust		Navigate			Overlays		
Contigs: 1	18.14 k rea	ids (more)		Bran ( M and a state of the sta					
Contig Length Reads	Features Miss	match % 1 to 10,807 (10.8	Kb) ( Internet internet					1	to 7,617 (7.6 Kb)
KJ776791.2 10,807 18,1	43 22,374	7.4							
		CICAR-D							
		CICAR-RICHT-CLIP					n		
		OCAR-LEFT-CLP			+11+111111-++				
									7,616 07,616
									D D
									Y
Filter by: Name									

**Figura 20 -** Gráfico de cobertura da amostra referência de ZIKV revelado pelo ensaio SMART-9N com input de 5 pg gerado por Tablet.



**Figura 21 -** Gráfico de cobertura da amostra referência de ZIKV revelado pelo ensaio SMART-9N mostrando a leitura de 10,2 Kb gerado por Tablet.

Sequenciamento multiplex PCR do isolado de ZIKV e amostras clínicas YFV e SARS-CoV-2

O teste seguinte foi realizado em uma diluição em série de 10x do isolado de ZIKV BeH815744, que foi submetido à abordagem padrão-ouro multiplex PCR seguida pelo sequenciamento MinION [40]. Esta amostra tinha um valor de Ct=15,1 e um título de RNA de 6e07 FFU / mL (Tabela 13). A porcentagem de leituras mapeadas foi de 55,99% com uma profundidade média de 326,98x cobrindo 98,74% do genoma viral coberto com pelo menos 1 *read* (Tabela 15) (Figura 22).

**Tabela 15** - Resumo dos dados de sequenciamento de nanoporos de vírus usando a abordagem de multiplex PCR utilizando. Amostras - cepa de referência do vírus Zika BeH815744 (ZIKV) (n = 1), vírus da febre amarela (YFV) (n = 21) e de síndrome respiratória aguda grave coronavírus 2 (SARS-CoV-2) (n = 10) com os valores dos Cts correspondentes.

Amostra	Vírus	Valor do Ct	% <i>reads</i> mapeada	Média Profundidad	% Genoma	%	N50	%20x
ZIKV Ref		15.1	55.99	326.08	08.74	83.3	445	97 10
		07.0	53.55	70.05	30.74	00.0	445	97.10
1003	YEV	37.0	51.85	76.65	99.84	88.9	659	86.36
1005	YFV	20.1	24.72	782.25	99.89	88.5	599	91.64
1009	YFV	19.1	91.08	852.99	99.89	88.9	593	94.09
1019	YFV	17.2	83.20	2,11	99.95	89.1	595	91.88
1020	YFV	29.5	1.71	100.93	99.82	88.7	656	89.11
1023	YFV	28.0	69.62	1,95	92.77	91.5	563	61.13
1024	YFV	33.0	15.14	1,22	87.63	91.3	560	55.70
1029	YFV	17.4	91.41	1,51	99.89	89.1	585	91.68
1030	YFV	4.6	97.47	3,10	99.89	88.9	595	91.85
1032	YFV	22.9	42.48	1,20	99.82	89.0	583	91.51
1042	YFV	23.3	6.25	82.43	99.82	88.8	643	87.95
1045	YFV	23.7	90.14	2,11	99.83	88.9	585	91.73
1053	YFV	26.6	91.50	3,37	99.85	89.0	594	93.92
1055	YFV	26.9	4.83	238.17	78.43	91.9	587	31.66
1059	YFV	31.0	5.87	72.50	99.82	88.8	660	88.77
1063	YFV	32.0	4.17	77.05	99.82	88.8	637	91.02
1065	YFV	33.0	5.51	104.05	99.82	88.9	625	91.35
1068	YFV	27.3	2.40	221.91	86.40	91.6	525	44.53
1073	YFV	32.5	11.24	247.57	78.11	90.7	567	35.01

1074	YFV	27.9	3.09	117.94	99.82	88.7	634	91.55
1075	YFV	24.1	21.26	1,57	87.22	91.5	551	65.64
2159	SARS-CoV-2	32.2	96.03	864.03	99.03	92.7	501	87.32
2163	SARS-CoV-2	27.6	94.48	928.92	96.66	92.7	500	87.17
2168	SARS-CoV-2	33.3	96.36	821.77	99.04	92.7	500	89.23
2173	SARS-CoV-2	25.5	94.51	1,34	95.90	93.4	505	84.49
2182	SARS-CoV-2	28.7	97.04	1,52	99.90	93.5	505	94.70
2187	SARS-CoV-2	31.1	95.47	852.58	98.16	93.5	506	84.86
2192	SARS-CoV-2	27.5	96.34	1,45	99.92	93.4	504	95.62
2207	SARS-CoV-2	28.9	96.57	1,53	99.89	93.5	506	97.62
5148	SARS-CoV-2	22.6	97.27	1,57	99.89	92.2	501	98.17
5157	SARS-CoV-2	21.8	97.04	1,57	99.90	92.1	498	96.81

Legenda: ZIKV = vírus Zika;

YFV = virus da febre amarela;

SARS-CoV-2 = síndrome respiratória aguda grave coronavírus 2;

Ct = limite de ciclo

N50: Metade da sequência do genoma é coberta por contigs maiores ou iguais ao tamanho do contig N50. Identidade: Identidade comparada ao genoma referência.

O ensaio também foi realizado em 41 amostras clínicas humanas positivas para YFV por RT-qPCR da epidemia de YFV de 2018 em São Paulo, Brasil. O valor mediano do Ct foi de 27,74, variando de 4,6 a 37 correspondendo a 1,00e00 a 1,50e10 cópias do genoma por mL de plasma [57]. Após a quantificação do produto de PCR e a visualização do gel de agarose E-gel, 21 amostras apresentaram bandas específicas distinguíveis por transiluminação e foram selecionadas para preparação da biblioteca e sequenciamento por nanoporos (Tabela 15). As amostras sequenciadas de YFV (n = 21) tiveram valores medianos de Ct = 25,57, entre 5 e 37 gerados em uma biblioteca de 24 *barcodes*. A porcentagem de *reads* mapeadas variou de 1,71% a 97,47%, com uma profundidade média entre 72,50x a 3370x, e a maioria das amostras apresentou cobertura do genoma em torno de 99,82% sendo a mais baixa 78,11% (Tabela 15) (Figura 22). As regiões do genoma com uma profundidade de cobertura <20x foram representadas com N caracteres.

Para SARS-CoV-2, o ensaio foi realizado em 10 amostras residuais nasofaríngeas positivas por RT-qPCR em abril de 2020 em São Paulo, Brasil. O valor mediano do Ct foi de 26,9, variando de 21,8 a 33,3 correspondendo a 2,40e05 a 1,30e02 cópias do genoma por mL (Tabela 13). A porcentagem de leituras mapeadas variou de 94,51% a 97,27%, com uma

profundidade média entre 821,77x a 1570x e mediana de cobertura do genoma de 98,8%, variando de 95,90% a 99,92% (Tabela 15) (Figura 23).

SMART-9N e Rapid SMART-9N de amostras de cultura isolada de ZIKV e limite de detecção

Para o ZIKV, o RNA isolado titulado foi diluído utilizando diluições seriadas de dez vezes, até 1:1.000.000 correspondendo de 6e07 a 6e00 FFU / mL, e submetido ao SMART-9N (Tabela 16). O teste resultou em uma mediana de cobertura genômica de 99,7% para as diluições testadas com a menor cobertura de 99,02% para a diluição 1:1.000.000. A porcentagem de cobertura do genoma em 20x foi de 90,70% com 6e00 FFU / mL até 99,73% com 6e07 FFU / mL de *input* de material. A profundidade de cobertura média foi de até 10010x e 154,25x com 6e00 FFU / mL de material, compatível com ensaios *single-cells* [69]. A média de leituras mapeadas variou de 56,29% a 0,52%. O N50 mediano foi de 1,7 kb e quando as leituras foram analisadas individualmente, o teste obteve coberturas completas do genoma do ZIKV em uma única *read* (aproximadamente 11 kb leitura mais longa) (Figura 22).

Amostra	<i>Input</i> (FFU/mL)	% <i>Reads</i> mapeadas	Profundidade média	% Cobertura	% Identidad e	N50	<i>read</i> mais longa	%20x cobertura
ZIKV Ref	6x10 <sup>7</sup>	48.27	10,01	99.97	84.9	853	10,07	99.73
ZIKV Ref	6x10⁰	46.14	4,73	99.80	84.7	741	8,05	99.67
ZIKV Ref	6x10⁵	39.74	4,56	99.96	84.4	867	9,91	99.57
ZIKV Ref	6x10⁴	56.29	7,48	99.68	83.9	1,15	7,82	99.18
ZIKV Ref	6x10³	25.55	7,02	99.57	83.5	1,66	10,83	99.00
ZIKV Ref	6x10²	6.51	719.08	99.67	83.4	2,14	6,16	97.54
ZIKV Ref	6x10 <sup>1</sup>	0.52	154.25	99.02	83.1	1,74	9,64	90.70

**Tabela 16** - Resumo do sequenciamento e resultados das estatísticas de alinhamento para a cepa de amostra de referência Zika vírus BeH815744 (ZIKV) usando o método SMART-9N durante o desenvolvimento (n = 1 amostra) de acordo com a *input* de material (FFU / mL).

Legenda: ZIKV = Zika vírus; FFU = focus forming units. *N50:* Metade da sequência do genoma coberta por contigs maiores ou iguais ao tamanho do contig N50.

Longest read: tamanho da leitura mais longa alinhada

Identidade: Identidade das leituras alinhadas em comparação com o genoma referência.

A mesma diluição 1:1.000.000 foi usada para testar a abordagem Rapid SMART-9N. As menores proporções de *reads* mapeadas observadas foram 0,06% e as maiores 86,15%. A maioria das amostras retornou um percentual de 99,87%, com a menor de 87,58% para o teste de diluição de 6e00 FFU / mL. A mediana da porcentagem de cobertura do genoma em 20x foi de 90,73% e o N50 de 2,27kb (Figura 22).

O método foi realizado usando *barcodes* RLB de 1  $\mu$ l e 0,5  $\mu$ l do kit SQK-RPB004 (ONT) com 6e07 de *input* de material. O teste resultou em 99,7% de cobertura genômica para 1 $\mu$ l e 0,5  $\mu$ l, e N50 de 1,84kb e 2,11kb, respectivamente (Tabela 17).

**Tabela 17** - Resumo dos resultados de sequenciamento e estatísticas de alinhamento para a cepa referência Zika vírus BeH815744 (ZIKV) usando o método Rapid SMART-9N durante o desenvolvimento (n = 1 amostra) de acordo com o *input* de material (FFU / mL).

Amostra	<i>Input</i> (FFU/mL)	% <i>Reads</i> mapeadas	Profundid ade média	% Cobertur a	% Identidade	N50	<i>Read</i> mais longa	%20x cobertura
ZIKV Ref 1µl	6x10 <sup>7</sup>	84.51	5,55	99.62	83	1,84	7,68	99.00
ZIKV Ref 0,5µl	6x10 <sup>7</sup>	86.15	1,58	99.69	82.6	2,11	8,43	98.82
ZIKV Ref	6x10⁰	85.11	5,55	99.96	83	2,08	7,20	99.02
ZIKV Ref	6x10⁵	72.10	2,88	99.47	83.1	2,14	7,09	99.02
ZIKV Ref	6x10⁴	49.67	1,87	99.78	83.2	2,50	7,62	99.06
ZIKV Ref	6x10³	9.48	7,42	99.30	83.2	2,34	8,44	99.02
ZIKV Ref	6x10²	0.16	1,76	99.56	83.3	2,10	5,85	94.73
ZIKV Ref	6x10¹	0.06	57	87.58	83.1	2,91	6,01	44.60

Legenda: ZIKV = Zika vírus;

FFU = focus forming units.

*N50:* Metade da sequência do genoma coberta por contigs maiores ou iguais ao tamanho do contig N50.

Longest read: tamanho da leitura mais longa alinhada

Identidade: Identidade das leituras alinhadas em comparação com o genoma referência.



**Figura 22** - Comparação das abordagens multiplex PCR, SMART-9N e Rapid SMART-9N A) Cobertura do genoma do isolado ZIKV mapeado contra o genoma referência quanto à cobertura para a posição de referência do genoma comparando com os ensaios multiplex PCR, SMART-9N e Rapid SMART-9N. B) Limite de detecção dos métodos SMART-9N e Rapid SMART-9N analisando as proporções de *reads* mapeadas contra a sequência viral de referência apropriada em uma faixa de *input* de material (FFU / mL) no gráfico esquerdo e porcentagem do genoma de referência sequenciado a uma profundidade mínima de 20 vezes nos dados gerados em uma faixa de *input* de material (FFU / mL) no gráfico esquerdo e da cobertura; Genome position = posição do genoma; Percentage of mapped reads = porcentagem de *reads* mapeadas; Percentage 20x coverage = porcentagem 20x de cobertura; Sample input = *input* de material.

### SMART-9N e Rapid SMART-9N em amostras clínicas de YFV

O ensaio foi avaliado em amostras clínicas de YFV com Cts variando entre 8 e 33. Neste teste primeiro foram comparados a utilização do tratamento com DNase para depleção do DNA do hospedeiro e a não utilização do tratamento das amostras com DNase. A tabela 18 mostra os resultados apresentados no teste, indicando que o tratamento com DNase é essencial para obter uma melhor cobertura genômica e melhores resultados em amostras clínicas.

**Tabela 18** - Resumo dos resultados de sequenciamento e estatísticas de alinhamento para amostras clínicas de YFV (n=4) usando o método SMART-9N durante o desenvolvimento do método comparando a tratamento prévio com DNase e não tratamento.

Amostras	Valor do Ct	n⁰ <i>reads</i>	% <i>reads</i> mapeadas	Profundidade cobertura	Cobertura do genoma	%identidade	N50
1053	26.6	81	11	0.64	42%	92.6	1,536
1045	23.7	1720	3	4.515	94%	91.8	1,090
1029	17.4	338	8	2.333	70%	92.9	1,161
1030	4.6	9248	100	1,030.83	100%	92.7	1,284
1053 DNase	26.6	3711	31	131,265	100%	92.4	1,696
1045 DNase	23.7	2410	34	93.546	100%	92.5	1,691
1029 DNase	17.4	8329	92	990.64	100%	92.4	1,456
1030 DNase	4.6	14600	100	1,651.54	100%	92.5	1,257

#### Ct: Cycle threshold;

*N50:* Metade da sequência do genoma coberta por contigs maiores ou iguais ao tamanho do contig N50.

Identidade: Identidade das leituras alinhadas em comparação com o genoma referência.

Depois de validar os métodos no isolado de ZIKV, aplicamos os ensaios validados em amostras clínicas. Começando com o SMART-9N, sete amostras clínicas humanas representativas positivas para RNA de YFV, já sequenciadas com o método multiplex PCR, com valores de Ct entre 4,6 e 33 foram selecionadas (Tabela 19).

**Tabela 19** - Resumo dos resultados de sequenciamento e estatísticas de alinhamento para amostras de plasma do vírus da febre amarela (YFV) (n = 7) usando o protocolo SMART-9N durante a validação do método de acordo com os valores de Ct.

Amostr a	Valor do Ct	n⁰ <i>read</i> s	% reads mapeada S	Profundidad e cobertura	Cobertura do genoma	% Identidade	N50	<i>Read</i> mais longa	%20x cobertur a
1019	17.2	30,05	32.2	630.1	99.91	91.6	998	3,92	98.84
1020	29.5	53,8	2.0	40.6	99.91	92.9	659	4,46	64.28
1029	17.4	16,83	85.7	1,87	99.99	92.3	1,58	7,95	99.91
1030	4.6	31,2	98.6	3,48	99.93	92.4	1,36	7,72	99.91
1045	23.7	981	21.4	240.0	96.96	92.3	1,39	6,94	63.81
1053	26.6	1,74	20.4	40.4	99.91	92.2	1,4	10,08	80.50
1065	33.0	15,82	0.3	3.2	95.11	90.1	1,01	2,57	0.00

Legenda: Ct = cycle threshold;

N50: Metade da sequência do genoma é coberta por contigs maiores ou iguais ao tamanho do contig N50.

Leitura mais longa: tamanho da leitura mais longa alinhada nos dados.

Identidade: Identidade das leituras alinhadas em comparação com a referência do genoma.

Um total de 86% das amostras apresentaram cobertura do genoma superior a 99,9% variando de 95,11% a 99,99% com valores Ct de 33 e 18 respectivamente, e uma profundidade média mínima de 3,2x, e máxima de 3480x (Figura 24 - A). As mesmas amostras foram selecionadas e submetidas ao método Rapid SMART-9N (Tabela 20).

**Tabela 20** - Resumo dos resultados de sequenciamento e estatísticas de alinhamento para amostras de plasma do vírus da febre amarela (YFV) (n = 7) usando o protocolo Rapid SMART-9N durante a validação do método de acordo com os valores de Ct.

Amostr a	Valor nº do Ct <i>reads</i>	% <i>reads</i> mapeada s	Profundidad e cobertura	Cobertura do genoma	% Identidade	N50	<i>Read</i> mais longa	%20x cobertur a
-------------	-----------------------------------	-----------------------------------	-------------------------------	---------------------------	-----------------	-----	------------------------------	-----------------------

1019	17.2	368,59	0.44	66.50	99.91	89.9	705	6,7	95.58
1020	29.5	441,13	0.05	26.35	99.10	89.4	2,05	6,15	81.68
1029	17.4	9,21	38.18	175.87	99.91	90.2	792	9,12	98.95
1030	4.6	212,07	98.26	25,3	99.95	89.8	1,8	3,96	99.91
1045	23.7	29,97	5.13	21.44	99.91	91.0	2,1	6,16	99.04
1053	26.6	1,005, 93	0.31	323.37	99.92	91.3	1,61	6,86	99.74
1065	33.0	109,77 4	0.16	22.81	94.28	91.8	2,16	7,21	50.50

Legenda: Ct = cycle threshold;

N50: Metade da sequência do genoma é coberta por contigs maiores ou iguais ao tamanho do contig N50.

Leitura mais longa: tamanho da leitura mais longa alinhada nos dados.

Identidade: Identidade das leituras alinhadas em comparação com a referência do genoma.

As maiores porcentagens de *reads* mapeadas observadas foram 98,26% e 38,18% para os valores de Cts 4,6 e 17,4, respectivamente. Um total de 86% das amostras apresentaram cobertura do genoma superior a 99,9% com a menor de 94,28% com Ct de 33, e a profundidade média de cobertura variou de 21,44x a 2530x (Figura 24 - A). Comparamos a profundidade de cobertura com diferentes amostras de valores de Cts em todo o genoma para cada método (PCR multiplex, SMART-9N e Rapid SMART-9N) (Figura 23). A profundidade média de cobertura revelou maior profundidade do genoma e melhor padrão de cobertura em todo o genoma para os métodos metagenômicos quando comparados ao método de multiplex PCR.



**Figura 23** - Comparação da profundidade de cobertura do genoma através do genoma do vírus da febre amarela (YFV) para diferentes métodos (Multiplex PCR, SMART-9N e Rapid SMART-9N) em todas as amostras clínicas testadas com diferentes valores de Cts (n = 7). Legenda: Coverage depth = profundidade da cobertura; Genome position = posição do genoma; Ct - value = Valor do Ct.

Todas as sete amostras sequenciadas com ambos os métodos foram comparadas com os resultados do ensaio multiplex PCR. Apesar da diminuição na proporção de *reads* virais mapeadas em toda a faixa de valores de Ct (Figura 24 - B) com o SMART-9N e o Rapid SMART-9N, obtivemos uma correlação comparável (SMART-9N R = 0,91, p = 0,005; Rapid SMART-Metagenomics R = -0,86, p = 0,012,). A correlação mostrou uma proporção menor de *reads* virais conforme o aumento dos valores dos Cts, com um nível considerável de variação (0,3% a 98,6% com SMART-9N e 0,16% a 98,26% com método rápido SMART-9N) entre as amostras e os métodos. (Tabelas 19, 20)

A cobertura genômica de 20 vezes foi analisada através dos valores de Ct entre todos os métodos, apresentada na Figura 24 - B. Na abordagem multiplex PCR, a média da cobertura do genoma foi de 78,9% com um mínimo de 35,01% para Ct 33 em comparação com 71,5% e 89,3% para SMART-9N e Rapid SMART-9N com um mínimo de 0% e 50,5%, respectivamente (Tabelas 19, 20).

Para este subconjunto de amostras, também comparamos os resultados do valor N50 das abordagens para cada amostra (Figura 24 - C). Aqui, descobrimos que a faixa foi de 525 bp a 660 bp para o ensaio multiplex PCR, 659 bp a 1,58 kbp para o SMART-9N, 705 bp a 2,16 bp para o Rapid SMART-9N. A mediana foi 599,8 pb, 1,6 kbp e 1,2 kbp para o multiplex PCR, SMART-9N e Rapid SMART-9N, respectivamente (Tabelas 19, 20). Para as amostras clínicas de YFV, as leituras mais longas observadas foram 10,08kb e 9,12kb para o SMART-9N e Rapid SMART-9N, respectivamente, 93,33% e 84,44% de todo o genoma viral.



Figura 24 - Comparação dos resultados dos ensaios multiplex PCR, SMART-9N e Rapid SMART-9N para amostras clínicas de YFV. A) Profundidade média da cobertura do genoma e 95% CI das *reads* mapeadas para a

posição de referência do genoma de cada amostra. B) Proporção de mapeamento das *reads* contra o genoma de referência em uma gama de valores de Ct (esquerda) e porcentagem de sequenciamento mapeados ao genoma de referência a uma profundidade mínima de 20 vezes em uma gama de valores de Ct (direita). C) N50 de cada amostra em bp. (n = 7 amostras). Legenda: Coverage depth = profundidade da cobertura; Genome position = posição do genoma; Percentage of mapped reads = porcentagem de *reads* mapeadas; Percentage 20x coverage = porcentagem 20x de cobertura.

Amostras clínicas SARS-CoV-2 para o ensaio Rapid SMART-9N

As amostras clínicas de SARS-CoV-2 foram submetidas à abordagem Rapid SMART-9N. Quando mapeadas as *reads* geradas durante validação do método ao genoma do vírus de referência (isolado: Wuhan-Hu-1, GenBank No. de acesso MN908947), genoma viral de SARS-CoV-2 estava presente em todas as dez sete amostras até o valor Ct 34 (*reads* totais variaram de 6480 a 93.570). As amostras sequenciadas foram comparadas com os resultados do ensaio multiplex PCR e não mostraram uma correlação significativa (R = 0,49, p = 0,15) entre a proporção de *reads* virais com o aumento do valor Ct (12,15 - 98,22%) (Tabela 21).

**Tabela 21** - Resumo dos resultados de sequenciamento e estatísticas do alinhamento para amostras clínicas de síndrome respiratória aguda grave de coronavírus 2 (SARS-CoV-2) (n = 10) usando o protocolo Rapid SMART-9N durante a validação do método de acordo com os valores de Ct.

Amostra	Valo r do Ct	nº reads	% <i>reads</i> mapeada s	Profundidad e cobertura	Cobertur a do genoma	% Identidade	N50	<i>Read</i> mais Ionga	%20x cobertur a
2159	32.2	187,35	68.30	556.96	100	91.9	2,42	6,50	99.80
2163	27.6	127,93	97.87	368.68	100	91.9	2,38	8,50	99.84
2168	33.3	82,53	66.01	281.04	100	91.9	2,54	7,21	99.79
2173	25.5	117,53	52.10	97.51	100	91.5	1,13	5,05	99.57
2182	28.7	142,95	64.40	333.64	100	92.0	2,38	8,77	99.80
2187	31.1	64,77	44.54	109.53	100	91.7	2,56	8,09	99.80
2192	27.5	130,99	48.16	98.49	100	91.3	1,09	5,92	99.81
2207	28.9	191,79	98.22	628.12	100	91.4	2,04	8,65	99.87
5148	22.6	842,57	12.15	291.97	100	91.7	1,10	6,30	99.81
5157	21.8	935,75	41.75	1,28	100	91.0	1,24	18,48	99.89

Legenda: Ct = cycle threshold;

N50: Metade da sequência do genoma é coberta por contigs maiores ou iguais ao tamanho do contig N50.

Leitura mais longa: tamanho da leitura mais longa alinhada nos dados.

Identidade: Identidade das leituras alinhadas em comparação com a referência do genoma.

A cobertura do genoma foi de 100% em todas as 7 amostras e a profundidade de cobertura mais baixa de 97,51x (Figura 25-A). Ao comparar cada profundidade de cobertura em diferentes amostras com uma variedade de valores de Ct para o PCR multiplex e métodos Rapid SMART-9n (Figura 26), pudemos observar uma profundidade de cobertura concordante e um padrão de cobertura homogénea em todo o genoma para ambos os métodos.

A comparação da cobertura do genoma em 20 vezes entre os métodos multiplex PCR e Rapid SMART-9N em toda a faixa de título viral é mostrada na Figura 24-B. A mediana revelada foi de 91,59%, mínimo de 84,49%, e o Rapid SMART-9N 99,79%, mínimo de 99,57%. Uma comparação do N50 em todas as 10 amostras foi feita resultando em um N50 mais alto de todas as amostras para a abordagem Rapid SMART-9N, até 2,56 kb. A *read* mais longa foi de 18,48 kb, a *read* mais longa obtida neste estudo, compreendendo aproximadamente 62% do genoma SARS-CoV-2 (29.903 bp) (Figura 25-C).



Figura 25 - Comparação dos resultados dos ensaios multiplex PCR e Rapid SMART-9N para amostras clínicas de SARS-CoV-2. A) Profundidade média da cobertura do genoma e 95% CI das *reads* mapeadas para a posição

de referência do genoma de cada amostra. B) Proporção de mapeamento das *reads* contra o genoma de referência em uma gama de valores de Ct (esquerda) e porcentagem de sequenciamento mapeados ao genoma de referência a uma profundidade mínima de 20 vezes em uma gama de valores de Ct (direita). C) N50 de cada amostra em bp. (n = 10 amostras). Legenda: Coverage depth = profundidade da cobertura; Genome position = posição do genoma; Percentage of mapped reads = porcentagem de *reads* mapeadas; Percentage 20x coverage = porcentagem 20x de cobertura.



**Figura 26** - Comparação da profundidade de cobertura do genoma do vírus da síndrome respiratória aguda grave coronavírus 2 (SARS-CoV-2) para diferentes métodos (Multiplex PCR e Rapid SMART-9N) em todas as amostras clínicas testadas com diferentes valores de Cts (n = 10). Legenda: Coverage depth = profundidade da cobertura; Genome position = posição do genoma; Ct - value = Valor do Ct.

Detecção de outros vírus de RNA em amostras clínicas e classificação de Kraken

Para testar a capacidade dos nossos métodos de detectar outros vírus nas amostras clínicas, avaliamos a classificação taxonômica de leituras usando Kraken. Isso permitiu a identificação de um vírus dsDNA do gênero Pa6virus, família Siphoviridae, presente em uma amostra de YFV. Após a identificação, as leituras foram mapeadas para a sequência de referência (NC 018838.1) obtendo 197 leituras do vírus, 84,78% de cobertura do genoma

com uma profundidade de cobertura máxima de 32x e identidade de 81,4%. A sequência consenso foi gerada e bases com uma profundidade de <10 vezes foram representadas com caracteres N (github.com/CADDE-CENTRE/Rapid-RNA-SMART-Metagenomics). A proporção de leituras não classificadas, de eucariotos, bactérias, arquéias e vírus para cada amostra pode ser encontrada na Tabela 22.

**Tabela 22 -** Proporção da porcentagem de *reads* não classificadas, eucariotas, bactérias, archaea e vírus, para cada amostra de acordo com a distribuição da classificação Kraken e metodologias metagenômicas.

Amostra	Tipo do vírus	Metodologia	nº <i>reads</i>	Não classificadas (%)	Eukaryot a (%)	Bacteri a (%)	Archae a (%)	Vírus (%)
1019	YFV	SMART-9N	30,05	58.12	37.63	1.37	0.01	2.18
1020	YFV	SMART-9N	53,8	46.17	45.23	5.08	0.01	0.14
1029	YFV	SMART-9N	16,83	72.75	11.03	1.97	0.02	13.87
1030	YFV	SMART-9N	31,2	84.46	2.40	1.91	0.02	11.17
1045	YFV	SMART-9N	981	30.48	35.27	4.79	0	29.26
1053	YFV	SMART-9N	1,74	28.90	42.31	1.15	0.06	27.00
1065	YFV	SMART-9N	15,82	46.91	52.23	0.56	0.01	0.01
1019	YFV	Rapid SMART-9N	368,59	83.10	1.70	16.85	0.01	0.04
1020	YFV	Rapid SMART-9N	441,13	78.60	18.85	2.26	0.01	0.04
1029	YFV	Rapid SMART-9N	9,21	84.76	12.74	1.55	0	0.73
1030	YFV	Rapid SMART-9N	212,07	92.86	0.54	2.33	0.02	4.79
1045	YFV	Rapid SMART-9N	29,97	80.47	18.86	0.84	0.01	0.65
1053	YFV	Rapid SMART-9N	1,005,9	77.88	21.15	0.73	0.01	0.03
1065	YFV	Rapid SMART-9N	109,77	75.42	23.77	0.41	0.01	0.01
2159	SARS-CoV- 2	Rapid SMART-9N	187,35	58.72	2.84	20.35	0.04	17.91
2163	SARS-CoV- 2	Rapid SMART-9N	127,93	62.76	3.69	3.22	0.04	30.13
2168	SARS-CoV- 2	Rapid SMART-9N	82,53	63.46	6.59	14.42	0.02	15.34
2173	SARS-CoV- 2	Rapid SMART-9N	117,53	74.84	17.37	1.68	0	5.90

2182	SARS-CoV- 2	Rapid SMART-9N	142,95	56.45	6.00	17.96	0.04	19.37
2187	SARS-CoV- 2	Rapid SMART-9N	64,77	61.90	12.43	12.01	0	13.39
2192	SARS-CoV- 2	Rapid SMART-9N	130,99	74.93	18.21	1.51	0.03	5.24
2207	SARS-CoV- 2	Rapid SMART-9N	191,79	74.11	2.25	2.10	0.04	21.43
5148	SARS-CoV- 2	Rapid SMART-9N	842,57	75.46	1.43	22.16	0.02	0.88
5157	SARS-CoV- 2	Rapid SMART-9N	935,75	75.31	1.70	19.37	0.01	3.48

Legenda: YFV = vírus da febre amarela SARS-CoV-2 = síndrome respiratória aguda grave coronavírus 2

## **3.4 Capítulo 4** - Otimização do custo do sequenciamento MinION.

Os dados gerados do sequenciamento a base de nanoporos devem ser convertidos do sinal de corrente elétrica (FAST5) em bases nucleotídicas (FASTQ), um processo denominado de *basecalling*. Este processo ocorre em servidores locais ou laptops em que tem um custo computacional significativo. Para a realização de basecalling em tempo real, ou seja, à medida que os dados vão sendo gerados, é necessário um computador com GPU e placa de NVIDIA. Caso contrário, esse processo demora em torno de 1-2 dias. No entanto, em muitos casos, muitas *reads* geradas não são necessárias, já que alguns vírus já foram sequenciados suficientemente, ou as reads não vêm de genomas virais (off target). Para resolver isso, empregamos técnicas de 'read until' denominada RAMPART para normalizar os amplicons até um nível fixo de cobertura (isto é, 100 vezes) e remover qualquer amplificação "off target", de modo que as análises possam prosseguir mais rapidamente, e sem um custo computacional tão grande. Ao utilizar um conjunto de dados de tamanho reduzido, o processo de base é executado no menor conjunto de dados necessário para dar resultados corretos, permitindo o direcionamento da capacidade de sequenciamento para moléculas específicas e, assim, em teoria, otimizar a capacidade de sequenciamento de reads desejadas. A equipe formada por pesquisadores da Universidade de Birmingham foram os primeiros a demonstrar que esta técnica funciona e demonstrou a capacidade de assegurar uma cobertura uniforme em todos os *amplicons* - um processo chamado de "balanceamento" (Figura 27) [70].



**Figura 27** - Exemplo de "balanceamento de *barcodes*" em tempo real - acima: distribuição dos *amplicon* antes do Rampart, embaixo: distribuição de *amplicon* após Rampart. Legenda: Coverage depth = profundidade da cobertura; Genome position = posição do genoma; Read type = tipo da *read*; Library control = biblioteca controle; Normalised = normalizada.

Para esta aplicação o Rampart tem os benefícios de:

1) Normalizar ("equilibrar") amplicons para um nível fixo de cobertura (isto é, 100 vezes);

2) Aplicar estes mesmos algoritmos a amostras com *barcodes* para assegurar uma cobertura uniforme tanto por *amplicons* como por barcodes e

3) rejeição de amplicons "off target".

O RAMPART é executado simultaneamente com o sequenciamento MinION seguindo as instruções disponiveis em <u>https://github.com/artic-network/rampart</u> e funciona seguindo os passos seguintes em tempo real (Figura 28):

- MinKNOW & guppy (via dogfish) produzem fastqs;
- Rampart demultiplexa os fastqs usando porechop e extrai os barcodes;
- Mapeia duas vezes, uma vez para uma única sequência de referência para obter coordenadas, uma contra um painel de referências para encontrar a correspondência mais próxima.

ATTIC NOTING	RAMPART Read Assignment, Map LHFV Current Outbreak Simulation Run time: 03m225 998 reads. 1 fastes. 6 samples	ping, and Phylogenetic Analysis in F	iteal Time	
	Read Depth	Total reads over time	Reads per Sample	Betweenie Masches           944<
1 - OB_C	204 reads. 0% > 1000s. 19% > 100s,	76% > 10x coverage.		dara separa
2 - case260	) 211 reads. 0% > 1000x, 2% > 100x, 8	9% = 30e coverage.		
3 - case321	181 reads. 0% > 1000x, 0% > 100x, 8	9% > 50x coverage.		
4 - contamir	nant 90 reads. 0% > 1000x. 0% > 100x. 60	% > 10x coverage.		
5 - case352	114 reads. 0% > 1000x, 0% > 100x, 8	2% > 10x coverage.		
6 - case370	198 reads. 0% > 1000x. 0% > 100x. 8	9% > 10x coverage.		

Figura 28 - Exemplo da execução do RAMPART para análise em tempo real.

À medida que o sequenciamento é realizado, os dados gerados são mapeados, podemos detectar o patógeno e interromper a execução assim que obtivermos a cobertura desejada do genoma alvo e, assim, excluir todas as sequências não virais, para que possamos realizar a análise bioinformática em computadores comuns e de maneira mais simples e rápida.

A utilização do RAMPART para detecção inicial e sequenciamento significa que ao gerar os dados suficientes e parar a corrida, não será utilizada totalmente a capacidade de sequenciamento da *flow cell* da ONT. Portanto, durante a validação dos métodos descritos anteriormente, testamos a reutilização das *flow cells* após um tratamento com tampão de lavagem contendo DNAse I (Tabela 23).

## Protocolo de digestão:

Antes do próximo sequenciamento, prepare o seguinte tampão de digestão:

Componente	Volume (µl)
Running buffer (RBF)	100
Nuclease free water	100

Tabela 23 - Realização do protocolo de digestão

CaCl2 (0.1M – dilute 1M stock 1/10)	4
DNase I (M0303L)	4
Volume total	208

Remova os resíduos da corrida de sequenciamento anterior da *flow cell* com as portas fechadas. Remova o ar pipetando da porta de entrada. Adicione 200  $\mu$ l de RBF diluído (proporção de 576 RBF + 672 NFW) à porta de entrada e, em seguida, 150  $\mu$ l do tampão de digestão. Aguarde 30-60 minutos. Antes de carregar a nova biblioteca, inicie um novo experimento para se certificar de que não há vestígios da biblioteca anterior. Verifique as leituras e os poros ativos (normalmente você pode ver um número alto de poros ativos, mas verifique se estes são apenas os adaptadores). É extremamente importante não introduzir nenhuma bolha de ar para dentro da *flow cell* durante esse processo. Prepare a *flow cell* de acordo com as instruções do fabricante e carregue a nova biblioteca normalmente.

A reutilização das *flow cells* diminuiu drasticamente o custo do sequenciamento chegando a menos de US \$10 por amostra. Com o RAMPART, criamos um painel contendo os arbovírus de maior importância para a saúde pública (painel disponível: <a href="https://www.dropbox.com/sh/27jp5s4ynko9i0m/AAC5VTNdM2nhqimXWRbYV2cJa?dl=0">https://www.dropbox.com/sh/27jp5s4ynko9i0m/AAC5VTNdM2nhqimXWRbYV2cJa?dl=0</a>., e estabelecemos uma cobertura genômica de 100x. Durante os experimentos, foi possível reutilizar as *flow cells* até 7 vezes em 3 horas, com detecção das primeiras *reads* virais nos primeiros 5 minutos do sequenciamento.

# 4 DISCUSSÃO

A vigilância genômica em tempo real se tornou uma abordagem para monitoramento da introdução de doenças infecciosas emergentes e de vírus circulantes. Protocolos rápidos e acessíveis e o posterior compartilhamento de dados podem fornecer informações genômicas, permitindo que as sequências sejam comparadas entre laboratórios ou autoridades de saúde pública, gerando assim, informações sobre as rotas de transmissão e a eficácia das medidas de contenção, podendo assim preparar para ou até mesmo prevenir novos surtos [80].

Uma variedade de abordagens diferentes têm sido usadas para enriquecimento viral e sequenciamento genômico, incluindo metagenômica de RNA, tecnologia usada para gerar o primeiro genoma de referência de SARS-CoV-2 [75], captura por hibridização, enriquecimento do alvo baseado em PCR, entre outros. O enriquecimento do alvo baseado em PCR produz uma alta taxa dos produtos alvo com um fluxo de trabalho laboratorial simples e barato quando comparado a outros métodos disponíveis. Antes do SARS-CoV-2, o surto do vírus Ebola na África Ocidental de 2013-2016 foi o surto viral mais densamente amostrado, com mais genomas gerados usando sequenciamento de amplicons do que todos os outros métodos combinados [39, 81]. Em 2016, o projeto ZiBRA (Zika no Brasil Real Time Analysis) usou essa abordagem para a vigilância de ZIKV a partir de um método adaptado de enriquecimento do alvo baseado em PCR denominado multiplex PCR [40]. Combinamos essa metodologia com a tecnologia de sequenciamento portátil baseada em nanopores MinION (ONT) para obtenção de genomas completos do vírus Zika em tempo real - chamada "epidemiologia genômica". Desde então esse protocolo vem sendo utilizado amplamente para a vigilância genômica de diversos vírus de importância pública como YFV [25], CHIKV [23], DENV [46, 61] e recentemente SARS-CoV-2 [82, 83].

No entanto, mesmo sendo a tecnologia de sequenciamento mais barata, acessível e mais adequada para a vigilância genômica de vírus nos dias de hoje, o custo por amostra do sequenciamento ainda é alto, em torno de US \$50, devido aos reagentes empregados. Além disso, tais abordagens são adequadas apenas para detectar surtos virais já conhecidos, sendo necessário diagnóstico molecular prévio para que sejam utilizados os esquemas de primers de Multiplex PCR já disponíveis e validados para cada vírus separadamente, não permitindo uma rápida detecção de novos vírus e rastreio dos casos com resultados negativos nos testes

moleculares mas com sintomatologia de infecções virais. Também, há uma complexidade significativa associada à operação do instrumento, criação de bibliotecas de sequenciamento e análise de dados e o tempo para realização do protocolo ainda é muito maior quando comparado a outros métodos moleculares, o que ainda torna improvável a implementação e metodologia de escolha nos centros de diagnóstico e vigilância epidemiológica do Brasil.

Neste trabalho buscamos gerar novas metodologias a partir das dificuldades encontradas durante o período de implementação de protocolos de epidemiologia genômica em estudos a partir do surto de ZIKV em 2016, até os dias de hoje com SARS-CoV-2. Esses protocolos foram desenvolvidos para a vigilância completa de vírus de importância pública, incluindo um painel de detecção e sequenciamento do genoma completo dos principais arbovírus de importância pública, um protocolo rápido de detecção e sequenciamento a partir de alvos específicos e duas abordagens de sequenciamento por metagenômica para rastreio e potencial descoberta de novos vírus. Além disso, introduzimos a análise em tempo real dos dados e reutilização dos chips - *flow cells* de sequenciamento para rápida detecção e diagnóstico e redução dos custos dos métodos já disponíveis.

Começamos com o desenvolvimento de um esquema de *primers* para detecção, amplificação e sequenciamento do genoma completo de múltiplos vírus em um único ensaio -*Superpool.* A partir do ensaio molecular multiplex PCR já validado e amplamente utilizado para sequenciamento de regiões alvos baseado em PCR [40], desenvolvemos um painel contendo 14 alvos virais, incluindo os vírus Zika, Chikungunya, Febre Amarela, Mayaro, Oropouche, vírus do Nilo Ocidental (WNV), Dengue 1-4 (DENV1, DENV2, DENV3, DENV4), Rocio (ROCV), Ilhéus (ILHV), Tacaiuma (TCV) e Saint Louis (STLV), para auxiliar no diagnóstico e vigilância genômica de infecções arbovirais. A concepção de ensaios multiplex PCR com um grande número de pares de *primers* é dificultado devido ao aumento da probabilidade de interações entre eles, que reduzem a amplificação e eficiência dos testes, podendo levar a falhas pelos genomas e consequentemente inferências filogenéticas. Testamos o efeito de diferentes esquemas de primers para todos os vírus estudados e avaliamos sua capacidade de amplificar genomas completos de arbovírus presentes individualmente e em misturas.

Os esquemas dos vírus ZIKV, YFV, CHIKV e DENV 2 já tinham sido gerados e validados em estudos anteriores [16, 23, 25, 46]. Geramos então os esquemas para os vírus Mayaro, Oropouche, vírus do Nilo Ocidental (WNV), Dengue 1, 3 e 4 (DENV1, DENV3,

DENV4), Rocio (ROCV), Ilhéus (ILHV), Tacaiuma (TCV) e Saint Louis (STLV) a partir da ferramenta *Primal Scheme*, com tamanho de amplicons variando entre 400-1000pb (referencias usadas e esquemas gerados disponíveis em <u>https://www.dropbox.com/sh/bgubs2pkvfba3t8/AAApFzahOnTPibQlgdV4geSda?dl=0</u>). Os esquemas individuais gerados foram validados para cada vírus individualmente, e em seguida foram testados combinando todos os esquemas em um único ensaio totalizando cerca de 1000 pares de *primers* com amostras clínicas e amostras de isolados virais titulados.

Os resultados demonstraram que os ensaios multiplex PCR com grande número de primers podem gerar uma cobertura genômica semelhante ao teste apenas com os primers dos arbovírus presentes individualmente, mesmo quando estamos lidando com vírus geneticamente semelhantes. Isso significa que o ensaio não aumentou a probabilidade de interações entre os primers e a eficiência de amplificação. Portanto, aqui apresentamos um painel de arbovírus que pode ser usado para identificar e consequentemente sequenciar o genoma completo dos principais arbovírus em circulação no Brasil. Este painel pode ser modificado de acordo com a demanda necessária, sendo possível a adição de novos esquemas de multiplex PCR virais, ou a validação de novos painéis para vírus encefalíticos, vírus respiratórios por exemplo.

Junto com o desenvolvimento do método Superpool, utilizamos durante os sequenciamentos a ferramenta apresentada no capítulo 4 desta tese - RAMPART (https://artic-network.github.io/rampart/). A partir de genomas de referência de cada vírus adicionado no método, fizemos um painel (https://www.dropbox.com/s/ntj8hs06blo7216/arbovirus reference genomes.fasta?dl=0) em que durante o sequenciamento, as reads geradas no formato FASTQ eram mapeadas em tempo real nos genomas referências e assim, à medida que os sequenciamentos geravam 200x de cobertura (cobertura suficiente para análises genômicas e filogenéticas), a corrida era interrompida. No momento em que este estudo foi realizado, ainda não havia o kit para a digestão e lavagem das flow cells, disponíveis hoje pela Oxford Nanopore Technologies -Flow Cell Wash Kit EXP-WSH004. Validamos então o kit para lavagem das flow cells também apresentado no capítulo 4, permitindo a reutilização de cada chip de acordo com a sua capacidade de utilização, diminuindo então drasticamente o custo do sequenciamento de acordo com o número de re-utilizações. Como teste final, combinando a utilização do ensaio Superpool, RAMPART e reutilização das *flow cells*, foi possível realizar o sequenciamento de

mais de 100 genomas em 3,5 horas e reutilizar apenas uma *flow cell* até 7 vezes.

Para o painel, ainda está sendo testado o limite de detecção do ensaio, diluindo todas as amostras de RNA em um exponencial de até 10-1 e sequenciando as amostras para analisar a cobertura final, podendo assim comparar os resultados em amostras clínicas com baixas caixas virais. Este método é mais adequado para uso em uma situação de surto onde os isolados são altamente relacionados e o baixo custo por amostra e tempo de resposta rápido são necessários.

Em seguida, desenvolvemos uma abordagem rápida utilizando a fusão de *barcodes* para amplificação do genoma completo de vírus num único ensaio denominado *SMART-plex*. No método anterior vimos que era possível a realização de um ensaio com um grande número de *primers* multiplexados. Portanto, buscamos desenvolver um método universal e menos complexo para preparação de bibliotecas e sequenciamento pela metodologia MinION (ONT, UK).

SMART (*Switching mechanism at 5' end of RNA templates*) é um método popular para RNA-Seq, que depende de *primers* oligo-dT para anelamento e amplificação de moléculas de mRNA poliadeniladas. Nós desenvolvemos uma abordagem denominada SMART-plex utilizando apenas *primers* reversos gene específicos de cada vírus gerados no capítulo 1 para anelamento inicial no RNA alvo, seguido da utilização de *primers strand switching* (SSP) para geração do cDNA *double-strand* marcado (*tagged*). A amplificação é então realizada com uma versão compatível, complementar às *tags* adicionadas durante a geração do cDNA, de *primers* de PCR já contendo os *barcodes* disponíveis na Oxford Nanopore Technologies. O preparo da biblioteca baseia-se apenas na ligação dos adaptadores de sequenciamento que são adicionados rapidamente via *click chemistry*.

O ensaio SMART-plex foi comparado com o Multiplex PCR em amostras de isolados virais de ZIKV e amostras clínicas de YFV, onde foram avaliados: cobertura genômica, sensibilidade, profundidade, custo, tempo e complexidade dos testes. O PCR Multiplex embora seja o método de escolha no momento, é um método demorado (>10 horas totais). A ligação dos *barcodes* durante o ensaio PCR convencional permite que o preparo da biblioteca de sequenciamento que antes era de 6 horas, passe a ser 10 minutos, sendo necessário apenas a ligação dos adaptadores, via *click chemistry*, permitindo uma rápida detecção, caracterização e sequenciamento do genoma completo dos vírus alvos. O tempo do fluxo de trabalho

laboratorial caiu de 14 horas totais para 5 horas e 40 minutos (diminuição de 62%), com cobertura genômica de 100% em 5 das 6 amostras testadas, demonstrando que o teste SMART-plex é uma metodologia que poderá ser utilizada para sequenciamento de arbovírus após finalização dos testes e validação. Também, a partir da ligação dos adaptadores via *click chemistry*, o ensaio reduziu a quantidade de enzima de ligação usada, reduziu o tempo e a complexidade da biblioteca etapa de preparação e reduziu significativamente os custos do ensaio, permitindo o uso de até 96 barcodes por ensaio utilizando o kit recém lançado *Rapid PCR barcoding kit* (SQK-RPB004) chegando a um custo de até US\$ 10 por amostra, sendo possível também a utilização do software RAMPART para otimização da utilização das *flow cells* e reutilização. Com otimização, será possível realizar este ensaio em três horas em único tubo e possibilitar a caracterização completa do genoma de diversos vírus utilizados no painel em um único ensaio.

Por último, desenvolvemos os ensaios SMART-9N e Rapid SMART-9N para rastreio de vírus de RNA e potencial descoberta de novos vírus.

Desenvolvemos aqui duas abordagens metagenômicas virais, SMART-9N e Rapid SMART-9N como métodos metagenômicos não direcionados para detecção e caracterização de RNA viral. As duas técnicas demonstraram excelente especificidade (100%) quando testadas em amostras isoladas e clínicas e comparadas a um método padrão ouro de amplificação alvo multiplex PCR [5-9], que permitiram a identificação de uma co-infecção desconhecida em uma amostra clínica de YFV. Para amostras de isolados de ZIKV, foi possível obter 99,02% da cobertura do genoma com um input de 6.00 FFU / mL, uma quantidade comparável a outros métodos single-cell disponíveis [76, 77] para a abordagem SMART-9N. Para o ensaio Rapid SMART-9N, 87,58% do genoma do ZIKV foi sequenciado para a mesma diluição de 1: 1.000.000, equivalente a 6.00 FFU / mL. A alta sensibilidade e alto rendimento de sequências virais provenientes de amostras clínicas de YFV e SARS-CoV-2 revelados nesse estudo, nos mostrou ser potencialmente viável a realização de sequenciamento por metagenômica de genomas completos utilizando a tecnologia MinION, mesmo para valores de Ct mais elevados. Amostras clínicas representativas com valores de Ct entre 4,6 e 33 para YFV e entre 21,8 e 33,3 para SARS-CoV-2 foram selecionadas para testar a recuperação do genoma dos vírus testados. Notavelmente, os métodos SMART-9N e Rapid SMART-9N foram eficazes no sequenciamento direto do genoma de amostras clínicas para
ambos os vírus testados, uma vez que as leituras virais foram detectadas em todas as amostras, até em amostras com 1,00e00 cópias do genoma por mL.

Avaliando o comprimento das *reads* durante a validação, observamos que nossa abordagem gerou *reads* muito longas quando comparada a outras abordagens metagenômicas [62, 63]. Neste estudo, geramos o genoma completo de ZIKV e YFV e aproximadamente 60% do genoma SARS-CoV-2 em uma única *read*. O N50 dos métodos foi de até 2,91kb com as amostras isoladas e até 2,56kb com amostras clínicas SARS-CoV-2. Ao observar a profundidade média de cobertura das amostras e o CI entre as amostras e métodos estudados, pudemos observar uma amplificação consistente em todos os genomas devido a este alto N50 gerado pelos métodos metagenômicos. O N50 aumentado fornece maior confiança na atribuição taxonômica de *reads* individuais, melhorando a confiança no mapeamento, montagem de novo e a capacidade de detectar recombinações virais [78,79]. Pelo que sabemos, 18,5 kb é o cDNA viral mais longo publicado até o momento produzido pelo método Rapid SMART-9N, provavelmente devido ao fato de que a polimerase LongAmp é usada para a amplificação dos *primers* já contendo os *barcodes*, de acordo com as recomendações da ONT, enquanto a polimerase Q5 foi usada para SMART-9N .

Também comparamos a complexidade, os custos e o tempo necessário de trabalho laboratorial com a abordagem já padronizada multiplex PCR [40]. Assim como o ensaio SMART-plex, o ensaio Rapid SMART-9N reduziu a complexidade, o tempo e o custo da amostra até o sequenciamento. A possibilidade da adição de *barcodes* durante a reação de PCR, reduziu a probabilidade de erro manual e o custo devido à não necessidade de enzimas *end-prep* e de ligação, que também dependem de congelamento e descongelamento (*cold-chain*), sendo inconvenientes para uso no campo. O tempo total de trabalho laboratorial caiu 15% e 61% para o SMART-9N e Rapid SMART-9N, respectivamente, quando comparado ao ensaio multiplex PCR. Os custos ao usar o Rapid SMART-9N caíram 45% e 40% em comparação com o SMART-9N e multiplex PCR, respectivamente. Por último, ao utilizar metade do volume para a amplificação com os *primers* contendo os *barcodes* da ONT, dobramos o número de amostras que podem ser processadas com cada kit, diminuindo ainda mais o custo. Este protocolo tem potencial para ser otimizado ainda mais e a partir da utilização de enzimas com formulação liofilizada, podemos eliminar qualquer cadeia *cold-chain*. Esses resultados demonstram que o Rapid SMART-9N é uma abordagem

importante tanto para trabalhos em ambientes laboratoriais quanto em campo.

Limitações do método: A grande maioria das *reads* geradas são derivadas do hospedeiro humano, principalmente em amostras clínicas com altos valores de Ct (com baixa abundância relativa do material genômico viral) ou em amostras com material genético degradado devido ao transporte incorreto ou longo/incorreto armazenamento. Este é um fator limitante da sensibilidade analítica da abordagem para detecção de agentes infecciosos, que pode resultar em profundidade de sequenciamento insuficiente para análise do organismo de interesse. Essa desvantagem é mostrada em nosso estudo ao comparar as *reads* mapeadas virais dos isolados de ZIKV com as amostras clínicas de YFV e SARS-CoV-2. A menor proporção de *reads* virais de acordo com o aumento dos valores de Ct é provavelmente devido ao nível total de *background* de *reads* não virais / ácido nucleico do hospedeiro presentes, e também devido à variabilidade entre os pacientes/tipo de amostras e/ou no manuseio das amostra durante a coleta e armazenamento.

O protocolo SMART-9N descrito acima foi utilizado na identificação do terceiro caso natural de infecção fatal pelo vírus Sabiá em casos suspeitos de febre amarela com febre hemorrágica São Paulo, 2019. em Os casos foram reportados em (https://github.com/CADDE-CENTRE/Miscellaneous/blob/master/Arenavirus%20Report.pdf ). O caso foi primeiro notificado ao Centro de Vigilância Epidemiológica Prof. Alexandre Vranjac, Secretário de Saúde do Estado de São Paulo, em 21 de janeiro de 2020. Posteriormente, realizamos uma caracterização mais extensa do vírus utilizando o sequenciamento por metagenômica SMART-9N e ferramentas comumente utilizadas e já validadas para identificar a infecção por SABV no Brasil (PCR). Descobrimos então, quando analisamos as sequências geradas pelo protocolo de metagenômica, que as ferramentas de PCR usadas atualmente podem não ser capazes de detectar cepas de SABV circulantes devido à diversidade genética do vírus nos locais de ligação do primers. Usando os dados genômicos gerados por nossa equipe com o protocolo descrito aqui, projetamos um novo esquema de PCR e um protocolo nested PCR mais sensível que detectou com sucesso SABV em uma nova amostra de baixo título viral que tinha sido também testada negativa para YFV. A estrutura descrita aqui, do teste metagenômico ao sequenciamento do vírus, desenvolvimento e validação de ferramentas de diagnóstico laboratorial de PCR padrão e nested para cepas divergentes e viremia variável, foi conduzido em um total de 10 dias

# 5 CONCLUSÃO

Aqui, demonstramos 4 novos métodos para fluxo de trabalho de alta sensibilidade em amostras virais isoladas e clínicas que tiram proveito da tecnologia de sequenciamento por nanoporos. Geramos um painel para detecção e sequenciamento de 14 arbovírus em uma única reação, gerando coberturas semelhantes ou até maiores quando comparados com os ensaios contendo apenas o vírus alvo. Geramos uma abordagem rápida de sequenciamento contendo *primers* alvos gene específicos e ligação de *barcodes* durante um ensaio de PCR convencional, que diminuiu a complexidade do sequenciamento e o tempo em 62% e o custo em até US\$ 10 por amostra. Por último, geramos uma abordagem rápida de sequenciamento gerando produtos de amplificação de cDNA ultra longos (até 18,5 kb) para metagenômica viral. Nossas abordagens de sequenciamento oferecem uma oportunidade para identificação e caracterização sensível de vírus de RNA diretamente de isolados ou amostras clínicas com uma variedade de cargas virais e permite também a detecção e identificação de vírus não identificados a partir dos métodos metagenômicos. Além disso, os protocolos rápidos demonstraram ser metodologias simples, de baixo custo e mais rápidos, promissores para uso rotineiro em laboratórios de pesquisa e também em campo.

# REFERÊNCIAS

1. Gubler DJ. Human arbovirus infections worldwide. Ann N Y Acad Sci. 2001 Dec;951:13-24. doi: 10.1111/j.1749-6632.2001.tb02681.x. PMID: 11797771.

2. Instituto Brasileiro de Geografia e Estatística [homepage na internet]. Areas territoriais [acesso em 25 agos 2021]. Disponível em: <u>http://www.ibge.gov.br</u>

3. Rosa Amelia Paes de Andrade Travassos da. The history of Arbovirology at Instituto Evandro Chagas, Belém, Pará, Brazil, from 1954 to 1998. Rev Pan-Amaz Saude [Internet]. 2016 Dez [citado 2021 Ago 03] ; 7( esp ): 61-70. Disponível em: http://scielo.iec.gov.br/scielo.php?script=sci\_arttext&pid=S2176-62232016000500061&lng=p t. http://dx.doi.org/10.5123/s2176-62232016000500007.

4. Figueiredo LT. Emergent arboviruses in Brazil. Rev Soc Bras Med Trop. 2007 Mar-Apr;40(2):224-9. doi: 10.1590/s0037-86822007000200016. PMID: 17568894.

5. Calisher CH, Gutiérrez E, Maness KS, Lord RD. Isolation of Mayaro virus from a migrating bird captured in Louisiana in 1967. Bull Pan Am Health Organ. 1974;8(3):243-8. PMID: 4418030.

6. da Rosa JF, da Rosa AP, Dégallier N, Vasconcelos PF. Caracterização e relacionamento antigênico de três novos Bunyavirus no grupo Anopheles A (Bunyaviridae) dos arbovirus [Characterization and antigenic relationship of 3 new Bunyaviruses in the group Anopheles A (Bunyaviridae) of arboviruses]. Rev Saude Publica. 1992 Jun;26(3):173-8. Portuguese. doi: 10.1590/s0034-89101992000300008. PMID: 1342498.

7. Porterfield JS, Casals J, Chumakov MP, Gaidamovich SY, Hannoun C, Holmes IH, Horzinek MC, Mussgay M, Oker-Blom N, Russell PK, Trent DW. Togaviridae. Intervirology. 1978;9(3):129-48. doi: 10.1159/000148930. PMID: 618831.

8. Westaway EG, Brinton MA, Gaidamovich SYa, Horzinek MC, Igarashi A, Kääriäinen L, Lvov DK, Porterfield JS, Russell PK, Trent DW. Flaviviridae. Intervirology. 1985;24(4):183-92. doi: 10.1159/000149642. PMID: 3000978.

9. Calisher CH, Karabatsos N, Dalrymple JM, Shope RE, Porterfield JS, Westaway EG, Brandt WE. Antigenic relationships between flaviviruses as determined by cross-neutralization tests with polyclonal antisera. J Gen Virol. 1989 Jan;70 (Pt 1):37-43. doi: 10.1099/0022-1317-70-1-37. PMID: 2543738.

10. Devaux CA. Emerging and re-emerging viruses: A global challenge illustrated by Chikungunya virus outbreaks. World J Virol. 2012 Feb 12;1(1):11-22. doi: 10.5501/wjv.v1.i1.11. PMID: 24175207; PMCID: PMC3782263.

11. Campos GS, Bandeira AC, Sardi SI. Zika Virus Outbreak, Bahia, Brazil. Emerg Infect Dis. 2015 Oct;21(10):1885-6. doi: 10.3201/eid2110.150847. PMID: 26401719; PMCID: PMC4593454.

12. Zanluca C, Melo VC, Mosimann AL, Santos GI, Santos CN, Luz K. First report of autochthonous transmission of Zika virus in Brazil. Mem Inst Oswaldo Cruz. 2015 Jun;110(4):569-72. doi: 10.1590/0074-02760150192. Epub 2015 Jun 9. PMID: 26061233; PMCID: PMC4501423.

13. Pan American Health Organization. Zika Virus Infection. Step-by-Step Guide to Risk Communication and Community Engagement. Washington, D.C. : PAHO; 2017.

14. Portal da Saúde [homepage na internet]. Situação epidemiológica da infecção pelo vírus Zika no Brasil, de 2015 a 2017 [acesso em 25 agos 2021]. Disponível em: <u>http://portalsaude.saude.gov.br</u>

15. de Oliveira WK, Carmo EH, Henriques CM, Coelho G, Vazquez E, Cortez-Escalante J, Molina J, Aldighieri S, Espinal MA, Dye C. Zika Virus Infection and Associated Neurologic Disorders in Brazil. N Engl J Med. 2017 Apr 20;376(16):1591-1593. doi: 10.1056/NEJMc1608612. Epub 2017 Mar 29. PMID: 28402236; PMCID: PMC5544116.

16. Faria NR, Quick J, Claro IM, Thézé J, de Jesus JG, Giovanetti M, Kraemer MUG, Hill SC, Black A, da Costa AC, Franco LC, Silva SP, Wu CH, Raghwani J, Cauchemez S, du Plessis L, Verotti MP, de Oliveira WK, Carmo EH, Coelho GE, Santelli ACFS, Vinhal LC, Henriques CM, Simpson JT, Loose M, Andersen KG, Grubaugh ND, Somasekar S, Chiu CY, Muñoz-Medina JE, Gonzalez-Bonilla CR, Arias CF, Lewis-Ximenez LL, Baylis SA, Chieppe AO, Aguiar SF, Fernandes CA, Lemos PS, Nascimento BLS, Monteiro HAO, Siqueira IC, de Queiroz MG, de Souza TR, Bezerra JF, Lemos MR, Pereira GF, Loudal D, Moura LC, Dhalia R, França RF, Magalhães T, Marques ET Jr, Jaenisch T, Wallau GL, de Lima MC, Nascimento V, de Cerqueira EM, de Lima MM, Mascarenhas DL, Neto JPM, Levin AS, Tozetto-Mendoza TR, Fonseca SN, Mendes-Correa MC, Milagres FP, Segurado A, Holmes EC, Rambaut A, Bedford T, Nunes MRT, Sabino EC, Alcantara LCJ, Loman NJ, Pybus OG. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature. 2017 Jun 15;546(7658):406-410. doi: 10.1038/nature22401. Epub 2017 May 24. PMID: 28538727; PMCID: PMC5722632.

17. Kasprzykowski JI, Fukutani KF, Fabio H, Fukutani ER, Costa LC, Andrade BB, Queiroz ATL. A recursive sub-typing screening surveillance system detects the appearance of the ZIKV African lineage in Brazil: Is there a risk of a new epidemic? Int J Infect Dis. 2020 Jul;96:579-581. doi: 10.1016/j.ijid.2020.05.090. Epub 2020 Jun 1. PMID: 32497802.

18. Paula Rodrigues de Almeida, Luiza Presser Ehlers, Meriane Demoliner, Ana Karolina Antunes Eisen, Viviane Girardi, Cíntia De Lorenzo, Matheus Viezzer Bianchi, LaurenMello, Saulo Pettinati Pavarini, David Driemeier, Luciana Sonne, Fernando RosadoSpilki. Detection of a novel African-lineage-like Zika virus naturally infecting free-living neotropical primates in Southern Brazil. bioRxiv 828871; doi: <u>https://doi.org/10.1101/828871</u>

19. Jeronimo Alencar, Cecilia Ferreira de Mello, Carlos Brisola Marcondes, Anthony ÉricoGuimarães, Helena Keiko Toma, Amanda Queiroz Bastos, Shayenne Olsson Freitas Silva, Sergio Lisboa Machado. Natural infection and vertical transmission of two flaviviruses

(Yellow fever and Zika) in mosquitoes in primary forests in the Brazilian state of Rio de Janeiro (Diptera: Culicidae)bioRxiv 688713; doi: <u>https://doi.org/10.1101/688713</u>

20. Portal da Saúde [homepage na internet]. Monitoramento dos casos de arboviroses urbanas causados por vírus transmitidos pelo mosquito Aedes (dengue, chikungunya e zika), semanas epidemiológicas 1 a 8, 2021. [acesso em 25 agos 2021]. Disponível em: <u>http://portalsaude.saude.gov.br</u>

21. Nunes MR, Faria NR, de Vasconcelos JM, Golding N, Kraemer MU, de Oliveira LF, Azevedo Rdo S, da Silva DE, da Silva EV, da Silva SP, Carvalho VL, Coelho GE, Cruz AC, Rodrigues SG, Vianez JL Jr, Nunes BT, Cardoso JF, Tesh RB, Hay SI, Pybus OG, Vasconcelos PF. Emergence and potential for spread of Chikungunya virus in Brazil. BMC Med. 2015 Apr 30;13:102. doi: 10.1186/s12916-015-0348-x. PMID: 25976325; PMCID: PMC4433093.

22. Chen R, Puri V, Fedorova N, Lin D, Hari KL, Jain R, Rodas JD, Das SR, Shabman RS, Weaver SC. Comprehensive Genome Scale Phylogenetic Study Provides New Insights on the Global Expansion of Chikungunya Virus. J Virol. 2016 Nov 14;90(23):10600-10611. doi: 10.1128/JVI.01166-16. PMID: 27654297; PMCID: PMC5110187.

23. Naveca FG, Claro I, Giovanetti M, de Jesus JG, Xavier J, Iani FCM, do Nascimento VA, de Souza VC, Silveira PP, Lourenço J, Santillana M, Kraemer MUG, Quick J, Hill SC, Thézé J, Carvalho RDO, Azevedo V, Salles FCDS, Nunes MRT, Lemos PDS, Candido DDS, Pereira GC, Oliveira MAA, Meneses CAR, Maito RM, Cunha CRSB, Campos DPS, Castilho MDC, Siqueira TCDS, Terra TM, de Albuquerque CFC, da Cruz LN, Abreu AL, Martins DV, Simoes DSMV, Aguiar RS, Luz SLB, Loman N, Pybus OG, Sabino EC, Okumoto O, Alcantara LCJ, Faria NR. Genomic, epidemiological and digital surveillance of Chikungunya virus in the Brazilian Amazon. PLoS Negl Trop Dis. 2019 Mar 7;13(3):e0007065. doi: 10.1371/journal.pntd.0007065. PMID: 30845267; PMCID: PMC6424459.

24. Petersen LR, Powers AM. Chikungunya: epidemiology. F1000Res. 2016 Jan 19;5:F1000 Faculty Rev-82. doi: 10.12688/f1000research.7171.1. PMID: 26918158; PMCID: PMC4754000.

25. Hill SC, de Souza R, Thézé J, Claro I, Aguiar RS, Abade L, Santos FCP, Cunha MS, Nogueira JS, Salles FCS, Rocco IM, Maeda AY, Vasami FGS, du Plessis L, Silveira PP, de Jesus JG, Quick J, Fernandes NCCA, Guerra JM, Réssio RA, Giovanetti M, Alcantara LCJ, Cirqueira CS, Díaz-Delgado J, Macedo FLL, Timenetsky MDCST, de Paula R, Spinola R, Telles de Deus J, Mucci LF, Tubaki RM, de Menezes RMT, Ramos PL, de Abreu AL, Cruz LN, Loman N, Dellicour S, Pybus OG, Sabino EC, Faria NR. Genomic Surveillance of Yellow Fever Virus Epizootic in São Paulo, Brazil, 2016 - 2018. PLoS Pathog. 2020 Aug 7;16(8):e1008699. doi: 10.1371/journal.ppat.1008699. PMID: 32764827; PMCID: PMC7437926.

26. European Centre for Disease Prevention and Control [homepage na internet]. Yellow fever, 2017. [acesso em 25 agos 2021]. Disponível em: <u>https://ecdc.europa.eu/en/yellow-fever</u>

27. Portal da Saúde [homepage na internet]. Situação epidemiológica da febre amarela -

Monitoramento 2020/2021 [acesso em 25 agos 2021]. Disponível em: <u>https://www.gov.br/saude/</u>

28. Fares RC, Souza KP, Añez G, Rios M. Epidemiological Scenario of Dengue in Brazil. Biomed Res Int. 2015;2015:321873.

29. Portal da Saúde [homepage na internet]. Monitoramento dos casos de dengue, febre de chikungunya e febre pelo vírus Zika até a Semana Epidemiológica 49, 2016 [acesso em 25 agos 2021]. Disponível em: <u>https://portalarquivos2.saude.gov.br/</u>

30. Portal da Saúde [homepage na internet]. Monitoramento dos casos de arboviroses urbanas causados por vírus transmitidos pelo mosquito Aedes (dengue, chikungunya e zika), semanas epidemiológicas 1 a 8, 2021. [acesso em 25 agos 2021]. Disponível em: <u>http://portalsaude.saude.gov.br</u>

31. Romero-Alvarez D, Escobar LE. Oropouche fever, an emergent disease from the Americas. Microbes Infect. 2018 Mar;20(3):135-146. doi: 10.1016/j.micinf.2017.11.013. Epub 2017 Dec 14. PMID: 29247710.

32. Esposito DLA, Fonseca BALD. Will Mayaro virus be responsible for the next outbreak of an arthropod-borne virus in Brazil? Braz J Infect Dis. 2017 Sep-Oct;21(5):540-544. doi: 10.1016/j.bjid.2017.06.002. Epub 2017 Jul 7. PMID: 28688628.

33. Almeida MAB, Santos ED, Cardoso JDC, Noll CA, Lima MM, Silva FAE, Ferreira MS, Martins LC, Vasconcelos PFDC, Bicca-Marques JC. Detection of antibodies against Icoaraci, Ilhéus, and Saint Louis Encephalitis arboviruses during yellow fever monitoring surveillance in non-human primates (Alouatta caraya) in southern Brazil. J Med Primatol. 2019 Aug;48(4):211-217. doi: 10.1111/jmp.12417. Epub 2019 Apr 29. PMID: 31032984.

34. Castro-Jorge LA, Siconelli MJL, Ribeiro BDS, Moraes FM, Moraes JB, Agostinho MR, Klein TM, Floriano VG, Fonseca BALD. West Nile virus infections are here! Are we prepared to face another flavivirus epidemic? Rev Soc Bras Med Trop. 2019 Mar 28;52:e20190089. doi: 10.1590/0037-8682-0089-2018. PMID: 30942263.

35. Miller S, Naccache SN, Samayoa E, Messacar K, Arevalo S, Federman S, Stryke D, Pham E, Fung B, Bolosky WJ, Ingebrigtsen D, Lorizio W, Paff SM, Leake JA, Pesano R, DeBiasi R, Dominguez S, Chiu CY. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. Genome Res. 2019 May;29(5):831-842. doi: 10.1101/gr.238170.118. Epub 2019 Apr 16. PMID: 30992304; PMCID: PMC6499319.

36. Washington JA. Principles of Diagnosis. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 10.Available from: <u>https://www.ncbi.nlm.nih.gov/books/NBK8014/</u>

37. Vasconcelos PF and Calisher CH. 2016. "Emergence of Human Arboviral Diseases in the Americas, 2000-2016." Vector Borne Zoonotic Dis. doi:10.1089/vbz.2016.1952.

38. Miten Jain, Sergey Koren, Karen H Miga, Josh Quick, Arthur C Rand, Thomas A Sasani. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nature. 2018.

# DOI:10.1038/nbt.4060.

39. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, Bore JA, Koundouno R, Dudas G, Mikhail A, Ouédraogo N, Afrough B, Bah A, Baum JH, Becker-Ziaja B, Boettcher JP, Cabeza-Cabrerizo M, Camino-Sanchez A, Carter LL, Doerrbecker J, Enkirch T, Dorival IGG, Hetzelt N, Hinzmann J, Holm T, Kafetzopoulou LE, Koropogui M, Kosgey A, Kuisma E, Logue CH, Mazzarelli A, Meisel S, Mertens M, Michel J, Ngabo D, Nitzsche K, Pallash E, Patrono LV, Portmann J, Repits JG, Rickett NY, Sachse A, Singethan K, Vitoriano I, Yemanaberhan RL, Zekeng EG, Trina R, Bello A, Sall AA, Faye O, Faye O, Magassouba N, Williams CV, Amburgey V, Winona L, Davis E, Gerlach J, Washington F, Monteil V, Jourdain M, Bererd M, Camara A, Somlare H, Camara A, Gerard M, Bado G, Baillet B, Delaune D, Nebie KY, Diarra A, Savane Y, Pallawo RB, Gutierrez GJ, Milhano N, Roger I, Williams CJ, Yattara F, Lewandowski K, Taylor J, Rachwal P, Turner D, Pollakis G, Hiscox JA, Matthews DA, O'Shea MK, Johnston AM, Wilson D, Hutley E, Smit E, Di Caro A, Woelfel R, Stoecker K, Fleischmann E, Gabriel M, Weller SA, Koivogui L, Diallo B, Keita S, Rambaut A, Formenty P, Gunther S, Carroll MW. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016 Feb 11;530(7589):228-232. doi: 10.1038/nature16996. Epub 2016 Feb 3. PMID: 26840485; PMCID: PMC4817224.

40. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, Oliveira G, Robles-Sikisaka R, Rogers TF, Beutler NA, Burton DR, Lewis-Ximenez LL, de Jesus JG, Giovanetti M, Hill SC, Black A, Bedford T, Carroll MW, Nunes M, Alcantara LC Jr, Sabino EC, Baylis SA, Faria NR, Loose M, Simpson JT, Pybus OG, Andersen KG, Loman NJ. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc. 2017 Jun;12(6):1261-1276. doi: 10.1038/nprot.2017.066. Epub 2017 May 24. PMID: 28538739; PMCID: PMC5902022.

41. de Oliveira EC, Fonseca V, Xavier J, Adelino T, Morales Claro I, Fabri A, Marques Macario E, Viniski AE, Campos Souza CL, Gomes da Costa ES, Soares de Sousa C, Guimarães Dias Duarte F, Correia de Medeiros A, Campelo de Albuquerque CF, Venancio Cunha R, Oliveira De Moura NF, Bispo de Filippis AM, Oliveira T, Lourenço J, de Abreu AL, Alcantara LCJ, Giovanetti M. Short report: Introduction of chikungunya virus ECSA genotype into the Brazilian Midwest and its dispersion through the Americas. PLoS Negl Trop Dis. 2021 Apr 16;15(4):e0009290. doi: 10.1371/journal.pntd.0009290. PMID: 33861753; PMCID: PMC8051810.

42. Xavier J, Fonseca V, Bezerra JF, do Monte Alves M, Mares-Guia MA, Claro IM, de Jesus R, Adelino T, Araújo E, Cavalcante KRLJ, Tosta S, de Souza TR, Moreira da Cruz FE, de Araújo Fabri A, de Oliveira EC, de Moura NFO, do Carmo Said RF, de Albuquerque CFC, Azevedo V, de Oliveira T, de Filippis AMB, Venâncio da Cunha R, Luz KG, Giovanetti M, Alcantara LCJ. Chikungunya virus ECSA lineage reintroduction in the northeasternmost region of Brazil. Int J Infect Dis. 2021 Apr;105:120-123. doi: 10.1016/j.ijid.2021.01.026. Epub 2021 Jan 18. PMID: 33476757.

43. Gräf T, Vazquez C, Giovanetti M, de Bruycker-Nogueira F, Fonseca V, Claro IM, de Jesus JG, Gómez A, Xavier J, de Mendonça MCL, Villalba S, Torales J, Gamarra ML, Thézé J, de Filippis AMB, Azevedo V, de Oliveira T, Franco L, de Albuquerque CFC, Irala S, Holmes EC, Méndez Rico JA, Alcantara LCJ. Epidemiologic History and Genetic Diversity Origins of Chikungunya and Dengue Viruses, Paraguay. Emerg Infect Dis. 2021

May;27(5):1393-1404. doi: 10.3201/eid2705.204244. PMID: 33900172; PMCID: PMC8084490.

44. Faria NR, Kraemer MUG, Hill SC, Goes de Jesus J, Aguiar RS, Iani FCM, Xavier J, Quick J, du Plessis L, Dellicour S, Thézé J, Carvalho RDO, Baele G, Wu CH, Silveira PP, Arruda MB, Pereira MA, Pereira GC, Lourenço J, Obolski U, Abade L, Vasylyeva TI, Giovanetti M, Yi D, Weiss DJ, Wint GRW, Shearer FM, Funk S, Nikolay B, Fonseca V, Adelino TER, Oliveira MAA, Silva MVF, Sacchetto L, Figueiredo PO, Rezende IM, Mello EM, Said RFC, Santos DA, Ferraz ML, Brito MG, Santana LF, Menezes MT, Brindeiro RM, Tanuri A, Dos Santos FCP, Cunha MS, Nogueira JS, Rocco IM, da Costa AC, Komninakis SCV, Azevedo V, Chieppe AO, Araujo ESM, Mendonça MCL, Dos Santos CC, Dos Santos CD, Mares-Guia AM, Nogueira RMR, Sequeira PC, Abreu RG, Garcia MHO, Abreu AL, Okumoto O, Kroon EG, de Albuquerque CFC, Lewandowski K, Pullan ST, Carroll M, de Oliveira T, Sabino EC, Souza RP, Suchard MA, Lemey P, Trindade GS, Drumond BP, Filippis AMB, Loman NJ, Cauchemez S, Alcantara LCJ, Pybus OG. Genomic and epidemiological of yellow fever virus transmission potential. Science. monitoring 2018 Aug 31;361(6405):894-899. doi: 10.1126/science.aat7115. Epub 2018 Aug 23. PMID: 30139911; PMCID: PMC6874500.

45. Naveca FG, Claro I, Giovanetti M, de Jesus JG, Xavier J, Iani FCM, do Nascimento VA, de Souza VC, Silveira PP, Lourenço J, Santillana M, Kraemer MUG, Quick J, Hill SC, Thézé J, Carvalho RDO, Azevedo V, Salles FCDS, Nunes MRT, Lemos PDS, Candido DDS, Pereira GC, Oliveira MAA, Meneses CAR, Maito RM, Cunha CRSB, Campos DPS, Castilho MDC, Siqueira TCDS, Terra TM, de Albuquerque CFC, da Cruz LN, Abreu AL, Martins DV, Simoes DSMV, Aguiar RS, Luz SLB, Loman N, Pybus OG, Sabino EC, Okumoto O, Alcantara LCJ, Faria NR. Genomic, epidemiological and digital surveillance of Chikungunya virus in the Brazilian Amazon. PLoS Negl Trop Dis. 2019 Mar 7;13(3):e0007065. doi: 10.1371/journal.pntd.0007065. PMID: 30845267; PMCID: PMC6424459.

46. Hill SC, Neto de Vasconcelos J, Granja BG, Thézé J, Jandondo D, Neto Z, Mirandela M, Sebastião CDS, Cândido ALM, Clemente C, Pereira da Silva S, de Oliveira T, Pybus OG, Faria NR, Afonso JM. Early Genomic Detection of Cosmopolitan Genotype of Dengue Virus Serotype 2, Angola, 2018. Emerg Infect Dis. 2019 Apr;25(4):784-787. doi: 10.3201/eid2504.180958. PMID: 30882320; PMCID: PMC6433006.

47. Faria NR, Sabino EC, Nunes MR, Alcantara LC, Loman NJ, Pybus OG. Mobile real-time surveillance of Zika virus in Brazil. Genome Med. 2016 Sep 29;8(1):97. doi: 10.1186/s13073-016-0356-2. PMID: 27683027; PMCID: PMC5041528.

48. Houldcroft CJ, Beale MA, Breuer J. Clinical and biological insights from viral genome sequencing. Nat Rev Microbiol. 2017 Mar;15(3):183-192. doi: 10.1038/nrmicro.2016.182. Epub 2017 Jan 16. PMID: 28090077; PMCID: PMC7097211.

49. Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, Conlan S, Quan PL, Hui J, Marshall J, Simons JF, Egholm M, Paddock CD, Shieh WJ, Goldsmith CS, Zaki SR, Catton M, Lipkin WI. A new arenavirus in a cluster of fatal transplant-associated diseases. N Engl J Med. 2008 Mar 6;358(10):991-8. doi: 10.1056/NEJMoa073785. Epub 2008 Feb 6. Erratum in: N Engl J Med. 2008 Mar 13;358(11):1204. PMID: 18256387.

50. Nakamura S, Yang CS, Sakon N, Ueda M, Tougan T, Yamashita A, Goto N, Takahashi K, Yasunaga T, Ikuta K, Mizutani T, Okamoto Y, Tagami M, Morita R, Maeda N, Kawai J, Hayashizaki Y, Nagai Y, Horii T, Iida T, Nakaya T. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. PLoS One. 2009;4(1):e4219. doi: 10.1371/journal.pone.0004219. Epub 2009 Jan 19. PMID: 19156205; PMCID: PMC2625441.

51. Wilson, M.R. *et al.* Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N. Engl. J. Med.* **370**, 2408–2417 (2014).

52. Manso CF, Bibby DF, Mbisa JL. Efficient and unbiased metagenomic recovery of RNA virus genomes from human plasma samples. Sci Rep. 2017 Jun 23;7(1):4173. doi: 10.1038/s41598-017-02239-5. PMID: 28646219; PMCID: PMC5482852.

53. Gu W, Miller S, Chiu CY. Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. Annu Rev Pathol. 2019 Jan 24;14:319-338. doi: 10.1146/annurev-pathmechdis-012418-012751. Epub 2018 Oct 24. PMID: 30355154; PMCID: PMC6345613.

54. Frey KG, Herrera-Galeano JE, Redden CL, Luu TV, Servetas SL, Mateczun AJ, Mokashi VP, Bishop-Lilly KA. Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. BMC Genomics. 2014 Feb 4;15:96. doi: 10.1186/1471-2164-15-96. PMID: 24495417; PMCID: PMC3922542.

55. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. Biotechniques. 2001 Apr;30(4):892-7. doi: 10.2144/01304pf02. PMID: 11314272.

56. Joao Luiz Silva-Filho, Lilian G. de Oliveira, Leticia Monteiro, Pierina L. Parise, Nagela G. Zanluqui, Carolina M. Polonio, Carla L. de Freitas, Daniel A. Toledo-Teixeira, William M. de Souza, Najara Bittencourt, Mariene R. Amorim, Julia Forato, Stéfanie P. Muraro, Gabriela F. de Souza, Matheus C. Martini, Karina Bispo-dos-Santos, Aline Vieira, Carla C. Judice, Glaucia M. Pastore, Eliana Amaral, Renato Passini Junior, Helaine M.B.P. Mayer-Milanez, Carolina C. Ribeiro-do-Valle, Roseli Calil, João Renato Bennini Junior, Giuliane J. Lajos, Albina Altemani, Marcos T. Nolasco da Silva, Ana Carolina Coan, Maria Francisca Colella-Santos, Andrea P.B. von Zuben, Marco Aurélio R. Vinolo, Clarice Weis Arns, Rodrigo Ramos Catharino, Maria Laura Costa, Rodrigo N. Angerami, André R.R. Freitas, Mariangela R. Resende, Márcia T. Garcia, Maria Luiza Moretti, Laurent Renia, Lisa F.P. Ng, Carla V. Rothlin, Fabio T.M. Costa, Jean Pierre Schatzmann Peron, José Luiz Proença-Modena, Gas6 drives Zika virus-induced neurological complications in humans and congenital syndrome in immunocompetent mice, Brain, Behavior, and Immunity, 2021, ISSN 0889-1591, https://doi.org/10.1016/j.bbi.2021.08.008.

57. Kallas EG, D'Elia Zanella LGFAB, Moreira CHV, Buccheri R, Diniz GBF, Castiñeiras ACP, Costa PR, Dias JZC, Marmorato MP, Song ATW, Maestri A, Borges IC, Joelsons D, Cerqueira NB, Santiago E Souza NC, Morales Claro I, Sabino EC, Levi JE, Avelino-Silva VI, Ho YL. Predictors of mortality in patients with yellow fever: an observational cohort study. Lancet Infect Dis. 2019 Jul;19(7):750-758. doi: 10.1016/S1473-3099(19)30125-2. Epub 2019 May 16. Erratum in: Lancet Infect Dis. 2019 Nov;19(11):e370. PMID: 31104909.

58. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078-9. doi: 10.1093/bioinformatics/btp352. Epub 2009 Jun 8. PMID: 19505943; PMCID: PMC2723002.

59. Milne I, Bayer M, Cardle L, Shaw P, Stephen G, Wright F, Marshall D. Tablet--next generation sequence assembly visualization. Bioinformatics. 2010 Feb 1;26(3):401-2. doi: 10.1093/bioinformatics/btp666. Epub 2009 Dec 4. PMID: 19965881; PMCID: PMC2815658.

60. Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu G, Salamat SM, Somasekar S, Federman S, Miller S, Sokolic R, Garabedian E, Candotti F, Buckley RH, Reed KD, Meyer TL, Seroogy CM, Galloway R, Henderson SL, Gern JE, DeRisi JL, Chiu CY. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. N Engl J Med. 2014 Jun 19;370(25):2408-17. doi: 10.1056/NEJMoa1401268. Epub 2014 Jun 4. PMID: 24896819; PMCID: PMC4134948.

61. de Jesus JG, Dutra KR, Sales FCDS, Claro IM, Terzian AC, Candido DDS, Hill SC, Thézé J, Torres C, D'Agostini TL, Felix AC, Reis AFN, Alcantara LCJ, de Abreu AL, Croda JH, de Oliveira WK, de Filipis AMB, Camis MDCRDS, Romano CM. Genomic detection of a virus lineage replacement event of dengue virus serotype 2 in Brazil, 2019. Mem Inst Oswaldo Cruz. 2020;115:e190423. doi: 10.1590/0074-02760190423. Epub 2020 May 15. PMID: 32428189; PMCID: PMC7227788.

62. Kafetzopoulou LE, Efthymiadis K, Lewandowski K, Crook A, Carter D, Osborne J, Aarons E, Hewson R, Hiscox JA, Carroll MW, Vipond R, Pullan ST. Assessment of metagenomic Nanopore and Illumina sequencing for recovering whole genome sequences of chikungunya and dengue viruses directly from clinical samples. Euro Surveill. 2018 Dec;23(50):1800228. doi: 10.2807/1560-7917.ES.2018.23.50.1800228. PMID: 30563591; PMCID: PMC6299504.

63. Lewandowski K, Xu Y, Pullan ST, Lumley SF, Foster D, Sanderson N, Vaughan A, Morgan M, Bright N, Kavanagh J, Vipond R, Carroll M, Marriott AC, Gooch KE, Andersson M, Jeffery K, Peto TEA, Crook DW, Walker AS, Matthews PC. Metagenomic Nanopore Sequencing of Influenza Virus Direct from Clinical Respiratory Samples. J Clin Microbiol. 2019 Dec 23;58(1):e00963-19. doi: 10.1128/JCM.00963-19. PMID: 31666364; PMCID: PMC6935926.

64. Lanciotti RS, Kosoy OL, Laven JJ, Velez JO, Lambert AJ, Johnson AJ, Stanfield SM, Duffy MR. Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. Emerg Infect Dis. 2008 Aug;14(8):1232-9. doi: 10.3201/eid1408.080287. PMID: 18680646; PMCID: PMC2600394.

65. Fischer C, Torres MC, Patel P, Moreira-Soto A, Gould EA, Charrel RN, de Lamballerie X, Nogueira RMR, Sequeira PC, Rodrigues CDS, Kümmerer BM, Drosten C, Landt O, Bispo de Filippis AM, Drexler JF. Lineage-Specific Real-Time RT-PCR for Yellow Fever Virus Outbreak Surveillance, Brazil. Emerg Infect Dis. 2017 Nov;23(11):1867–71. doi: 10.3201/eid2311.171131. Epub 2017 Nov 17. PMID: 28949285; PMCID: PMC5652426.

66. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, Bleicker T, Brünink S, Schneider J, Schmidt ML, Mulders DG, Haagmans BL, van der Veer B, van den Brink S, Wijsman L, Goderski G, Romette JL, Ellis J, Zambon M, Peiris M, Goossens H, Reusken C, Koopmans MP, Drosten C. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill. Jan;25(3):2000045. 2020 doi. 10.2807/1560-7917.ES.2020.25.3.2000045. Erratum in: Euro Surveill. 2020 Apr;25(14): Erratum in: Euro Surveill. 2020 Jul;25(30): PMID: 31992387; PMCID: PMC6988269. 67. Tyson JR, James P, Stoddart D, Sparks N, Wickenhagen A, Hall G, Choi JH, Lapointe H, Kamelian K, Smith AD, Prystajecky N, Goodfellow I, Wilson SJ, Harrigan R, Snutch TP, Loman NJ, Quick J. Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. bioRxiv [Preprint]. 2020 Sep 4:2020.09.04.283077. doi: 10.1101/2020.09.04.283077. PMID: 32908977; PMCID: PMC7480024.

68. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019 Nov 28;20(1):257. doi: 10.1186/s13059-019-1891-0. PMID: 31779668; PMCID: PMC6883579.

69. Ramsköld D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, Daniels GA, Khrebtukova I, Loring JF, Laurent LC, Schroth GP, Sandberg R. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol. 2012 Aug;30(8):777-82. doi: 10.1038/nbt.2282. Erratum in: Nat Biotechnol. 2020 Mar;38(3):374. PMID: 22820318; PMCID: PMC3467340.

70. Loose M, Malla S, Stout M. Real-time selective sequencing using nanopore technology. Nat Methods. 2016 Sep;13(9):751-4. doi: 10.1038/nmeth.3930. Epub 2016 Jul 25. PMID: 27454285; PMCID: PMC5008457.

71. Gire SK, Goba A, Andersen KG, Sealfon RS, Park DJ, Kanneh L, Jalloh S, Momoh M, Fullah M, Dudas G, Wohl S, Moses LM, Yozwiak NL, Winnicki S, Matranga CB, Malboeuf CM, Qu J, Gladden AD, Schaffner SF, Yang X, Jiang PP, Nekoui M, Colubri A, Coomber MR, Fonnie M, Moigboi A, Gbakie M, Kamara FK, Tucker V, Konuwa E, Saffa S, Sellu J, Jalloh AA, Kovoma A, Koninga J, Mustapha I, Kargbo K, Foday M, Yillah M, Kanneh F, Robert W, Massally JL, Chapman SB, Bochicchio J, Murphy C, Nusbaum C, Young S, Birren BW, Grant DS, Scheiffelin JS, Lander ES, Happi C, Gevao SM, Gnirke A, Rambaut A, Garry RF, Khan SH, Sabeti PC. Genomic surveillance elucidates Ebola virus origin and transmission during 12;345(6202):1369-72. the 2014 outbreak. Science. 2014 Sep doi: 10.1126/science.1259657. Epub 2014 Aug 28. PMID: 25214632; PMCID: PMC4431643.

72. Salipante SJ, SenGupta DJ, Cummings LA, Land TA, Hoogestraat DR, Cookson BT. Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology. J Clin Microbiol. 2015 Apr;53(4):1072-9. doi: 10.1128/JCM.03385-14. Epub 2015 Jan 28. PMID: 25631811; PMCID: PMC4365209.

73. Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, García-Cobos S, Kooistra-Smid AM, Raangs EC, Rosema S, Veloo AC, Zhou K, Friedrich AW, Rossen JW. Application of next generation sequencing in clinical microbiology and infection prevention. J Biotechnol. 2017 Feb 10;243:16-24. doi: 10.1016/j.jbiotec.2016.12.022. Epub 2016 Dec 29. PMID: 28042011.

74. Carbo EC, Sidorov IA, Zevenhoven-Dobbe JC, Snijder EJ, Claas EC, Laros JFJ, Kroes ACM, de Vries JJC. Coronavirus discovery by metagenomic sequencing: a tool for pandemic preparedness. J Clin Virol. 2020 Oct;131:104594. doi: 10.1016/j.jcv.2020.104594. Epub 2020 Aug 21. PMID: 32866812; PMCID: PMC7441049.

75. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020 Mar;579(7798):270-273. doi: 10.1038/s41586-020-2012-7. Epub 2020 Feb 3. PMID: 32015507; PMCID: PMC7095418.

76. Nawy T. Single-cell sequencing. Nat Methods. 2014 Jan;11(1):18. doi: 10.1038/nmeth.2771. PMID: 24524131.

77. Eberwine J, Sul JY, Bartfai T, Kim J. The promise of single-cell sequencing. Nat Methods. 2014 Jan;11(1):25-7. doi: 10.1038/nmeth.2769. PMID: 24524134.

78. Stephens Z, Wang C, Iyer RK, Kocher JP. Detection and visualization of complex structural variants from long reads. BMC Bioinformatics. 2018 Dec 21;19(Suppl 20):508. doi: 10.1186/s12859-018-2539-x. PMID: 30577744; PMCID: PMC6302372.

79. Fuselli S, Baptista RP, Panziera A, Magi A, Guglielmi S, Tonin R, Benazzo A, Bauzer LG, Mazzoni CJ, Bertorelle G. A new hybrid approach for MHC genotyping: high-throughput NGS and long read MinION nanopore sequencing, with application to the non-model vertebrate Alpine chamois (Rupicapra rupicapra). Heredity (Edinb). 2018 Oct;121(4):293-303. doi: 10.1038/s41437-018-0070-5. Epub 2018 Mar 24. PMID: 29572469; PMCID: PMC6133961.

80. Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. Nat Rev Genet. 2018 Jan;19(1):9-20. doi: 10.1038/nrg.2017.88. Epub 2017 Nov 13. PMID: 29129921; PMCID: PMC7097748.

81. Holmes EC, Dudas G, Rambaut A, Andersen KG. The evolution of Ebola virus: Insights from the 2013-2016 epidemic. Nature. 2016 Oct 13;538(7624):193-200. doi: 10.1038/nature19790. PMID: 27734858; PMCID: PMC5580494.

82. Faria NR, Mellan TA, Whittaker C, Claro IM, Candido DDS, Mishra S, Crispim MAE, Sales FCS, Hawryluk I, McCrone JT, Hulswit RJG, Franco LAM, Ramundo MS, de Jesus JG, Andrade PS, Coletti TM, Ferreira GM, Silva CAM, Manuli ER, Pereira RHM, Peixoto PS, Kraemer MUG, Gaburo N Jr, Camilo CDC, Hoeltgebaum H, Souza WM, Rocha EC, de Souza LM, de Pinho MC, Araujo LJT, Malta FSV, de Lima AB, Silva JDP, Zauli DAG, Ferreira ACS, Schnekenberg RP, Laydon DJ, Walker PGT, Schlüter HM, Dos Santos ALP, Vidal MS, Del Caro VS, Filho RMF, Dos Santos HM, Aguiar RS, Proença-Modena JL, Nelson B, Hay JA, Monod M, Miscouridou X, Coupland H, Sonabend R, Vollmer M, Gandy A, Prete CA Jr, Nascimento VH, Suchard MA, Bowden TA, Pond SLK, Wu CH, Ratmann O, Ferguson NM, Dye C, Loman NJ, Lemey P, Rambaut A, Fraiji NA, Carvalho MDPSS, Pybus OG, Flaxman S, Bhatt S, Sabino EC. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science. 2021 May 21;372(6544):815-821. doi: 10.1126/science.abh2644. Epub 2021 Apr 14. PMID: 33853970; PMCID: PMC8139423.

83. Candido DS, Claro IM, de Jesus JG, Souza WM, Moreira FRR, Dellicour S, Mellan TA, du Plessis L, Pereira RHM, Sales FCS, Manuli ER, Thézé J, Almeida L, Menezes MT, Voloch CM, Fumagalli MJ, Coletti TM, da Silva CAM, Ramundo MS, Amorim MR, Hoeltgebaum HH, Mishra S, Gill MS, Carvalho LM, Buss LF, Prete CA Jr, Ashworth J, Nakaya HI, Peixoto PS, Brady OJ, Nicholls SM, Tanuri A, Rossi ÁD, Braga CKV, Gerber AL, de C Guimarães AP, Gaburo N Jr, Alencar CS, Ferreira ACS, Lima CX, Levi JE, Granato C, Ferreira GM, Francisco RS Jr, Granja F, Garcia MT, Moretti ML, Perroud MW Jr, Castiñeiras TMPP, Lazari CS, Hill SC, de Souza Santos AA, Simeoni CL, Forato J, Sposito AC, Schreiber AZ, Santos MNN, de Sá CZ, Souza RP, Resende-Moreira LC, Teixeira MM, Hubner J, Leme PAF, Moreira RG, Nogueira ML; Brazil-UK Centre for Arbovirus Discovery, Diagnosis, Genomics and Epidemiology (CADDE) Genomic Network, Ferguson NM, Costa SF, Proenca-Modena JL, Vasconcelos ATR, Bhatt S, Lemey P, Wu CH, Rambaut A, Loman NJ, Aguiar RS, Pybus OG, Sabino EC, Faria NR. Evolution and epidemic spread of SARS-CoV-2 in Brazil. Science. 2020 Sep 4;369(6508):1255-1260. doi: 10.1126/science.abd2161. Epub 2020 Jul 23. PMID: 32703910; PMCID: PMC7402630.

### Anexo I - Ficha do aluno com disciplinas realizadas e concluídas.

Janus - Sistema Administrativo da Pós-Graduação Universidade de São Paulo Faculdade de Medicina FICHA DO ALUNO 5134 - 7693003/1 - Ingra Morales Claro ome da Disciplina Inicio Término Carga Horária Cred. Freq. Conc. Exc. Situação Sigla Alividade Publicação do Artigo "Emergence of the Asian lineage of Zika virus in Angola: an outbreak do 10.1015/S1473-3099(19)30293-2. (1) investigation." The Lancet. Infectious diseases. October 2019, v. 19, Issue 10, p.1138-1147. DOI: 24/05/2017 15/06/2017 1 Atividade do Programa Publicação do Artigo "Establishment and cryptic transmission of Zika virus in Brazil and the Americas." Nature, Junho 2017, v. 546, Issue 7658, 406-410. DOI: 10.1038/nature22401. (2) 24/05/2017 15/06/2017 3 Allvidade Publicação do Artigo "Multiplex PCR method for MiniON and Illumina sequencing of Zika and other virus genomes directly from clinical samples". Nature Protocols, Junho 2017, v 12, Issue 6, Porgama 1261-1276. DOI: 10.1038/nprol.2017.066. (2) 24/05/2017 15/06/2017 3 Patogenia Genômica, Genética de Populações e Estudos de Associação 2/1 05/06/2017 11/06/2017 30 0 0 R N Concluida Alividade Participou do Curso Internacional de Bioinformática em Epidemiologia Molecular e Evolução Viral, ministrando as palestras intituladas "PCR pré-sequenciamento do genoma do virus Chikungury; do e Prepare da biblioteca: leonia", "Organização e alinhamento dos genomas", "PCR pré-sequenciamento do genoma do virus Chikungurya; Pratica" e "Preparo da Biblioteca: Pratica", Belo Programa Horizona, 2017, (3) 25/09/2017 29/09/2017 MIP5714-7/2 Malária e Doenças Tropicais Negligenciadas: Avanços no Conhecimento e Desaflos para o Controle - - N Matricula 14/08/2018 03/09/2018 75 0 BMH5778-1/2 Biologia Molecular: Conceitos e Aplicações (Instituto de Ciências Biomédicas - Universidade de São Paulo) 14/08/2018 17/09/2018 60 4 100 A N Concluida IMT5117-1/5 Tópicos avançados em doenças infecciosas relevantes para a Saúde Internacional (Instituto de Medicina Tropical de São Paulo - Universidade de São Paulo) 2 100 A N Concluida 17/08/2018 29/11/2018 30 Pré-- N matrícula indeferida ICB5752-1/5 Como Comunicar Sua Ciência: Melhorando a Oratória e a Empatia com o Público (Instituto de Ciências Biomédicas - Universidade de São Paulo) 27/08/2018 10/09/2018 30 0 BIO5789-Genômica Evolutiva (Instituto de Biociências - Universidade de São Paulo) N Matricula cancelada 06/05/2019 19/05/2019 60 1M5904-1/2 Interação Vírus-Célula e Seus Alvos Terapêuticos (Instituto de Ciências Biomédicas - Universidade de São Paulo) N Matrícula 08/05/2019 18/06/2019 60 0 Altividade Publicação do Artigo "Genomic- epidemiological and digital surveillance of Chikungunya virus in the Brazilian Amazon." PLoS neglected tropical diseases. March 2019, v. 13, Issue 3, e0007065. do DOI: 10.1371/journal.prtd.0007065. (1) 24/05/2019 15/06/2019 Atividade do Publicação do Artigo "Pre Programa ever: an observational cohort study" The Lancet. Infectious diseases. July 2019, v. 19, Issue 7, p.750-758. DOI: 10.1016/S1473-24/05/2019 15/06/2019 dictors of mortality in pa Alividade Publicação do Artigo 72/ka virus infection among symptomatic patients from two healthcare centers in Sao Paulo Stale, Brazit, prevalence, clinical characteristics, viral detection in body fluids and o serodynamics' Revista do Instituto de Medicina Tropical de São Paulo. April 2019, v. 61. DOI: 10.1590/s1678-9945201961019 : (1) 4 ICB5747-2/4 Ciências Ômicas em Doenças Infeccio: N Matricula cancelada 01/07/2019 21/07/2019 90 0 Crédito Externo 19/08/2019 30/08/2019 2 т.-Crédito Externo Tópicos em Bioinfor т nática IV - Epidemiologia De Surto De Dengue: Capacitação, Pesquisa, Vigilância e Divulgação Científica - Universidade Federal de Minas Gerais (4) 19/08/2019 30/08/2019 - N Matricula cancelada IBI5071-1/3 Abordagens Teóricas e Práticas de Metagenômica para a Descoberta de Vírus (Curso Interunidades: Bioinformática - Unir 1/3 21/10/2019 27/10/2019 0 rsidade de São Paulo) 30 IMT5111-2/2 Tópicos avançados em Medicina Tropical (Instituto de Medicina Tropical de São Paulo - Universidade de São Paulo) 2 100 A N Concluida 20/03/2020 02/07/2020 30 MGT5773-2/2 Desenvolvimento de Idéias, Marcas e Patentes em Medicina 0 - - N Matricula cancelada 06/04/2020 19/04/2020 60 IO5700-5/2 Docência Universitária e Estratégias de Ensino-Aprendizagem (Faculdade de Odontologia - Universidade de São Paulo) 03/08/2020 23/11/2020 75 5 100 A N Concluida

	Créditos mínim	Créditos mínimos exigidos	
	Para exame de qualificação	o Para depósito de tese	
Disciplinas:	12	24	30
Estágios:			
Total:	12	24	30
Créditos Atribuído	s à Tasa: 176		

Observaçõe

Observações: 1) Créditos atribuídos de acordo com o Artigo 60 do Regimento de Põs-Graduação e aprovados pela Comissão de Põs-Graduação, em Sessão de 29/11/2019 2) Créditos atribuídos de acordo com o Artigo 60 do Regimento de Põs-Graduação e aprovados pela Comissão de Põs-Graduação, em Sessão de 05/00/2018 3) Créditos atribuídos de acordo com o Artigo 60 do Regimento de Põs-Graduação e aprovados pela Comissão de Põs-Graduação, em Sessão de 05/00/2018 3) Créditos atribuídos de acordo com o Artigo 60 do Regimento de Põs-Graduação e aprovados pela Comissão de Põs-Graduação, em Sessão de 27/01/2020 4) Disciplinas(s) cursada(s) na(o) Universidade Federal de Minas Gerais. Atribuíção de créditos aprovada pelo(a) Comissão de Põs-Graduação em Sessão de 27/01/2020

Conceito a partir de 02/01/1997: A - Excelente, com direito a crédito, B - Born, com direito a crédito, C - Regular, com direito a crédito, R - Reprovado, T - Transferência. Um(1) crédito equivale a 15 horas de atividade programada. Faria NR, Quick J, Claro IM, Thézé J, de Jesus JG, Giovanetti M, Kraemer MUG, Hill SC, Black A, da Costa AC, Franco LC, Silva SP, Wu CH, Raghwani J, Cauchemez S, du Plessis L, Verotti MP, de Oliveira WK, Carmo EH, Coelho GE, Santelli ACFS, Vinhal LC, Henriques CM, Simpson JT, Loose M, Andersen KG, Grubaugh ND, Somasekar S, Chiu CY, Muñoz-Medina JE, Gonzalez-Bonilla CR, Arias CF, Lewis-Ximenez LL, Baylis SA, Chieppe AO, Aguiar SF, Fernandes CA, Lemos PS, Nascimento BLS, Monteiro HAO, Siqueira IC, de Queiroz MG, de Souza TR, Bezerra JF, Lemos MR, Pereira GF, Loudal D, Moura LC, Dhalia R, França RF, Magalhães T, Marques ET Jr, Jaenisch T, Wallau GL, de Lima MC, Nascimento V, de Cerqueira EM, de Lima MM, Mascarenhas DL, Neto JPM, Levin AS, Tozetto-Mendoza TR, Fonseca SN, Mendes-Correa MC, Milagres FP, Segurado A, Holmes EC, Rambaut A, Bedford T, Nunes MRT, Sabino EC, Alcantara LCJ, Loman NJ, Pybus OG. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature. 2017 Jun 15;546(7658):406-410. doi: 10.1038/nature22401. Epub 2017 May 24. PMID: 28538727; PMCID: PMC5722632.

Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, Oliveira G, Robles-Sikisaka R, Rogers TF, Beutler NA, Burton DR, Lewis-Ximenez LL, de Jesus JG, Giovanetti M, Hill SC, Black A, Bedford T, Carroll MW, Nunes M, Alcantara LC Jr, Sabino EC, Baylis SA, Faria NR, Loose M, Simpson JT, Pybus OG, Andersen KG, Loman NJ. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc. 2017 Jun;12(6):1261-1276. doi: 10.1038/nprot.2017.066. Epub 2017 May 24. PMID: 28538739; PMCID: PMC5902022.

Faria NR, Mellan TA, Whittaker C, Claro IM, Candido DDS, Mishra S, Crispim MAE, Sales FCS, Hawryluk I, McCrone JT, Hulswit RJG, Franco LAM, Ramundo MS, de Jesus JG, Andrade PS, Coletti TM, Ferreira GM, Silva CAM, Manuli ER, Pereira RHM, Peixoto PS, Kraemer MUG, Gaburo N Jr, Camilo CDC, Hoeltgebaum H, Souza WM, Rocha EC, de Souza LM, de Pinho MC, Araujo LJT, Malta FSV, de Lima AB, Silva JDP, Zauli DAG, Ferreira ACS, Schnekenberg RP, Laydon DJ, Walker PGT, Schlüter HM, Dos Santos ALP, Vidal MS, Del Caro VS, Filho RMF, Dos Santos HM, Aguiar RS, Proença-Modena JL, Nelson B, Hay JA, Monod M, Miscouridou X, Coupland H, Sonabend R, Vollmer M, Gandy A, Prete CA Jr, Nascimento VH, Suchard MA, Bowden TA, Pond SLK, Wu CH, Ratmann O, Ferguson NM, Dye C, Loman NJ, Lemey P, Rambaut A, Fraiji NA, Carvalho MDPSS, Pybus OG, Flaxman S, Bhatt S, Sabino EC. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage Manaus, Brazil. Science. 2021 May 21;372(6544):815-821. in doi: 10.1126/science.abh2644. Epub 2021 Apr 14. PMID: 33853970; PMCID: PMC8139423.

Candido DS, Claro IM, de Jesus JG, Souza WM, Moreira FRR, Dellicour S, Mellan TA, du Plessis L, Pereira RHM, Sales FCS, Manuli ER, Thézé J, Almeida L, Menezes MT, Voloch CM, Fumagalli MJ, Coletti TM, da Silva CAM, Ramundo MS, Amorim MR, Hoeltgebaum HH, Mishra S, Gill MS, Carvalho LM, Buss LF, Prete CA Jr, Ashworth J, Nakaya HI, Peixoto PS, Brady OJ, Nicholls SM, Tanuri A, Rossi ÁD, Braga CKV, Gerber AL, de C Guimarães AP, Gaburo N Jr, Alencar CS, Ferreira ACS, Lima CX, Levi JE, Granato C, Ferreira GM, Francisco RS Jr, Granja F, Garcia MT, Moretti ML, Perroud MW Jr, Castiñeiras TMPP, Lazari CS, Hill SC, de Souza Santos AA, Simeoni CL, Forato J, Sposito AC, Schreiber AZ, Santos

MNN, de Sá CZ, Souza RP, Resende-Moreira LC, Teixeira MM, Hubner J, Leme PAF, Moreira RG, Nogueira ML; Brazil-UK Centre for Arbovirus Discovery, Diagnosis, Genomics and Epidemiology (CADDE) Genomic Network, Ferguson NM, Costa SF, Proenca-Modena JL, Vasconcelos ATR, Bhatt S, Lemey P, Wu CH, Rambaut A, Loman NJ, Aguiar RS, Pybus OG, Sabino EC, Faria NR. Evolution and epidemic spread of SARS-CoV-2 in Brazil. Science. 2020 Sep 4;369(6508):1255-1260. doi: 10.1126/science.abd2161. Epub 2020 Jul 23. PMID: 32703910; PMCID: PMC7402630.

Claro IM, da Silva Sales FC, Ramundo MS, Candido DS, Silva CAM, de Jesus JG, Manuli ER, de Oliveira CM, Scarpelli L, Campana G, Pybus OG, Sabino EC, Faria NR, Levi JE. Local Transmission of SARS-CoV-2 Lineage B.1.1.7, Brazil, December 2020. Emerg Infect Dis. 2021 Mar;27(3):970-972. doi: 10.3201/eid2703.210038. Epub 2021 Jan 26. PMID: 33496249; PMCID: PMC7920684.

Jesus JG, Sacchi C, Candido DDS, Claro IM, Sales FCS, Manuli ER, Silva DBBD, Paiva TM, Pinho MAB, Santos KCO, Hill SC, Aguiar RS, Romero F, Santos FCPD, Gonçalves CR, Timenetsky MDC, Quick J, Croda JHR, Oliveira W, Rambaut A, Pybus OG, Loman NJ, Sabino EC, Faria NR. Importation and early local transmission of COVID-19 in Brazil, 2020. Rev Inst Med Trop Sao Paulo. 2020;62:e30. doi: 10.1590/s1678-9946202062030. Epub 2020 May 11. PMID: 32401959; PMCID: PMC7232955.

Giovanetti M, Faria NR, Lourenço J, Goes de Jesus J, Xavier J, Claro IM, Kraemer MUG, Fonseca V, Dellicour S, Thézé J, da Silva Salles F, Gräf T, Silveira PP, do Nascimento VA, Costa de Souza V, de Melo Iani FC, Castilho-Martins EA, Cruz LN, Wallau G, Fabri A, Levy F, Quick J, de Azevedo V, Aguiar RS, de Oliveira T, Bôtto de Menezes C, da Costa Castilho M, Terra TM, Souza da Silva M, Bispo de Filippis AM, Luiz de Abreu A, Oliveira WK, Croda J, Campelo de Albuquerque CF, Nunes MRT, Sabino EC, Loman N, Naveca FG, Pybus OG, Alcantara LC. Genomic and Epidemiological Surveillance of Zika Virus in the Amazon Region. Cell Rep. 2020 Feb 18;30(7):2275-2283.e7. doi: 10.1016/j.celrep.2020.01.085. PMID: 32075736.

Cunha MS, Faria NR, Caleiro GS, Candido DS, Hill SC, Claro IM, da Costa AC, Nogueira JS, Maeda AY, da Silva FG, de Souza RP, Spinola R, Tubaki RM, de Menezes RMT, Abade L, Mucci LF, Timenetsky MDCST, Sabino E. Genomic evidence of yellow fever virus in Aedes scapularis, southeastern Brazil, 2016. Acta Trop. 2020 May;205:105390. doi: 10.1016/j.actatropica.2020.105390. Epub 2020 Feb 7. PMID: 32044285.

de Jesus JG, Dutra KR, Sales FCDS, Claro IM, Terzian AC, Candido DDS, Hill SC, Thézé J, Torres C, D'Agostini TL, Felix AC, Reis AFN, Alcantara LCJ, de Abreu AL, Croda JH, de Oliveira WK, de Filipis AMB, Camis MDCRDS, Romano CM. Genomic detection of a virus lineage replacement event of dengue virus serotype 2 in Brazil, 2019. Mem Inst Oswaldo Cruz. 2020;115:e190423. doi: 10.1590/0074-02760190423. Epub 2020 May 15. PMID: 32428189; PMCID: PMC7227788.

Xavier J, Fonseca V, Bezerra JF, do Monte Alves M, Mares-Guia MA, Claro IM, de Jesus R, Adelino T, Araújo E, Cavalcante KRLJ, Tosta S, de Souza TR, Moreira da Cruz FE, de Araújo Fabri A, de Oliveira EC, de Moura NFO, do Carmo Said RF, de Albuquerque CFC, Azevedo V, de Oliveira T, de Filippis AMB, Venâncio da Cunha R, Luz KG, Giovanetti M, Alcantara LCJ. Chikungunya virus ECSA lineage reintroduction in the northeasternmost

region of Brazil. Int J Infect Dis. 2021 Apr;105:120-123. doi: 10.1016/j.ijid.2021.01.026. Epub 2021 Jan 18. PMID: 33476757.

Hill SC, Vasconcelos J, Neto Z, Jandondo D, Zé-Zé L, Aguiar RS, Xavier J, Thézé J, Mirandela M, Micolo Cândido AL, Vaz F, Sebastião CDS, Wu CH, Kraemer MUG, Melo A, Schamber-Reis BLF, de Azevedo GS, Tanuri A, Higa LM, Clemente C, da Silva SP, da Silva Candido D, Claro IM, Quibuco D, Domingos C, Pocongo B, Watts AG, Khan K, Alcantara LCJ, Sabino EC, Lackritz E, Pybus OG, Alves MJ, Afonso J, Faria NR. Emergence of the Asian lineage of Zika virus in Angola: an outbreak investigation. Lancet Infect Dis. 2019 Oct;19(10):1138-1147. doi: 10.1016/S1473-3099(19)30293-2. PMID: 31559967; PMCID: PMC6892302.

Amorim MR, Souza WM, Barros ACG Jr, Toledo-Teixeira DA, Dos-Santos KB, Simeoni CL, Parise PL, Vieira A, Forato J, Claro IM, Mofatto LS, Barbosa PP, Brunetti NS, França ESS, Pedroso GA, Carvalho BFN, Zaccariotto TR, Krywacz KCS, Vieira AS, Mori MA, Farias AS, Pavan MHP, Bachur LF, Cardoso LGO, Spilki FR, Sabino EC, Faria NR, Santos MNN, Angerami R, Leme PAF, Schreiber A, Moretti ML, Granja F, Proenca-Modena JL. Respiratory Viral Shedding in Healthcare Workers Reinfected with SARS-CoV-2, Brazil, 2020. Emerg Infect Dis. 2021;27(6):1737-1740. doi: 10.3201/eid2706.210558. Epub 2021 Apr 19. PMID: 33871331; PMCID: PMC8153890.

Tozetto-Mendoza TR, Avelino-Silva VI, Fonseca S, Claro IM, Paula AV, Levin AS, Sabino EC, Mendes-Correa MC, Figueiredo WM, Felix AC, Souza NCS, Costa AA, Inenami M, Silva RMGD, Levi JE, Romano CM, Paranhos-Baccalà G, Segurado AC, Mayaud P. Zika virus infection among symptomatic patients from two healthcare centers in Sao Paulo State, Brazil: prevalence, clinical characteristics, viral detection in body fluids and serodynamics. Rev Inst Med Trop Sao Paulo. 2019;61:e19. doi: 10.1590/S1678-9946201961019. Epub 2019 Apr 4. PMID: 30970110; PMCID: PMC6453419.

Gräf T, Vazquez C, Giovanetti M, de Bruycker-Nogueira F, Fonseca V, Claro IM, de Jesus JG, Gómez A, Xavier J, de Mendonça MCL, Villalba S, Torales J, Gamarra ML, Thézé J, de Filippis AMB, Azevedo V, de Oliveira T, Franco L, de Albuquerque CFC, Irala S, Holmes EC, Méndez Rico JA, Alcantara LCJ. Epidemiologic History and Genetic Diversity Origins of Chikungunya and Dengue Viruses, Paraguay. Emerg Infect Dis. 2021 May;27(5):1393-1404. PMID: 33900172; doi: 10.3201/eid2705.204244. PMCID: PMC8084490.

Vogels CBF, Breban MI, Alpert T, Petrone ME, Watkins AE, Ott IM, de Jesus JG, Claro IM, Ferreira GM, Crispim MAE; Brazil-UK CADDE Genomic Network, Singh L, Tegally H, Anyaneji UJ; NGS-SA, Hodcroft EB, Mason CE, Khullar G, Metti J, Dudley JT, MacKay MJ, Nash M, Wang J, Liu C, Hui P, Murphy S, Neal C, Laszlo E, Landry ML, Muyombwe A, Downing R, Razeq J, de Oliveira T, Faria NR, Sabino EC, Neher RA, Fauver JR, Grubaugh ND. PCR assay to enhance global surveillance for SARS-CoV-2 variants of concern. medRxiv [Preprint]. 2021 Mar 12:2021.01.28.21250486. doi: 10.1101/2021.01.28.21250486. Update in: PLoS Biol. 2021 May 7;19(5):e3001236. PMID: 33758901; PMCID: PMC7987060.

Nicolete VC, Rodrigues PT, Johansen IC, Corder RM, Tonini J, Cardoso MA, de Jesus JG, Claro IM, Faria NR, Sabino EC, Castro MC, Ferreira MU. Interacting Epidemics in Amazonian Brazil: Prior Dengue Infection Associated with Increased COVID-19 Risk in a Population-Based Cohort Study. Clin Infect Dis. 2021 May 6:ciab410. doi: 10.1093/cid/ciab410. Epub ahead of print. PMID: 33956939; PMCID: PMC8135953.

Souza WM, Amorim MR, Sesti-Costa R, Coimbra LD, Brunetti NS, Toledo-Teixeira DA, de Souza GF, Muraro SP, Parise PL, Barbosa PP, Bispo-Dos-Santos K, Mofatto LS, Simeoni CL, Claro IM, Duarte ASS, Coletti TM, Zangirolami AB, Costa-Lima C, Gomes ABSP, Buscaratti LI, Sales FC, Costa VA, Franco LAM, Candido DS, Pybus OG, de Jesus JG, Silva CAM, Ramundo MS, Ferreira GM, Pinho MC, Souza LM, Rocha EC, Andrade PS, Crispim MAE, Maktura GC, Manuli ER, Santos MNN, Camilo CC, Angerami RN, Moretti ML, Spilki FR, Arns CW, Addas-Carvalho M, Benites BD, Vinolo MAR, Mori MAS, Gaburo N, Dye C, Marques-Souza H, Marques RE, Farias AS, Diamond MS, Faria NR, Sabino EC, Granja F, Proença-Módena JL. Neutralisation of SARS-CoV-2 lineage P.1 by antibodies elicited through natural SARS-CoV-2 infection or vaccination with an inactivated SARS-CoV-2 vaccine: an immunological study. Lancet Microbe. 2021 Jul 8. doi: 10.1016/S2666-5247(21)00129-4. Epub ahead of print. PMID: 34258603; PMCID: PMC8266272.

Ricas Rezende H, Malta Romano C, Morales Claro I, Santos Caleiro G, Cerdeira Sabino E, Felix AC, Bissoli J, Hill S, Rodrigues Faria N, Cardoso da Silva TC, Brioschi Santos AP, Cerutti Junior C, Vicente CR. First report of Aedes albopictus infected by Dengue and Zika virus in a rural outbreak in Brazil. PLoS One. 2020 Mar 12;15(3):e0229847. doi: 10.1371/journal.pone.0229847. PMID: 32163449; PMCID: PMC7067471.

Vogels CBF, Breban MI, Ott IM, Alpert T, Petrone ME, Watkins AE, Kalinich CC, Earnest R, Rothman JE, Goes de Jesus J, Morales Claro I, Magalhães Ferreira G, Crispim MAE; Brazil-UK CADDE Genomic Network, Singh L, Tegally H, Anyaneji UJ; Network for Genomic Surveillance in South Africa, Hodcroft EB, Mason CE, Khullar G, Metti J, Dudley JT, MacKay MJ, Nash M, Wang J, Liu C, Hui P, Murphy S, Neal C, Laszlo E, Landry ML, Muyombwe A, Downing R, Razeq J, de Oliveira T, Faria NR, Sabino EC, Neher RA, Fauver JR, Grubaugh ND. Multiplex qPCR discriminates variants of concern to enhance global surveillance of SARS-CoV-2. PLoS Biol. 2021 May 7;19(5):e3001236. doi: 10.1371/journal.pbio.3001236. PMID: 33961632; PMCID: PMC8133773.

de Oliveira EC, Fonseca V, Xavier J, Adelino T, Morales Claro I, Fabri A, Marques Macario E, Viniski AE, Campos Souza CL, Gomes da Costa ES, Soares de Sousa C, Guimarães Dias Duarte F, Correia de Medeiros A, Campelo de Albuquerque CF, Venancio Cunha R, Oliveira De Moura NF, Bispo de Filippis AM, Oliveira T, Lourenço J, de Abreu AL, Alcantara LCJ, Giovanetti M. Short report: Introduction of chikungunya virus ECSA genotype into the Brazilian Midwest and its dispersion through the Americas. PLoS Negl Trop Dis. 2021 Apr 16;15(4):e0009290. doi: 10.1371/journal.pntd.0009290. PMID: 33861753; PMCID: PMC8051810.

Adelino TÉR, Giovanetti M, Fonseca V, Xavier J, de Abreu ÁS, do Nascimento VA, Demarchi LHF, Oliveira MAA, da Silva VL, de Mello ALES, Cunha GM, Santos RH, de Oliveira EC, Júnior JAC, de Melo Iani FC, de Filippis AMB, de Abreu AL, de Jesus R, de Albuquerque CFC, Rico JM, do Carmo Said RF, Silva JA, de Moura NFO, Leite P, Frutuoso

LCV, Haddad SK, Martínez A, Barreto FK, Vazquez CC, da Cunha RV, Araújo ELL, de Oliveira Tosta SF, de Araújo Fabri A, Chalhoub FLL, da Silva Lemos P, de Bruycker-Nogueira F, de Castro Lichs GG, Zardin MCSU, Segovia FMC, Gonçalves CCM, Grillo ZDCF, Slavov SN, Pereira LA, Mendonça AF, Pereira FM, de Magalhães JJF, Dos Santos Júnior ACM, de Lima MM, Nogueira RMR, Góes-Neto A, de Carvalho Azevedo VA, Ramalho DB, Oliveira WK, Macario EM, de Medeiros AC, Pimentel V; Latin American Genomic Surveillance Arboviral Network, Holmes EC, de Oliveira T, Lourenço J, Alcantara LCJ. Field and classroom initiatives for portable sequence-based monitoring of dengue virus in Brazil. Nat Commun. 2021 Apr 16;12(1):2296. doi: 10.1038/s41467-021-22607-0. PMID: 33863880; PMCID: PMC8052316.

Domingues RB, Mendes-Correa MC, de Moura Leite FBV, Sabino EC, Salarini DZ, Claro I, Santos DW, de Jesus JG, Ferreira NE, Romano CM, Soares CAS. First case of SARS-COV-2 sequencing in cerebrospinal fluid of a patient with suspected demyelinating disease. J Neurol. 2020 Nov;267(11):3154-3156. doi: 10.1007/s00415-020-09996-w. Epub 2020 Jun 20. PMID: 32564153; PMCID: PMC7305694.

Hill SC, de Souza R, Thézé J, Claro I, Aguiar RS, Abade L, Santos FCP, Cunha MS, Nogueira JS, Salles FCS, Rocco IM, Maeda AY, Vasami FGS, du Plessis L, Silveira PP, de Jesus JG, Quick J, Fernandes NCCA, Guerra JM, Réssio RA, Giovanetti M, Alcantara LCJ, Cirqueira CS, Díaz-Delgado J, Macedo FLL, Timenetsky MDCST, de Paula R, Spinola R, Telles de Deus J, Mucci LF, Tubaki RM, de Menezes RMT, Ramos PL, de Abreu AL, Cruz LN, Loman N, Dellicour S, Pybus OG, Sabino EC, Faria NR. Genomic Surveillance of Yellow Fever Virus Epizootic in São Paulo, Brazil, 2016 - 2018. PLoS Pathog. 2020 Aug 7;16(8):e1008699. doi: 10.1371/journal.ppat.1008699. PMID: 32764827; PMCID: PMC7437926.

Nicolete VC, Rodrigues PT, Johansen IC, Corder RM, Tonini J, Cardoso MA, de Jesus JG, Claro IM, Faria NR, Sabino EC, Castro MC, Ferreira MU. Interacting Epidemics in Amazonian Brazil: Prior Dengue Infection Associated with Increased COVID-19 Risk in a Population-Based Cohort Study. Clin Infect Dis. 2021 May 6:ciab410. doi: 10.1093/cid/ciab410. Epub ahead of print. PMID: 33956939; PMCID: PMC8135953.

Evelyn Patricia Sánchez Espinoza, Marina Cortes Farrel, Saidy Vasconez Nogueira, Anderson Vicente de Paula, Thais Guimarães, Lucy Santos Villas Boas, Marcelo Park, Cristina Carvalho da Silva, Ingra Morales, Lauro Vieira Perdigão Neto. Tania ReginaTozetto-Mendoza, Icaro Boszczowski, Ester Sabino, Maria Cássia Mendes-Correa, Anna Sara Shafferman Levin, Silvia Figueiredo Costa. Are Mobile Phones part of the chain of transmission of SARS-CoV-2 in the hospital? medRxiv 2020.11.02.20224519; doi:https://doi.org/10.1101/2020.11.02.20224519

### Anexo III - Parecer Consubstanciado do CEP



#### PARECER CONSUBSTANCIADO DO CEP

#### DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Metagenômica viral de Dengue, Chikungunya e Zika vírus: Acompanhar, explicar e prever a transmissão e distribuição espaço-temporal no Brasil.

Pesquisador: Ester cerdeira Sabino Área Temática: Versão: 3 CAAE: 53153916.7.0000.0065 Instituição Proponente: Faculdade de Medicina da Universidade de São Paulo Patrocinador Principal: FUNDACAO DE AMPARO A PESQUISA DO ESTADO DE SAO PAULO

#### DADOS DO PARECER

Número do Parecer: 1.448.178

#### Apresentação do Projeto:

Trata-se de estudo relevante em que será realizada o sequenciamento viral para identificação de vírus da dengue DENV, chikungunya CHIKV e zika ZIKV de amostras de sangue obtidas de indivíduos em várias cidades brasileiras.

Espera-se captar no mínimo 400 DENV, 200 CHIKV e 200 ZIKV, amostras que serão sequenciadas os acidos nucleicos totais, visando o sequenciamento de um ou mais virus.

#### Objetivo da Pesquisa:

O principal objetivo do estudo é prever padrões espaço temporais de transmissão de arbovirus (vírus da dengue DENV, chikungunya CHIKV e zika ZIKV) em grandes centros urbanos brasileiros.

#### Avaliação dos Riscos e Benefícios:

A pesquisa apresenta baixo risco para o participante: coleta de sangue por punção venosa. Benefícios: sem benefícios diretos para o participante; a previsão de padrões espaço temporais de transmissão de arbovirus em grandes centros urbanos brasileiros pode melhorar a compreensão da epidemia e influenciar ações públicas

#### Comentários e Considerações sobre a Pesquisa:

Consta no projeto de pesquisa que serão analisadas amostras disponíveis no biobanco do IMT (pag

Endereço: DOUTOR ARNALDO 251 21º andar sala 36				
Bairro: PACAEMBU	CEP:	: 01.246-903		
UF: SP Municip	o: SAO PAULO			
Telefone: (11)3893-4401		E-mail: cep.fm@usp.br		

Página 01 de 03

# Thank you for Submitting your Article

Rapid viral metagenomics using SMART-9N amplification and nanopore sequencing Ingra M. Claro et al.

# WHAT WE DO NEXT

#### Before accepting your article

We will check: content suitability, readability and manuscript format; adherence to ethical standards for the type of study; that the underlying data have been supplied (where appropriate); and that there is sufficient detail to enable others to replicate the study (if applicable).

### Before publishing your article

If we accept your article, we will be in touch in the next two or three working days with any issues that need addressing. You will then receive a final proof of your article for approval prior to publication.

### WHAT YOU NEED TO DO NEXT

#### Suggest Reviewers

We will need you to suggest at least 5 suitable reviewers to peer review your manuscript following publication (in accordance with our publishing model) and we will need these names before we can publish the article. We will be in touch shortly with a link to a page where you can suggest reviewers.

Wellcome Open Research strongly supports diversity and inclusion. Providing your demographic profile will help us accurately monitor our progress and achieve a more diverse and inclusive community.

In accordance with our privacy policy, we will not share your personal information with third parties.