UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa de Pós-Graduação em Genética

Joilson Xavier dos Santos Junior

DETECÇÃO E ANÁLISE FILOGENÉTICA DO *CHIKUNGUNYA VIRUS* E DE *ORTHOFLAVIVIRUS* DURANTE A VIGILÂNCIA DE ARBOVÍRUS NO BRASIL

Belo Horizonte
2023

Joilson Xavier dos Santos Junior

DETECÇÃO E ANÁLISE FILOGENÉTICA DO *CHIKUNGUNYA VIRUS* E DE *ORTHOFLAVIVIRUS* DURANTE A VIGILÂNCIA DE ARBOVÍRUS NO BRASIL

Tese de Doutorado submetida ao Programa de Pós-Graduação em Genética da Universidade Federal de Minas Gerais como requisito parcial à obtenção do título de Doutor em Genética.

Orientador: Prof. Dr. Luiz Carlos Junior Alcantara
Coorientadora: Dra. Marta Giovanetti

Belo Horizonte
2023

043 Santos Junior, Joilson Xavier dos.
Detecção e análise filogenética do Chikungunya virus e de Orthoflavivirus durante a vigilância de arbovírus no Brasil [manuscrito] / Joilson Xavier dos Santos Junior. – 2023.
156 f. : il. ; 29,5 cm.

Orientador: Prof. Dr. Luiz Carlos Junior Alcantara. Coorientadora: Dra. Marta Giovanetti.
Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Genética.

1. Genética. 2. Epidemiologia. 3. Vírus Chikungunya. 4. Genômica. 5. Sequenciamento por Nanoporos. I. Alcantara, Luiz Carlos Junior. II. Giovanetti, Marta. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 575

Ficha catalográfica elaborada pela Bibliotecária Fabiane C M Reis. CRB/6 - 2680

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa de Pós-Graduação em Genética

ATA DE DEFESA DE TESE

ATA DA DEFESA DE TESE	**179/2023** **entrada**
Joilson Xavier dos Santos Junior	**1º/2019** **CPF: 055.359.595-45**

Às nove horas do dia **07 de novembro de 2023**, reuniu-se, no Instituto de Ciências Biológicas da UFMG, a Comissão Examinadora de Tese, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: **"Detecção e análise filogenética do *Chikungunya virus* e de *Orthoflavivirus* durante a vigilância de arbovírus no Brasil"**, requisito para obtenção do grau de Doutor em **Genética.** Abrindo a sessão, o Presidente da Comissão, **Luiz Carlos Júnior Alcantara**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa do candidato. Logo após, a Comissão se reuniu, sem a presença do candidato e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Prof./Pesq.	Instituição	CPF	Indicação
Luiz Carlos Júnior Alcantara	FIOCRUZ	576.261.896-04	APROVADO
Renan Pedra de Souza	UFMG	064.488.066-01	APROVADO
Pedro Augusto Alves	FIOCRUZ-MG	073.267.056-07	APROVADO
Filipe Ferreira de Almeida Rego	Escola Bahiana de Medicina e Saúde Pública	011.054.515-00	APROVADO
Erna Geessien Kroon	UFMG	290.320.679-15	APROVADO

Pelas indicações, o candidato foi considerado: APROVADO

O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 07 de novembro de 2023.

Luiz Carlos Júnior Alcantara
Renan Pedra de Souza
Pedro Augusto Alves
Filipe Ferreira de Almeida Rego
Erna Geessien Kroon

Documento assinado eletronicamente por **Renan Pedra de Souza**, **Professor do Magistério Superior**, em 09/11/2023, às 22:08, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Pedro Augusto Alves**, **Usuário Externo**, em 10/11/2023, às 11:33, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Luiz Carlos Junior Alcantara**, **Usuário Externo**, em 12/11/2023, às 19:17, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Filipe Ferreira de Almeida Rego**, **Usuário Externo**, em 13/11/2023, às 11:26, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Erna Geessien Kroon**, **Professora do Magistério Superior**, em 14/11/2023, às 08:08, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2775004** e o código CRC **B12929FC**.

Referência: Processo nº 23072.267465/2023-03 SEI nº 2775004

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa de Pós-Graduação em Genética

FOLHA DE APROVAÇÃO

"Detecção e análise filogenética do *Chikungunya virus* e de *Orthoflavivirus* durante a vigilância de arbovírus no Brasil"

Joilson Xavier dos Santos Junior

Tese aprovada pela banca examinadora constituída pelos Professores:

Luiz Carlos Júnior Alcantara
FIOCRUZ

Renan Pedra de Souza
UFMG

Pedro Augusto Alves
FIOCRUZ-MG

Filipe Ferreira de Almeida Rego

Escola Bahiana de Medicina e Saúde Pública

Erna Geessien Kroon

UFMG

Belo Horizonte, 07 de novembro de 2023.

Documento assinado eletronicamente por **Renan Pedra de Souza**, **Professor do Magistério Superior**, em 09/11/2023, às 22:08, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Pedro Augusto Alves**, **Usuário Externo**, em 10/11/2023, às 11:33, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Luiz Carlos Junior Alcantara**, **Usuário Externo**, em 12/11/2023, às 19:18, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Filipe Ferreira de Almeida Rego**, **Usuário Externo**, em 13/11/2023, às 11:25, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Erna Geessien Kroon**, **Professora do Magistério Superior**, em 14/11/2023, às 08:08, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2775141** e o código CRC **74550D16**.

Referência: Processo nº 23072.267465/2023-03 SEI nº 2775141

AGRADECIMENTOS

Os meus sinceros agradecimentos

À toda minha família que tem oferecido todo o suporte imensurável para o meu desenvolvimento ao longo da minha vida.

Aos meus orientadores Dr. Luiz Alcantara e a Dra. Marta Giovanetti pela parceria, confiança depositada no meu trabalho e pelas inúmeras oportunidades oferecidas a mim ao longo desses últimos anos. Sou imensamente grato pelas contribuições que vocês tiveram na minha formação acadêmica.

Aos meus colegas de laboratório pela parceria e suporte durante a realização deste trabalho. Agradeço especialmente a Talita, Vagner, Carla, Hegger e Natalia pelos bons momentos proporcionados pela companhia de vocês no lab e nas nossas viagens pelo Brasil.

Aos meus amigos do “Achados e Perdidos”, cujo apoio, companhia e paciência marcaram positivamente a minha experiência durante o doutorado.

À equipe da Fundação Ezequiel Dias, em especial ao Felipe Iani, Alana, Ludmila, André e Adriana por me acolherem com toda a atenção e paciência no Laboratório de Virologia. A companhia e o apoio de vocês foram muito importantes para o desenvolvimento deste trabalho.

Aos professores e colaboradores do programa de Pós-Graduação em Genética da Universidade Federal de Minas Gerais.

Aos meus queridos amigos espalhados por esse grande Brasil. Obrigado pelo incentivo.

A todos aqueles que direta ou indiretamente contribuíram para a concretização deste trabalho.

RESUMO

O cenário de cocirculação de diferentes espécies de arbovírus que já causaram extensos surtos no Brasil, tem imposto constantes desafios ao sistema de saúde pública relacionados a identificação precisa e controle da disseminação de vírus epidêmicos. Diante deste contexto, com o objetivo de analisar a diversidade genética e as relações filogenéticas do *Chikungunya virus* (CHIKV) este trabalho utilizou a tecnologia de sequenciamento por nanoporos para gerar um total de 454 genomas do CHIKV a partir de amostras clínicas coletadas em 13 estados brasileiros entre 2021 e 2022. As análises filogenéticas revelaram a reintrodução da linhagem Leste-Central-Sul-Africano (ECSA) do CHIKV no estado do Rio Grande do Norte, destacando sua dispersão do Nordeste para o Centro-Oeste, e sugerindo a região Centro-Oeste como fonte de disseminação do CHIKV para o Paraguai. Os dados genômicos do CHIKV permitiram a reconstrução da história evolutiva atualizada da linhagem ECSA no Brasil e revelaram que a região Nordeste continua atuando como principal fonte de dispersão do CHIKV no Brasil, alimentando uma rede frequente de intercâmbio viral entre as regiões Nordeste e Sudeste. Este trabalho descreve a emergência de dois subclados da linhagem ECSA, o clado I, mais diverso, abrange amostras de 14 estados brasileiros, enquanto o clado II concentra-se na região Nordeste em 2022. A análise comparativa revelou a presença de variantes de nucleotídeo único entre as sequências de ambos os clados, incluindo as mutações E1-T98A e E2-V264A, previamente associadas ao aumento da adaptação do CHIKV a mosquitos *Aedes spp*. Além disso, este trabalho também descreve os resultados do desenvolvimento de um protocolo que integra a RT-PCR e o sequenciamento por nanoporos para a detecção multiplex do CHIKV e de vários vírus pertencente ao gênero *Orthoflavivirus*, possibilitando a identificação acurada de diferentes linhagens virais em uma única reação. A validação do protocolo com amostras clínicas positivas para DENV-1, DENV-2, CHIKV, YFV e ZIKV resultou na identificação correta das linhagens virais em 83% das amostras testadas.

Palavras-chave: *Chikungunya virus*; vigilância genômica; sequenciamento; epidemiologia genômica.

ABSTRACT

The scenario of co-circulation of different arbovirus species, that have already caused extensive outbreaks in Brazil, has imposed constant challenges to the public health system related to the precise identification and control of the spread of epidemic viruses. Given this context, in order to analyze the genetic diversity and phylogenetic relationships of the *Chikungunya virus* (CHIKV) this work used the nanopore sequencing technology to generate a total of 454 CHIKV genomes from clinical samples collected in 13 Brazilian states between 2021 and 2022. The phylogenetic analyzes revealed the reintroduction of the East-Central-South-African lineage (ECSA) of CHIKV in the state of Rio Grande do Norte, highlighting its dispersion from the Northeast to the Midwest, and suggesting the Midwest as a source of dissemination of CHIKV to Paraguay. The CHIKV genomic data allowed the reconstruction of the updated evolutionary history of the ECSA lineage in Brazil and revealed that the Northeast region continues to act as the main source of CHIKV dispersion in Brazil, fueling a frequent network of viral exchange between the Northeast and Southeast regions. This work also describes the emergence of two subclates of the ECSA lineage, clade I, more diverse, covers samples from 14 Brazilian states, while clade II is concentrated in the Northeast region in 2022. The comparative analysis revealed the presence of single nucleotide variants among the sequences of both clades, including mutations E1-T98A and E2-V264A, previously associated with increased adaptation of CHIKV to *Aedes spp.* mosquitoes. In addition, this work also describes the results of the development of a protocol that integrates RT-PCR and nanopore sequencing for the multiplex detection of CHIKV and several viruses belonging to the genus *Orthoflavivirus*, enabling the accurate identification of different viral lineages in a single reaction. Validation of the protocol with clinical samples positive for DENV-1, DENV-2, CHIKV, YFV and ZIKV resulted in the correct identification of viral lineages in 83% of the samples tested.

Keywords: *Chikungunya virus*; genomic surveillance; sequencing; genomic epidemiology.

LISTA DE ILUSTRAÇÕES

LISTA DE ABREVIATURAS E SIGLAS

A. aegypti	*Aedes aegypti*
ADAR	adenosine deaminase acting on RNA
AID/APOBEC	activation-induced cytidine deaminase/apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like
cDNA	complementary DNA
CHIKV	vírus chikungunya (*Chikungunya virus*)
COVID-19	coronavirus disease 2019
Ct	cycle threshold
ddNTP	Dideoxynucleotide triphosphates
DENV	vírus dengue (*Orthoflavivirus denguei*)
DNA	Ácido desoxirribonucleico
dNTP	Deoxynucleotide triphosphates
ECSA	East-Central-South-African
EEEV	vírus encefalite equina do Leste (*Eastern equine encephalitis virus*)
ELISA	Enzyme-linked immunosorbent assay
HIV/AIDS	human immunodeficiency virus/ Acquired immunodeficiency syndrome
IFA	Indirect Fluorescent Antibody
IgG	Imunoglobulina G
IgM	Imunoglobulina M
JEV	vírus da encefalite Japonesa (*Orthoflavivirus japonicum*)
LACEN	Laboratório Central de Saúde Pública
LACV	vírus La Crosse (*Orthobunyavirus lacrosseense*)
MAYV	vírus Mayaro (*Mayaro virus*)
MVEV	vírus da encefalite do Vale Murray (*Orthoflavivirus murrayense*)
NGS	Next-Generation Sequencing
NRIV	vírus Ngari (*Orthobunyavirus bunyamweraense*)
NS1	non-structural protein 1
ONNV	vírus O'nyong-Nyong (*o'nyong-nyong virus*)
OROV	vírus Oropouche (*Orthobunyavirus oropoucheense*)
PCR	Polymerase chain reaction

PRNT	Plaque-reduction neutralization test
RBD	receptor-binding domain
RNA	Ribonucleic acid
RRV	vírus do rio Ross (*Ross River virus*)
RT-qPCR	Reverse transcription-quantitative polymerase chain reaction
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2 (*Severe acute respiratory syndrome-related coronavirus*)
SCZ	Síndrome Congênita associada à infecção pelo vírus Zika
SINV	vírus Sindbis (*Sindbis virus*)
SLEV	vírus encefalite de Saint Louis (*Orthoflavivirus louisense*)
TBEV	vírus da encefalite transmitida por carrapatos (*Orthoflavivirus encephalitidis*)
USUV	vírus Usutu (*Orthoflavivirus usutuense*)
VEEV	vírus encefalite equina venezuelana (*venezuelan equine encephalitis virus*)
WEEV	vírus encefalite equina do Oeste (*Western equine encephalitis virus*)
WNV	vírus do Nilo Ocidental (*Orthoflavivirus nilense*)
YFV	vírus da febre amarela (*Orthoflavivirus flavi*)
ZIKV	vírus Zika (*Orthoflavivirus zikaense*)

SUMÁRIO

1 INTRODUÇÃO

Os vírus transmitidos por artrópodes (arbovírus) são causas frequentes de doenças no Brasil. As arboviroses (infecções causadas por arbovírus) podem ser consideradas doenças tropicais negligenciadas, pois tendem a afetar populações mais pobres de países tropicais e subtropicais de regiões subdesenvolvidas e recebem atenção inadequada em termos de financiamento para pesquisas (Bharadwaj *et al.*, 2021). Entretanto, o século XXI tem presenciado a rápida e preocupante reemergência destes vírus que está ligada a fatores como: adaptação dos vírus a novos vetores, crescimento da população mundial, urbanização de áreas de floresta, mudanças climáticas, globalização e facilidade de deslocamentos populacionais (Wilder-Smith *et al.*, 2017) (Figura 1).

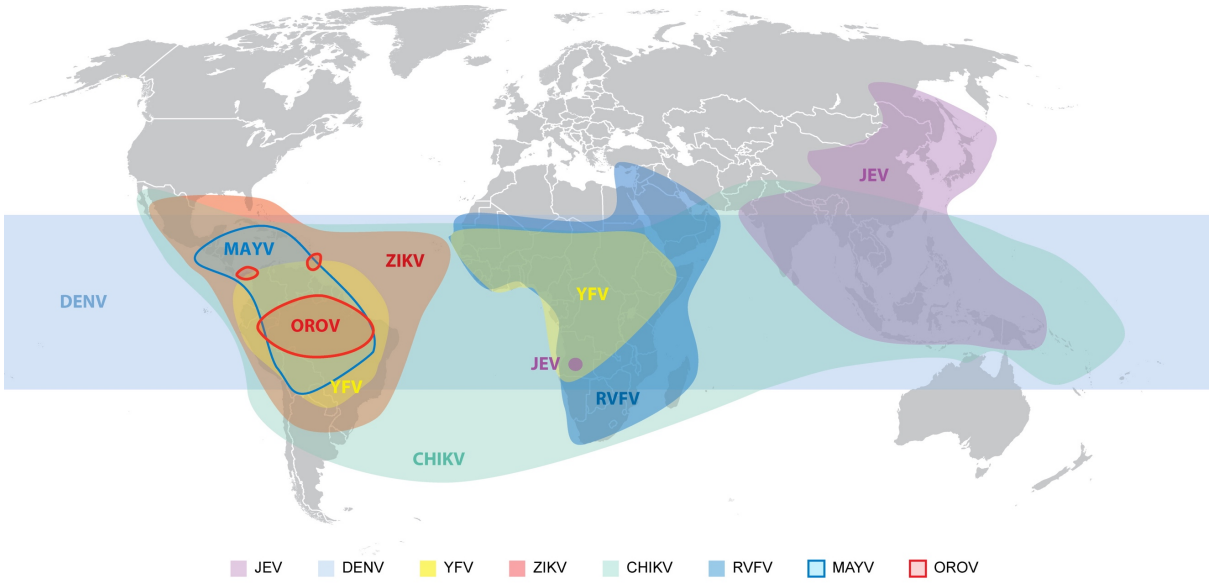


Figura 1. Mapa do mundo representa as distribuições espaciais relatadas para alguns arbovírus emergentes. Abreviaturas: CHIKV, Chikungunya virus; DENV, Dengue virus; JEV, vírus da encefalite japonesa; MAYV, vírus Mayaro; OROV, vírus Oropouche; RVFV, vírus da febre de Vale do Rift; YFV, vírus da febre amarela; ZIKV, vírus Zika. Adaptado de (Weaver *et al.*, 2018).

Os arbovírus são mantidos na natureza principalmente através da transmissão biológica entre hospedeiros vertebrados suscetíveis e artrópodes hematófagos infectados (Artsob, Lindsay and Drebot, 2017). Embora os mosquitos e os carrapatos

sejam considerados os vetores predominantes, os flebotomíneos também podem transmitir arbovírus (Young, 2018). Duas espécies de mosquitos, *Aedes aegypti* e *Aedes albopictus,* estão entre os vetores arbovirais mais proeminentes do mundo (Jones *et al.*, 2020). Esses dois mosquitos são vetores competentes de quatro arbovírus principais, sendo eles o vírus da dengue (*Orthoflavivirus denguei;* DENV), vírus chikungunya (*Chikungunya vírus*; CHIKV), vírus da febre amarela (*Orthoflavivirus flavi*; YFV), e o vírus Zika (*Orthoflavivirus zikaense*; ZIKV), que tem causado prejuízos à saúde e à economia em todo o mundo. O *A. aegypti*, vetor mais estudado, é encontrado principalmente nos trópicos e subtrópicos e a sua eficiência de transmissão é atribuída à sua alta susceptibilidade à infecção pelos quatro arbovírus mencionados anteriormente, à sua preferência de se alimentar a partir do sangue de seres humanos, à sua atividade diurna, à sua prosperidade em ambientes peridomésticos, e à sua capacidade de picar várias pessoas em um curto período de tempo (Wilder-Smith *et al.*, 2017). Por muitos anos, esses vetores e vírus têm sido endêmicos e restritos a regiões específicas, mas atualmente eles estão se espalhando para novas regiões tropicais, subtropicais e áreas temperadas, ampliando a cobertura global e criando uma situação que pode agravar epidemias em larga escala (Lwande *et al.*, 2020). A distribuição de mosquitos *Aedes spp.* atualmente é a mais ampla já registrada, se estendendo a todos os continentes, incluindo a América do Norte e Europa, com mais de 3 bilhões de pessoas vivendo em regiões infestadas por esta espécie. Tal expansão geográfica parece ser impulsionada pela globalização e pelas alterações antrópicas no meio ambiente (incluindo as mudanças climáticas) (Lwande *et al.*, 2020; Young, 2018).

Os mosquitos do gênero *Culex spp.* também podem atuar como vetores de arbovírus de importância médica para os animais, incluindo os humanos. Arbovírus emergentes e reemergentes transmitidos por *Culex spp.* incluem o vírus do Nilo Ocidental (*Orthoflavivirus nilense*; WNV), o vírus da encefalite Japonesa (*Orthoflavivirus japonicum*; JEV), o vírus encefalite de Saint Louis (*Orthoflavivirus louisense*; SLEV), o vírus Usutu (*Orthoflavivirus usutuense*; USUV) e o vírus da encefalite do Vale Murray (*Orthoflavivirus murrayense*; MVEV) (Gould *et al.*, 2017). Estes e outros vírus pertencentes ao sorocomplexo da encefalite Japonesa são geralmente dispersos via aves migratórias infectadas que atuam como hospedeiros amplificadores (Young, 2018). Dessa forma os padrões de dispersão geográfica desses arbovírus são influenciados pelas aves que eles infectam, cujas rotas

migratórias incluem longas distâncias (Gould *et al.*, 2017). No entanto, a introdução e dispersão de arbovírus podem também ser impulsionadas pelo movimento de pessoas por via aérea, por exemplo. Argumenta-se que a introdução do WNV na América do Norte ocorreu possivelmente através de um voo internacional carregando mosquitos infectados que partiu do Oriente Médio ou do Norte da África para Nova York em 1999 (Gould *et al.*, 2017, 2003).

O termo arbovírus não tem significado taxonômico, pois inclui vírus com diferentes características e pertencentes a diversas famílias virais (Artsob, Lindsay and Drebot, 2017). Existem mais de 500 espécies de arbovírus conhecidas e este número cresce à medida que surgem novos estudos utilizando as modernas tecnologias de sequenciamento de DNA para a caracterização genética de novos vírus. Estima-se que apenas 1% da diversidade total de arbovírus do mundo é conhecida, sendo apenas 150 espécies conhecidas por causar doenças em seres humanos (Young, 2018).

1.1 Principais famílias de arbovírus registrados no Brasil

A família *Flaviviridae* inclui pequenos vírus envelopados com genoma de RNA não-segmentado de sentido positivo de aproximadamente 9-13 mil bases (Rico-Hesse et al., 2023). Até o momento, estão identificadas 89 espécies distribuídas em quatro gêneros: *Orthoflavivirus* (anteriormente denominado *Flavivirus*) (Postler *et al.*, 2023), *Pestivirus, Hepacivirus e Pegivirus.* O gênero *Orthoflavivirus* inclui 53 espécies, cuja maioria é transmitida a hospedeiros vertebrados (mamíferos e aves) por mosquitos ou carrapatos, nos quais o vírus se replica ativamente (Rico-Hesse et al., 2023). Os ortoflavivírus podem ser considerados o grupo mais importante do ponto de vista médico-epidemiológico, pois incluem uma variedade de espécies causadoras de doenças em humanos de preocupação para a saúde global, nas quais observam-se manifestações clínicas variáveis, desde febre e exantema até encefalite (Artsob, Lindsay and Drebot, 2017). Entre os vírus mais conhecidos estão DENV, WNV, ZIKV e YFV (Mayer, Tesh and Vasilakis, 2017). Entre os outros membros de relevância médica, por causarem doenças em humanos e animais domésticos ou selvagens, estão o JEV, o vírus da encefalite transmitida por carrapatos (*Orthoflavivirus*

encephalitidis; TBEV), o SLEV, o USUV e o MVEV (Artsob, Lindsay and Drebot, 2017; Mayer, Tesh and Vasilakis, 2017).

A família *Togaviridae*, por sua vez*,* inclui pequenos vírus envelopados com genoma de RNA não-segmentado, fita-simples de sentido positivo compreendendo entre 10-12 mil bases nitrogenadas (Chen *et al.*, 2018). A família inclui apenas um gênero, *Alphavirus*, que contém mais de 30 espécies de arbovírus com distribuição geográfica ampla, presente em todos os continentes, exceto Antártica, e que causam doenças em seres humanos e animais (Chen *et al.*, 2018; Powers *et al.*, 2001). Entre os vírus causadores de encefalite destacam-se o vírus encefalite equina venezuelana (*Venezuelan equine encephalitis vírus*; VEEV), vírus encefalite equina do Leste (*Eastern equine encephalitis virus*; EEEV) e o vírus encefalite equina do Oeste (*Western equine encephalitis virus*; WEEV). As manifestações clínicas como artralgia e febre alta estão frequentemente presentes em infecções causadas pelos vírus do rio Ross (*Ross River virus*; RRV), vírus o'nyong-nyong (*Onyong-nyong virus*; ONNV), vírus Mayaro (*Mayaro virus*; MAYV), vírus Sindbis (*Sindbis virus*; SINV) e o CHIKV. Além disso, doenças neurológicas também têm sido associadas à infecção por CHIKV (Acevedo *et al.*, 2017; Mehta *et al.*, 2018; Schuffenecker *et al.*, 2006; Vieira *et al.*, 2018).

A família *Peribunyaviridae* compreende vírus ecologicamente diversos que podem ser encontrados em todos os continentes, exceto na Antártica, e que são mantidos na natureza em ciclos de transmissão vertebrados-artrópodes (Hughes *et al.*, 2020). Os peribunyavírus apresentam um genoma segmentado com três moléculas de RNA de fita simples negativa com tamanho que varia entre 11,2−12,5 mil bases (Hughes *et al.*, 2020). A família *Peribunyaviridae* inclui 97 espécies divididas em quatro gêneros distribuídos globalmente: *Orthobunyavirus, Herbevirus, Pacuvirus e Shangavirus*. O gênero *Orthobunyavirus* é o maior e mais diverso dessa família, incluindo 88 espécies, que tem isolados a partir de muitos hospedeiros vertebrados diferentes, incluindo vírus La Crosse (*Orthobunyavirus lacrosseense*; isolado em esquilos), vírus Mojuí dos Campos (*Orthobunyavirus nyandoense*; isolado em morcegos), vírus da lebre da neve (*Orthobunyavirus khatangaense*; isolado em coelhos), vírus Akabane (*Orthobunyavirus akabaneense*; isolado em ungulados), vírus Oropouche (*Orthobunyavirus oropoucheense*; isolado em preguiças) e vírus Mermet (*Orthobunyavirus mermetense*; isolado em pássaros) (Hughes *et al.*, 2020). Infecções por ortobunyavírus têm sido associadas a doenças do sistema nervoso central em

humanos e outros vertebrados, tais como a encefalite causada pelo vírus La Crosse, a doença febril aguda causada pelo vírus Oropouche, e febre hemorrágica causada pelo vírus Ngari (*Orthobunyavirus bunyamweraense*; NRIV) (Edridge and Hoek, van der, 2020; Elliott, 2014).

1.2 Epidemiologia dos principais arbovírus circulantes no Brasil

No atual cenário epidemiológico brasileiro, os arbovírus mais comuns são DENV, CHIKV, ZIKV e YFV, embora existam outros arbovírus com potencial de disseminação no país (Donalisio, Freitas and Zuben, 2017). O DENV sorotipo 1 e 4 foram identificados pela primeira vez no Brasil no estado de Roraima após um surto registrado na região amazônica entre 1981 e 1982 (Osanai et al., 1983). No entanto, infecções pelo DENV apenas e se tornaram um problema nacional em saúde publica após a emergência do sorotipo 1 em grandes comunidades urbanas no estado do Rio de Janeiro em 1986 (Schatzmayr, Nogueira and Travassos da Rosa, 1986). Quatro anos depois, em 1990, o sorotipo 2 do DENV foi isolado também no estado do Rio de Janeiro após um surto registrado naquela região, o que agravou a situação da dengue no país com o aparecimento de casos graves e fatais associados a infecções secundárias (Nogueira *et al.*, 1993). Entre 1990 e 2022, o Brasil enfrentou diversas epidemias de dengue, caracterizadas por um padrão cíclico, totalizando mais de 18,5 milhões de casos notificados (Andrioli, Busato and Lutinski, 2020; Pan American Health Organization, 2023a). Destes, mais de 14,5 milhões de casos ocorreram na última década (2012 a 2022) (Andrioli, Busato and Lutinski, 2020; Pan American Health Organization, 2023a). A incidência da dengue tem aumentado nos últimos anos no Brasil, uma vez que em 2014 a incidência foi estimada em 292,5 casos por 100.000 habitantes e em 2019 a incidência aumentou para 1,078.48/100.000 até atingir a taxa de 1,104.47 casos por 100.000 habitantes em 2022, quando o país registrou uma grande epidemia com mais de 2,3 milhões de casos notificados que corresponderam a mais da metade dos casos de dengue na região das Américas (Pan American Health Organization, 2023a; e). No Brasil a dengue é endêmica e epidêmica, com circulação simultânea dos quatro sorotipos virais já descritos: DENV-1, DENV-2, DENV-3 e DENV-4. Na mais recente epidemia ocorrida em 2019, DENV-1 e DENV-2 predominaram entre os casos reportados, com o sorotipo 1 dominante na região Nordeste, e o sorotipo 2 dominante nas regiões Centro-Oeste e Sudeste (Adelino *et*

al., 2021). Análises da evolução dos casos de dengue entre os anos 2001 e 2012 indicaram um maior número de casos de dengue concentrado na costa atlântica e no interior do estado de São Paulo (Rodrigues *et al.*, 2016). Além disso, as regiões Sudeste e Norte do Brasil parecem ter desempenhado um papel relevante para a dispersão dos sorotipos 1 e 2 durante a epidemia de 2019 (Adelino *et al.*, 2021).

Historicamente, o arbovírus que tem recebido maior atenção devido o seu significado para a saúde humana é o vírus da febre amarela, um vírus que circula entre humanos e primatas não-humanos em um ciclo silvestre em partes da África e da América do Sul (Artsob, Lindsay and Drebot, 2017). Ocasionalmente, o vírus pode estabelecer um ciclo urbano no qual os humanos atuam como o único hospedeiro amplificador após transmissão mediada por mosquitos *A. aegypti* e *A. simpsoni* (Artsob, Lindsay and Drebot, 2017). No Brasil, o YFV é encontrado em surtos selvagens associados a epizootias, e esporadicamente causa surtos em humanos, devido ao fenômeno de transbordamento (*spillover* em inglês) (Donalisio, Freitas and Zuben, 2017). A última epidemia de febre amarela urbana no Brasil foi relatada em 1929 no estado do Rio de Janeiro e o último caso urbano foi documentado em 1942 no estado do Acre (Silva *et al.*, 2020). No entanto, tem sido observada uma expansão geográfica da circulação do YFV indicada pela reemergência extra-amazônica do vírus nas regiões Centro-Oeste, Sudeste e Sul desde 2014 (Donalisio, Freitas and Zuben, 2017; Ministério da Saúde do Brasil, 2020). Entre 2016 e 2019, o Brasil registrou mais de 2 mil casos confirmados de febre amarela em humanos, com 700 óbitos confirmados (Pan American Health Organization, 2021). Análises de epidemiologia genômica indicaram que o surto iniciado em 2016 apresentava características de transmissão silvestre do YFV que estava se expandindo em direção à áreas anteriormente livres de febre amarela (Faria, N R *et al.*, 2018). Outro estudo sugere que os últimos surtos ocorridos na região sudeste envolveram diversos eventos de introdução do vírus nos estados do Rio de Janeiro e Espírito Santo entre 2016 e 2019 (Giovanetti *et al.*, 2019). Ademais, novos dados apresentados por um recente estudo sugerem a existência de um corredor de dispersão do YFV entre as regiões Norte, Centro-Oeste e Sudeste, além da existência de áreas de risco contendo casos de febre amarela em humanos e baixa cobertura vacinal localizadas próximo a grande populações da região Sudeste (Figura 2) (Giovanetti *et al.*, 2023).

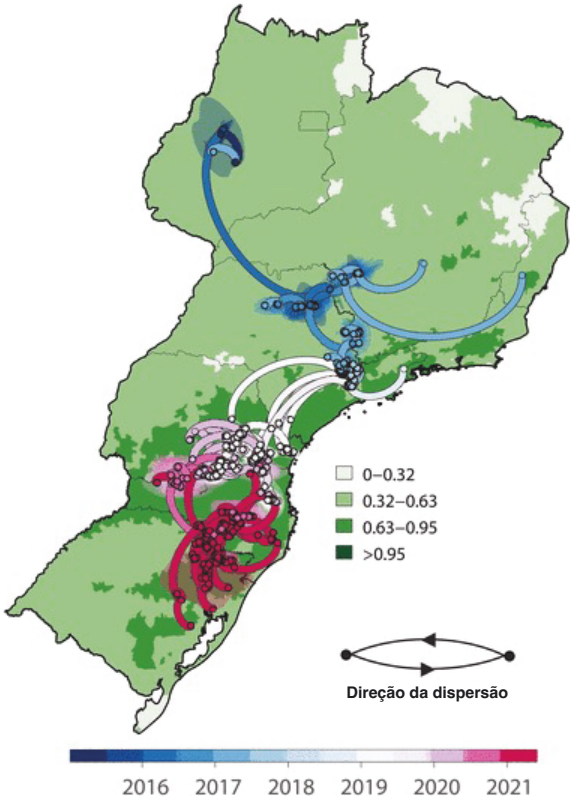


Figura 2. Distribuição espaço-temporal do YFV em diferentes estados brasileiros a partir de uma reconstrução filogeográfica. Linhas curvas sólidas denotam conexões e a direcionalidade da dispersão. As cores das linhas indicam o tempo estimado e as cores do mapa representam diferenças na cobertura florestal. Adaptado de (Giovanetti *et al.*, 2023).

Antes de chegar nas Américas, o vírus Zika, por sua vez, passou por um processo recente de dispersão por vários países, iniciado em 2007 na Micronésia e passando por países do Pacífico e Ásia (Faria, Azevedo, *et al.*, 2016; Faria, Sabino, *et al.*, 2016). Em 2015, o Brasil foi o primeiro país a reportar um surto de doença exantemática causada pelo ZIKV (Fantinato *et al.*, 2016; Zanluca *et al.*, 2015). A doença atingiu mais de 6 mil pessoas na região Nordeste e desde a introdução do vírus mais de 508 mil casos foram notificados em todo o Brasil, o que corresponde a 52% de todos os casos de Zika registrados nas Américas (Pan American Health Organization, 2023b). Os surtos causados pelo ZIKV ocorridos no Brasil de 2014 a 2016 provavelmente cessaram devido à imunidade de rebanho, porém, em vez de desaparecer o vírus ainda circula com taxas baixas em áreas que foram intensamente afetadas, como regiões metropolitanas no Nordeste do país (Magalhaes *et al.*, 2020). Desde o estabelecimento da correlação positiva entre a infecção pelo ZIKV e os casos

de microcefalia, mais de 1,8 mil casos de Síndrome Congênita associada à infecção pelo vírus Zika (SCZ) foram confirmados entre os anos de 2015 e 2022 (Brasil *et al.*, 2016; Cauchemez *et al.*, 2016; Ministério da Saúde do Brasil, 2023a). Apesar do encerramento do estado de emergência em saúde pública, novos casos de SCZ continuam ocorrendo no Brasil principalmente na região Nordeste que apresentou os maiores números de casos notificados entre 2018 e 2022 (Ministério da Saúde do Brasil, 2023a).

Outro arbovírus com destaque no Brasil é o CHIKV. Análises filogenéticas revelaram que o CHIKV apresenta quatro genótipos (ou linhagens), nomeadas de acordo com sua distribuição geográfica, sendo eles o genótipo África Ocidental, o Asiático, o Leste-Central-Sul-Africano (ECSA), e o Oceano Índico (Figura 3) (Powers, 2011; Powers *et al.*, 2000; Schuffenecker *et al.*, 2006). No Brasil, os primeiros casos autóctones causados pela linhagem Asiática foram registrados em setembro de 2014 no município de Oiapoque, estado do Amapá, na região norte do país (Nunes *et al.*, 2015). Naquele mesmo mês também foi identificado pela primeira vez no Brasil a presença da linhagem ECSA na cidade de Feira de Santana, Bahia, introduzida por um viajante infectado que retornou da Angola. A identificação da transmissão autóctone do genótipo ECSA na Bahia sinalizou a emergência e o estabelecimento deste genótipo no país (Nunes *et al.*, 2015). Desde então, o genótipo ECSA tem se espalhado para outros estados brasileiros das regiões Nordeste e Sudeste, exercendo uma ameaça à saúde pública, especialmente devido à cocirculação de outros arbovírus, tais como os vírus dengue e Zika, que causam manifestações clínicas semelhantes (Charlys da Costa *et al.*, 2017; Cunha, M. D. P. *et al.*, 2017; Cunha, M. S. *et al.*, 2017; Lessa-Aquino *et al.*, 2018; Tanabe *et al.*, 2018). Desde a introdução do vírus há 9 anos, o Brasil tem acumulado um total de mais de 1,5 milhão de casos notificados da febre chikungunya, dois quais 265 mil casos ocorreram apenas em 2022 (Pan American Health Organization, 2023c).

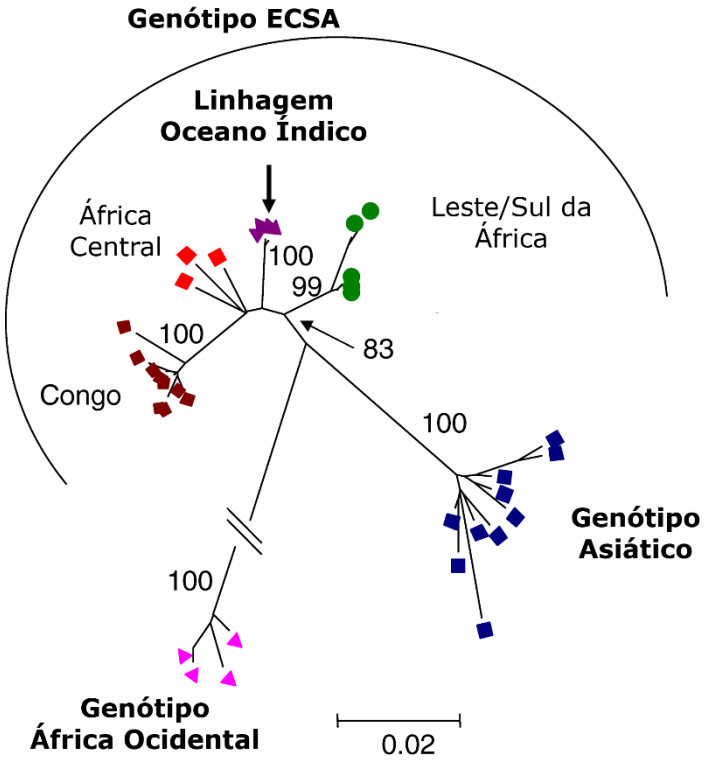


Figura 3. Filogenia das linhagens do *Chikungunya virus*. A árvore filogenética mostra as relações de parentesco evolutivo entre isolados representantes das quatro linhagens do CHIKV, destacando a descendência da linhagem Oceano Índico a partir do genótipo ECSA. Números ao longo dos ramos indicam os valores de reamostragem por bootstrap. Adaptado de (Schuffenecker *et al.*, 2006).

No cenário epidemiológico brasileiro atual, outros arbovírus com potencial de disseminação representam uma ameaça real de circulação epidêmica no país. O WNV que se expandiu rapidamente nas Américas, tem sido observado no Brasil a partir de evidências sorológicas da circulação do vírus em equinos e aves das regiões do pantanal e do Nordeste em 2009 e 2010 (Melandri *et al.*, 2012; Pauvolid-Corrêa *et al.*, 2011; Silva *et al.*, 2013). O primeiro caso de infecção pelo WNV no Brasil foi confirmado por sorologia em amostras de um homem residente em uma área rural do estado do Piauí em 2014. Recentemente, também em uma área rural no estado do Piauí, foi confirmado outro caso de febre do Nilo Ocidental caracterizado por encefalite aguda (Pan American Health Organization, 2023d; Vieira *et al.*, 2015). Em 2019 foi publicado o primeiro genoma completo do WNV isolado no Brasil a partir de amostras do sistema nervoso central coletadas de um cavalo no estado do Espírito Santo em 2018 (Martins *et al.*, 2019). Análises filogenéticas indicaram que este isolado pertencia

a linhagem 1a e estava proximamente relacionado a isolados da América do Norte, México, Colômbia, e Argentina (Martins *et al.*, 2019). Recentemente, novos genomas do WNV foram gerados a partir de amostras coletadas nos anos 2018, 2019, e 2020 em equinos infectados oriundos dos estados de Minas Gerais, São Paulo e Piauí (Azevedo Costa *et al.*, 2021).

Além do WNV, outro arbovírus com potencial epidêmico é o OROV (*Orthobunyavirus*) que tem sido frequentemente identificado na região amazônica em pacientes que apresentaram uma doença febril e exantemática aguda (Nascimento *et al.*, 2020). O OROV foi identificado pela primeira vez em 1960 em mosquitos e em um bicho preguiça próximos a estrada Belém-Brasília. Desde então o vírus já foi identificado em vários estados da região Norte e Nordeste (Nascimento *et al.*, 2020; Pinheiro et al., 2002; Vasconcelos *et al.*, 2011). Em 2019 foram detectados 26 casos de infecção pelo OROV em quatro estados da região Norte (Ministério da Saúde do Brasil, 2019). Ainda na região Norte do Brasil, encontra-se também o MAYV, um alfavírus que causa um quadro febril exantemático semelhante ao da dengue e que foi identificado pela primeira vez em 1955 no estado do Pará, onde também foi identificado um surto em 1978 com 55 casos humanos confirmados (Causey and Maroja, 1957; Pinheiro *et al.*, 1981). Desde então, surtos causados pelo MAYV foram registrados nos anos 1991, 2008 e 2016 nas regiões Norte e Centro-Oeste (Esposito and Fonseca, 2017). Recentemente, em 2019, no estado do Pará foram confirmados 5 casos de infecção pelo MAYV, com um caso de hospitalização infantil (Ministério da Saúde do Brasil, 2019).

1.3 Diagnóstico laboratorial das infecções por arbovírus

Embora a maioria das infecções arbovirais sejam assintomáticas ou autolimitadas, o diagnóstico correto é importante devido ao risco potencial de complicações como síndrome congênita por ZIKV, febre hemorrágica por DENV e manifestações neurológicas (Kerkhof *et al.*, 2020). O diagnóstico das infecções arbovirais pode ser realizado através da detecção direta do vírus por métodos de cultura viral, captura de antígenos virais ou detecção do ácido nucleico. Esses métodos de detecção direta requerem que as amostras sejam coletadas na fase aguda da doença, até o quinto ou sétimo dia após o início dos sintomas, quando há alta viremia (Young *et al.*, 2014). A maioria dos arbovírus pode ser isolada a partir de

amostras de sangue ou tecido inoculadas em cultura de células primárias ou linhagens celulares, onde a presença de efeitos citopáticos característicos, observados após alguns dia de incubação, são indicativos da infecção viral e ajudam na determinação do gênero e da família do vírus (Hunsperger, 2015).

Os antígenos virais podem ser detectados diretamente em amostras de sangue por meio do ensaio de imunoabsorção enzimática (ELISA - *enzyme-linked immunosorbent assay*) que tem sido frequentemente utilizado para o diagnóstico de dengue através da detecção da proteína não-estrutural 1 (NS1), uma glicoproteína expressa pelos ortoflavivírus e essencial para a replicação viral (Hunsperger, 2015; Kerkhof *et al.*, 2020). No entanto, esse método de detecção de antígenos tem apresentado baixa sensibilidade em infecções secundárias por DENV e apresenta pouca disponibilidade comercial para outros ortoflavivírus além de DENV e ZIKV (Kerkhof *et al.*, 2020).

Os testes de amplificação de ácido nucléico, por sua vez, permitem diagnosticar infecções por arbovírus com mais especificidade e sensibilidade (58.9–100%) quando comparado com outros métodos de diagnóstico (Lanciotti and Roehrig, 2016; Tang and Ooi, 2012). O teste molecular mais frequentemente utilizado é a RT-qPCR que combina a amplificação da reação em cadeia da polimerase pós transcrição reversa (RT-PCR) com o uso de sondas vírus-específicas marcadas com moléculas fluorescentes capazes de detectar o DNA amplificado durante a reação (Hunsperger, 2015; Lanciotti and Roehrig, 2016). O uso de sondas marcadas, que são quantificadas em tempo real, permite que múltiplos alvos sejam detectados ao mesmo tempo e na mesma reação. Dessa forma, os testes RT-qPCR têm apresentado sucesso em detectar os diferentes sorotipos de DENV a partir de uma única reação (Lanciotti and Roehrig, 2016).

Os tipos de testes usados para o diagnóstico de arbovírus com base na sorologia são os ensaios de ELISA, o ensaio de imunofluorescência (IFA) e o teste de neutralização por redução de placa (PRNT). Em contraste com o ELISA e o IFA que capturam anticorpos direcionados aos antígenos virais em meio a uma diversidade de anticorpos gerados em resposta à infecção, o PRNT é direcionado apenas aos anticorpos neutralizantes presentes em uma amostra (Kerkhof *et al.*, 2020). O PRNT, apesar de ser um teste laborioso e que demanda tempo, é considerado um teste confirmatório padrão ouro que permite quantificar o grau em que a infecção celular é inibida por anticorpos neutralizantes presentes no soro ou no líquido cefalorraquidiano

(Piantadosi and Kanjilal, 2020). Dessa forma, os resultados do PRNT fornecem uma estimativa direta da quantidade de anticorpos funcionais e são valiosos não apenas para confirmar a infecção recente, mas também para avaliar a eficácia de vacinas (Piantadosi and Kanjilal, 2020).

A detecção de imunoglobulinas fornece uma janela mais ampla de oportunidade para o diagnóstico, porque os anticorpos permanecem detectáveis no sangue por períodos mais longos (Imrie *et al.*, 2007; Tang and Ooi, 2012). No entanto, a presença de IgM e IgG deve ser interpretada cuidadosamente dependendo do contexto epidemiológico, pois numa típica infecção primária por ortoflavivírus o IgM é detectado entre 6 a 14 dias após infecção, e seus níveis declinam durante os próximos 2-3 meses. Em alguns casos os níveis de IgM persistem por longos períodos, podendo limitar seu uso na diferenciação entre uma infecção aguda ou passada recente (Kerkhof *et al.*, 2020; Ohst *et al.*, 2018). A detecção de IgG, por sua vez, acontece dentro de alguns dias após infecção primária. Todavia, durante uma infecção secundária os níveis de IgG aumentam rapidamente nos dois primeiros dias após infecção, tornando a interpretação dos resultados e, consequentemente o diagnóstico, mais complicado (OHST et al., 2018).

Apesar dos testes sorológicos serem comumente utilizados na rotina de diagnóstico de infecções por arbovírus, a existência de reação cruzada entre os anticorpos dentro dos principais subgrupos de arbovírus é um desafio, pois pode resultar em um diagnóstico incorreto (Piantadosi and Kanjilal, 2020; Young *et al.*, 2014). Anticorpos de alfavírus apresentam reação cruzada limitada e os testes padrão geralmente são suficientes para identificar o vírus infectante. No entanto, os anticorpos para os diferentes ortoflavivírus geralmente reagem de forma cruzada, de modo que a detecção de IgM e/ou IgG para um arbovírus nos testes de rotina não é evidência definitiva de infecção específica por aquele vírus (Young *et al.*, 2014).

1.4 Uso da tecnologia de sequenciamento de DNA para a identificação e caracterização de vírus

Avanços no desenvolvimento da tecnologia de sequenciamento de DNA permitiu a produção rápida de grandes quantidades de dados genômicos que têm sido utilizados para a identificação e caracterização de agentes infeciosos (Wilson, Musicha and Beale, 2023; Yin, Butler and Zhang, 2021). Os diferentes métodos de

sequenciamento podem ser classificados de acordo com a quantidade de dados de sequência geradas. O sequenciamento de primeira geração caracterizado pelo método de terminação de cadeia, também conhecido como sequenciamento Sanger, baseia-se na incorporação de didesoxinucleotídeo trifosfato (ddNTP) durante a síntese *in vitro* de DNA (Sanger, Nicklen and Coulson, 1977). Os ddNTPs incorporados não permitem a extensão da molécula de DNA, produzindo fragmentos de DNA com diferentes tamanhos que são, em seguida, separados por eletroforese capilar. Ao fim, a sequência de nucleotídeos é determinada a partir da detecção da fluorescência emitida pelos ddNTPs marcados com um corante fluorescente (Marziali and Akeson, 2001; Shendure *et al.*, 2017). Apesar do baixo rendimento, o sequenciamento Sanger possibilitou o sequenciamento de grandes genomas procarióticos e eucarióticos, incluindo o primeiro genoma humano publicado em 2001 (C. elegans Sequencing Consortium, 1998; International Human Genome Sequencing Consortium *et al.*, 2001).

A partir de 2005 as primeiras plataformas de sequenciamento, coletivamente denominadas de Sequenciamento de Nova Geração (NGS, do inglês *next generation sequencing*), aparecem como resultados de avanços tecnológicos que permitiram o sequenciamento paralelo de milhões a bilhões de moléculas de DNA em uma única reação (Shendure *et al.*, 2017). O sequenciamento NGS promoveu o aumento da capacidade de geração de dados a custos significativamente reduzidos (Yin, Butler and Zhang, 2021). As plataformas Illumina e Ion Torrent são atualmente as plataformas de NGS mais populares no mercado e ambas utilizam o método de sequenciamento por síntese, no qual há a amplificação *in vitro* de bibliotecas de fragmentos de DNA seguida da detecção da incorporação de nucleotídeos (Goodwin, McPherson and McCombie, 2016; Mardis, 2017). Mais especificamente, o sequenciamento no Ion Torrent utiliza sensores para detectar íons de hidrogênio liberados durante a incorporação de cada nucleotídeo à molécula de DNA sendo sintetizada. A Illumina, por sua vez, utiliza a leitura ótica da fluorescência emitida durante ciclos de incorporação de desoxinucleotídeo trifosfato (dNTPs) de terminação de cadeia marcados com um fluoróforo para determinar a sequência de nucleotídeos do DNA sequenciado (Goodwin, McPherson and McCombie, 2016; Hu *et al.*, 2021).

As plataformas Illumina e Ion Torrent compreendem uma abordagem geradora de sequências curtas entre 150 a 400 bases que pode ser limitante para o sequenciamento de regiões homólogas ou repetitivas que requerem o alinhamento correto das sequências (Mardis, 2017). Para lidar com tais limitações, novas

abordagens foram desenvolvidas para dar origem a plataformas de sequenciamento de molécula única capazes de gerar longas sequências da ordem de centenas a milhares de bases (Shendure *et al.*, 2017). A abordagem empregada pela PacBio baseia-se no registro ótico da síntese de DNA em tempo real mediada por uma única DNA polimerase fixada em nanoestruturas captadoras de sinais fluorescentes. Uma única molécula de DNA molde presente em cada nanoestrutura é sequenciada a medida que um nucleotídeo marcado é incorporado, liberando um sinal fluorescente correspondente a base nitrogenada adicionada (Goodwin, McPherson and McCombie, 2016). Uma segunda abordagem, empregada pela Oxford Nanopore em 2014, utiliza um princípio diferente ao detectar diretamente a composição de uma molécula de ácido nucleico (DNA ou RNA) de fita simples que passa através de um poro proteico que atua como biosensor (Figura 4). As disrupções na corrente iônica, gerada durante a passagem de uma molécula, produz padrões de corrente elétrica correspondentes a sequência nucleotídica que são identificados por sensores e decodificados por algoritmos (Yin, Butler and Zhang, 2021). Pesando aproximadamente 87 gramas, o sequenciador portátil da nanopore, MinION, possui um *chip* contendo mais de 2 mil poros que tem permitido a geração de sequências ultralongas de mais de 2,2 milhões de bases (Payne *et al.*, 2019).

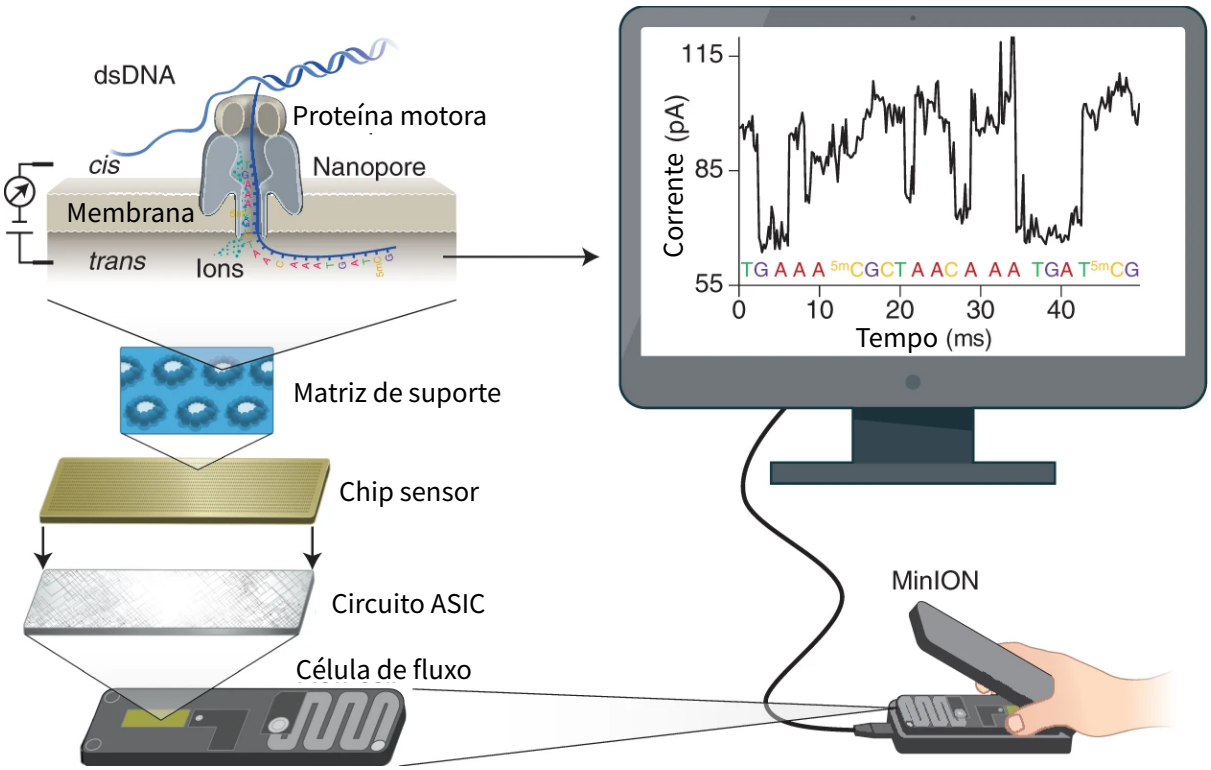


Figura 4. Princípio do sequenciamento por nanoporos. Os nanoporos são proteínas canais inseridas em uma membrana apoiada por uma matriz de suporte conectada ao um chip sensor. Os canais são controlados individualmente por circuitos ASIC presentes na célula de fluxo. Uma voltagem aplicada sobre a membrana cria uma corrente iônica que é interrompida quando uma molécula de ácido nucleico passa através do nanoporo do lado *cis* carregado negativamente para o lado *trans* carregado positivamente. Essas interrupções na corrente elétrica são registradas e decodificadas na sequência de bases nitrogenadas. Adaptado de (Wang *et al.*, 2021).

Embora as plataformas de sequenciamento de molécula única sejam capazes de gerar dados de forma mais rápida, as longas sequências geradas por essas plataformas apresentam taxas de erros (aproximadamente 14% durante os primeiros anos) maiores que as taxas das plataformas Ilumina e Ion Torrent, cujas sequências possuem uma acurácia acima de 99,9% (Hu *et al.*, 2021; Sahlin and Medvedev, 2021). No entanto, avanços tem sido feitos para melhorar a qualidade das sequências geradas pelas plataformas da Nanopore e PacBio que quando combinados com abordagens computacionais personalizadas tem permitido a geração de sequências com maiores porcentagens de acurácia (99% para Nanopore e 99.8% para PacBio) (Sahlin and Medvedev, 2021; Sereika *et al.*, 2022; Wan *et al.*, 2022; Wenger *et al.*, 2019).

Os avanços observados ao longo do desenvolvimento da tecnologia de sequenciamento acompanharam também a aplicação dessa tecnologia em diversas áreas de pesquisa, incluindo descoberta de novas espécies, análise evolutiva e genômica comparativa, investigações forenses e caracterização genética do câncer no contexto da medicina de precisão (Goodwin, McPherson and McCombie, 2016; Nogrady, 2020; Shendure *et al.*, 2017). No âmbito da pesquisa de monitoramento e caracterização de agentes infecciosos, tais como os arbovírus, o sequenciamento genômico tem emergido como uma abordagem promissora para o diagnóstico de doenças infecciosas através da identificação baseada na sequência de ácido nucléico dos agentes etiológicos (Lanciotti and Roehrig, 2016; Miao *et al.*, 2018). A diminuição dos custos, a alta sensibilidade e a velocidade da geração de dados permitem que tecnologia de sequenciamento forneça informações relevantes para a microbiologia clínica praticamente em tempo real, desde a testagem de fenótipos até o rastreamento de surtos (Balloux *et al.*, 2018).

Aliado à tecnologia de sequenciamento de equipamentos portátil como o MinION, o monitoramento genômico de agentes virais é capaz de produzir grandes quantidades de dados em um tempo relativamente curto (de 24h a 48h) que podem revelar aspectos epidemiológicos relevantes sobre a dinâmica de uma epidemia (Gardy and Loman, 2018; Lewandowski *et al.*, 2019). Isso é possível porque a análise de sequências genômicas virais, por meio de ferramentas de bioinformática, permite avaliar a diversidade genética da população viral, reconstruir as origens das epidemias, estimar taxas de transmissão e gerar informações genéticas para o desenvolvimento e aprimoramento de vacinas, novos medicamentos e diagnósticos moleculares (Moratorio and Vignuzzi, 2018; Pybus, Tatem and Lemey, 2015).

A epidemiologia genômica, ao combinar dados genômicos de patógenos virais com metadados individuais, tem desempenhado um papel de destaque na luta contra doenças infecciosas emergentes e reemergentes (Hill *et al.*, 2021; Houldcroft, Beale and Breuer, 2017). O emprego da tecnologia de sequenciamento durante a epidemia de Ebola no Oeste africano entre 2014 e 2016 representou o primeiro uso de genomas virais em larga escala para identificar as origens e disseminação de um patógeno durante uma epidemia (Gire *et al.*, 2014; Hill *et al.*, 2021; Quick *et al.*, 2016). Desde então, o uso de sequenciadores portáteis como o MinION, tem permitido a implementação rápida de esforços de sequenciamento para a identificação e caracterização da diversidade viral durante eventos epidêmicos. Após a emergência

do ZIKV nas américas em 2014, o MinION foi empregado pela primeira vez no Brasil para gerar 54 genomas que ajudaram a esclarecer a origem e padrão da transmissão críptica do ZIKV no país (Faria *et al.*, 2017; Faria, Sabino, *et al.*, 2016). Posteriormente, esta mesma abordagem foi aplicada nos estudos de vigilância genômica de outros vírus causadores de epidemias no Brasil, como durante a transmissão selvática do YFV na região sudeste entre 2016 e 2017 (Faria, Nuno Rodrigues *et al.*, 2018), o surto do CHIKV na região amazônica em 2018 (Naveca *et al.*, 2018), a epidemia de dengue em 2019 (Adelino *et al.*, 2021), além da emergência no Brasil do coronavírus da síndrome respiratória aguda grave 2 (*Severe acute respiratory syndrome-related coronavírus*; SARS-CoV-2) em 2020 e vírus monkeypox (*Monkeypox virus*) em 2022 (Candido *et al.*, 2020; Claro *et al.*, 2022; Jesus *et al.*, 2020).

2 JUSTIFICATIVA

As doenças infecciosas causadas por vírus emergentes e reemergentes, como DENV, ZIKV, CHIKV e mais recentemente o SARS-CoV-2, tem impactado a saúde, a economia e o modo de vida de bilhões de pessoas ao redor do mundo nos últimos dez anos. Somente as doenças transmitidas por vetores representam mais de 17% de todas as doenças infecciosas registradas no mundo, causando mais de 700 mil óbitos anualmente. Por exemplo, estima-se que 3,8 bilhões de pessoas vivem atualmente em áreas de risco de infecção pelo vírus da dengue.

O contexto brasileiro torna-se preocupante devido a circulação simultânea de várias espécies virais causadoras de surtos sazonais e que muitas vezes resultam em doenças de difícil diagnóstico clínico devido a presença de síndromes febris clinicamente indistinguíveis. Além disso, há a circulação de outros vírus com potencial epidêmico, porém menos conhecidos pela população geral, como o vírus do Nilo Ocidental e o vírus Mayaro. Somente em 2022 foram registrados no Brasil 2,3 milhões de casos de dengue, além de 3,3 mil casos de febre Zika e 265 mil casos de febre Chikungunya. No entanto, apesar do elevado número de casos, a quantidade de dados genômicos atualizados destes vírus ainda é limitada, e em relação ao CHIKV, especificamente, os dados genômicos disponíveis representavam surtos localizados e limitados a alguns estados, enquanto outras áreas encontravam-se sub-representadas.

O monitoramento genômico do CHIKV é relevante devido a rápida expansão geográfica do CHIKV no Brasil e ao risco de emergência de mutações que podem aumentar a adaptação do vírus aos mosquitos vetores do gênero *Aedes spp*., conforme observado na epidemia de Chikungunya registrada em ilhas do oceano Índico entre 2005 e 2006. Nesse sentido, um sistema de vigilância que integra a tecnologia de sequenciamento de DNA com dados demográficos tem sido sugerido como abordagem complementar para a prevenção e controle de epidemias, através da rápida geração de dados genômicos virais que podem ser utilizados para a atualização de *primers* e testes de diagnóstico, a identificação mais precisa de linhagens virais circulantes e a identificação de rotas de dispersão viral, onde estratégias focadas no controle do vetor podem ser implementadas para reduzir o risco de recorrência de surtos.

3 OBJETIVOS

3.1 Objetivo geral

Analisar a diversidade genética e as relações filogenéticas do *Chikungunya virus* e de *Orthoflavivirus* identificados em amostras clínicas coletadas durante atividades de vigilância genômica realizadas nos anos de 2021 e 2022 no Brasil.

3.2 Objetivos específicos

I. Avaliar a existência de variações nucleotídicas entre as novas sequências genômicas do *Chikungunya virus* obtidas em amostras clínicas coletadas em diferentes estados brasileiros durante a epidemia de febre Chikungunya reportada nos anos de 2021 e 2022.
II. Identificar as possíveis trajetórias de dispersão espacial do *Chikungunya virus* através de reconstruções filogeográficas.
III. Identificar os genótipos do *Chikungunya virus* e de diferentes vírus do gênero *Orthoflavivirus* usando sequências nucleotídicas geradas por um protocolo de detecção multiplex de arbovírus baseado na tecnologia de sequenciamento por nanoporos.

4 RESULTADOS

4.1 Vigilância genômica do *Chikungunya virus* no Brasil

Artigo 1: Chikungunya virus ECSA lineage reintroduction in the northeasternmost region of Brazil. Publicado na revista *International Journal of Infectious Diseases*: Xavier *et al*., 2021.

Artigo 2: Short report: Introduction of chikungunya virus ECSA genotype into the Brazilian Midwest and its dispersion through the Americas. Publicado na revista *PLoS Neglected Tropical Diseases*: de Oliveira, Fonseca, Xavier *et al*., 2021.

Artigo 3: Increased interregional virus exchange and nucleotide diversity outline the expansion of chikungunya virus in Brazil. Publicado na revista *Nature Communications*: Xavier *et al*., 2023.

International Journal of Infectious Diseases 105 (2021) 120–123

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

Short Communication

Chikungunya virus ECSA lineage reintroduction in the northeasternmost region of Brazil

Joilson Xavier[a], Vagner Fonseca[a,b,c], Joao Felipe Bezerra[d,e], Manoella do Monte Alves[f], Maria Angélica Mares-Guia[g], Ingra Morales Claro[h], Ronaldo de Jesus[a], Talita Adelino[a], Emerson Araújo[b], Karina Ribeiro Leite Jardim Cavalcante[b], Stephane Tosta[a], Themis Rocha de Souza[i], Flavia Emanuelle Moreira da Cruz[i], Allison de Araújo Fabri[g], Elaine Cristina de Oliveira[j], Noely Fabiana Oliveira de Moura[b], Rodrigo Fabiano do Carmo Said[k], Carlos Frederico Campelo de Albuquerque[k], Vasco Azevedo[a], Tulio de Oliveira[c], Ana Maria Bispo de Filippis[g], Rivaldo Venâncio da Cunha[l], Kleber Giovanni Luz[f], Marta Giovanetti[a,g], Luiz Carlos Junior Alcantara[a,g,*]

[a] Laboratório de Genética Celular e Molecular, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[b] Secretaria de Vigilância em Saúde, Ministério da Saúde, Brasília, Distrito Federal, Brazil
[c] KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), University of KwaZuluNatal, Durban, South Africa
[d] Laboratório Central de Saúde Pública do Rio Grande do Norte Dr. Almino Fernandes, Natal, Brazil
[e] Escola Técnica de Saúde, Universidade Federal da Paraíba, Brazil
[f] Departamento de Infectologia, Universidade Federal do Rio Grande do Norte, Natal, Brazil
[g] Laboratório de Flavivírus, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil
[h] Instituto de Medicina Tropical e Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil
[i] Secretaria de Saúde do Estado do Rio Grande do Norte, Brazil
[j] Laboratório Central de Saúde Pública do Estado de Mato Grosso, Cuiabá, Brazil
[k] Organização Pan-Americana da Saúde/Organização Mundial da Saúde, Brasília, Distrito Federal, Brazil
[l] Fundação Oswaldo Cruz Pantanal, Campo Grande, Brazil

ARTICLE INFO

Article history:
Received 7 January 2021
Accepted 13 January 2021

Keywords:
Chikungunya virus
ECSA lineage
Northeast Brazil
Genomic epidemiology

ABSTRACT

The Northeast region of Brazil registered the second-highest incidence proportion of Chikungunya fever in 2019. In that year, an outbreak consisting of patients presenting with febrile disease associated with joint pain was reported by the public primary health care service in the city of Natal, in the state of Rio Grande do Norte, in March 2019. At first, the aetiological agent of the disease was undetermined. Since much is still unknown about chikungunya virus' (CHIKV) genomic diversity and evolutionary history in this northeasternmost state, we used a combination of portable whole-genome sequencing, molecular clock, and epidemiological analyses that revealed the reintroduction of the CHIKV East-Central-South-African (ECSA) lineage into Rio Grande do Norte. We estimated that the CHIKV ECSA lineage was first introduced into Rio Grande do Norte in early June 2014, while the 2019 outbreak clade diverged around April 2018, during a period of increased Chikungunya incidence in the Southeast region, which might have acted as a source of virus dispersion towards the Northeast region. Together, these results confirm that the ECSA lineage continues to spread across the country through interregional importation events, likely mediated by human mobility.

Introduction

Since its emergence in East Africa, the chikungunya virus (CHIKV) has caused more than 70 epidemics worldwide, mainly in Southeast Asian and Latin American countries (Mascarenhas et al., 2018).

* Corresponding author at: Laboratório de Flavivírus, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil.
E-mail address: luiz.alcantara@ioc.fiocruz.br (L.C.J. Alcantara).

https://doi.org/10.1016/j.ijid.2021.01.026

Table 1
Characteristic and clinical symptoms of patients with Chikungunya fever.

	Patients (n = 13)
Mean year of age (SD)	42 ± 22
Gender	
Female	4 (31%)
Male	9 (69%)
Symptoms onset date range	19/Mar/2019–29/Mar/2019
Presenting symptoms	
Fever	13 (100%)
Myalgia	8 (62%)
Cephalea	7 (54%)
Skin rash	8 (62%)
Vomiting	2 (15%)
Arthralgia	12 (92%)
Retro-orbital pain	4 (31%)
Conjunctival hyperemia	2 (15%)
CHIKV RT-qPCR mean Ct value (cycle threshold)	28.15 ± 6.69

Data is presented as a number (%) unless otherwise indicated. SD, standard deviation. RT-qPCR, quantitative reverse transcription PCR.

Northeast Brazil presented with the second-highest incidence (59.4 cases per 100,000 population) of Chikungunya reported in 2019 in the country (Ministério da Saúde, 2020). In that year, an outbreak of 13,713 Chikungunya cases was recorded in the state of Rio Grande do Norte, in the far northeastern tip of Brazil, where a previous outbreak of almost 25,000 probable cases was reported in 2016 (Ministério da Saúde, 2018). Despite the high number of cases, much remains unknown about CHIKV genomic diversity and the evolutionary history in this northeastern state. We generated eight CHIKV genomes from the 2019 outbreak registered in Natal, the capital city of Rio Grande do Norte, and provide a genomic epidemiological report of the virus circulating in that state.

Methods

An outbreak consisting of patients presenting with febrile disease associated with joint pain was reported by the public primary health care service in the city of Natal, Rio Grande do Norte, in March 2019. Since the aetiological disease agent was undetermined at first, serum samples were provided to the State Central Laboratory of Public Health (Dr. Almino Fernandes) and then sent to the Laboratory of Flavivirus at the Oswaldo Cruz Foundation (Fiocruz) to confirm the laboratory diagnosis by RT-qPCR (Lanciotti et al., 2007). CHIKV-RNA positive samples were submitted to a MinION sequencing protocol (Quick et al., 2017), and newly generated genomes were used for phylogenetic analysis (more details in Supplementary files).

Results

Laboratory diagnostics confirmed thirteen positive samples of CHIKV RNA with a mean Ct value of 28.15. We generated eight CHIKV genomes from patients with an average age of 42 years. We

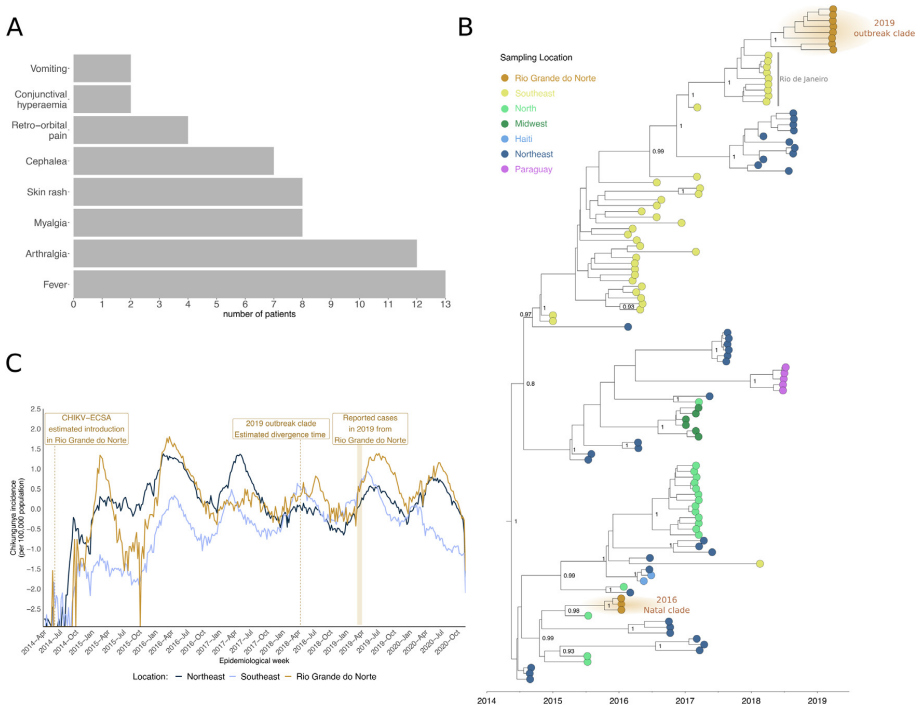


Figure 1. Symptoms frequency, epidemic curves, and phylogenetic analysis from Chikungunya outbreaks reported in Rio Grande do Norte, northeast Brazil. (A) Symptom frequency among thirteen patients with confirmed laboratory diagnosis of chikungunya virus infection. (B) Time-scaled maximum clade credibility tree of chikungunya virus East-Central-South African lineage in Brazil, including the eight new genomes generated in this study. Tips are colored according to the sample source location. Values around nodes represent posterior probability support of the tree nodes inferred under Bayesian Evolutionary Analysis using a molecular clock approach. (C) Time-series plot of Chikungunya incidence per 100,000 population calculated from notified cases from Rio Grande do Norte, Northeast, and Southeast regions of Brazil. Y-axis values were log-transformed.

found five (38%) patients who presented with fever associated with myalgia, arthralgia, and rash (see Table 1 and Figure 1A). The eight CHIKV genomes generated have an average genome coverage of 93% (ranging from 92% to 94%); they belonged to the East-Central-South-African (ECSA) lineage as indicated by our Maximum Likelihood (ML) phylogeny (see Supplementary files).

We inferred a time-measured phylogeny using a dataset comprising 110 Brazilian sequences plus two isolates from Haiti and five from Paraguay. Both our ML and Bayesian phylogenies showed the 2019 CHIKV isolates from Rio Grande do Norte formed a single well-supported clade (hereafter, the 2019 outbreak clade; posterior probability = 1.00), which was closely related to isolates from Rio de Janeiro (posterior probability = 1.00) (Figure 1B). Note that this 2019 outbreak clade did not cluster to the other three previously published isolates from Natal, Rio Grande do Norte, sampled in January 2016 (hereafter, the 2016 Natal clade; posterior probability = 1.00). We estimated the time of the most recent common ancestor (TMRCA) of the 2019 outbreak clade to be around mid-April 2018 (95% BCI: early December 2017 to late August 2018) (Figure 1B and C). Moreover, the TMRCA of the CHIKV-ECSA lineage in Rio Grande do Norte was estimated to be in early June 2014 (95% BCI: late December 2013 to early October 2014).

Weekly reported incidences revealed three major outbreaks in Rio Grande do Norte during early 2015, early 2016, and mid-2019. Smaller waves were observed throughout the years from 2017 to 2020, indicating the persistence of the virus in the state through year-round transmission cycles (see Figure 1C), which were also observed in other regions of Brazil (Supplementary files).

Discussion

In this study, we generated eight CHIKV genomes from symptomatic cases of Chikungunya fever reported during an outbreak in March 2019 in Natal, the capital city of the state of Rio Grande do Norte. We found fever, arthralgia, myalgia, and rash were the most common presenting symptoms, which are similar findings to those described in other studies of suspected cases during the 2016 Chikungunya outbreak in Natal (Monteiro et al., 2020) and on Chikungunya-associated neurological disease cases reported from 2014 to 2016 in another northeastern Brazilian state (Brito Ferreira et al., 2020).

Our phylogenetic analysis revealed the 2019 outbreak clade is distantly related to other isolates from the previous 2016 outbreak in Rio Grande do Norte. This suggests at least two importation events of CHIKV-ECSA lineage occurred in Rio Grande do Norte; the time of the first virus importation event estimated to be early June 2014. This estimate indicates the virus was circulating for several months before the first Chikungunya cases were confirmed in 2015 in Rio Grande do Norte, including one Chikungunya-associated death, as indicated by available epidemiological surveillance data from the State Health Department (Secretaria da Saúde Pública, 2017).

From the combined analysis of regional epidemic curves and time-scaled phylogeny, we hypothesize that the 2019 outbreak in Rio Grande do Norte was caused by a second event of CHIKV introduction imported from Rio de Janeiro. We estimate the divergence time of the 2019 outbreak clade to be mid-April 2018, which corresponds to a period of increased Chikungunya incidence in the southeast region, where Rio de Janeiro is located. This might be explained by virus interregional spread through people's movement, as discussed elsewhere (Candido et al., 2020; Churakov et al., 2019).

Our study increases the number of CHIKV genomic sequences and, consequently, provides more information on the evolution and genomic diversity of the ECSA lineage currently circulating in Brazil's Northeast region, where this lineage was first introduced.

Conflict of interest

The authors have no competing interests to disclose.

Funding

This work was supported by the ZIBRA-2 project grants (Decit/SCTIE/BrMoH/CNPq-440685/2016-8 and CAPES-88887.130716/2016-00), by the European Union's Horizon 2020 Research and Innovation Programme under ZIKAlliance Grant Agreement no. 734548, and by the Pan American Health Organization, Regional Office for the Americas of the World Health Organization (007-FEX-19-2-2-30). JX is supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Finance Code 001. MG and LCJA are supported by Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ). VF and TO are supported by a Flagship grant from the South African Medical Research Council (MRC-RFA-UFSP-01-2013/UKZN HIVEPI). Funders had no role in study design, data collection and analysis, writing, and/or decision to publish the manuscript.

Ethics statement

This project was reviewed and approved by the Comissão Nacional de Ética em Pesquisa (CONEP) [National Research Ethics Committee], as part of the arboviral genomic surveillance efforts within the terms of Resolution 510/2016 of CONEP, by the Pan American Health Organization Ethics Review Committee (PAHOERC) (Ref. No. PAHO-2016-08-0029), and by the Oswaldo Cruz Foundation Ethics Committee (CAAE: 90249218.6.1001.5248).

Data availability

Newly generated CHIKV sequences have been deposited in GenBank under accession numbers MW260512-MW260519.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ijid.2021.01.026.

References

Brito Ferreira ML, Militão de Albuquerque M de FP, de Brito CAA, de Oliveira França RF, Porto Moreira ÁJ, de Morais Machado MÍ, et al. Neurological disease in adults with Zika and Chikungunya virus infection in Northeast Brazil: a prospective observational study. Lancet Neurol 2020;19(10):826–39.

Candido DS, Claro IM, de Jesus JG, Souza WM, Moreira FRR, Dellicour S, et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. Science 2020;369(6508):1255–60.

Churakov M, Villabona-Arenas CJ, Kraemer MUG, Salje H, Cauchemez S. Spatio-temporal dynamics of dengue in Brazil: seasonal traveling waves and determinants of regional synchrony. PLoS Negl Trop Dis 2019;13(4)e0007012.

Lanciotti RS, Kosoy OL, Laven JJ, Panella AJ, Velez JO, Lambert AJ, et al. Chikungunya virus in US travelers returning from India, 2006. Emerg Infect Dis 2007;13(5):764–7.

Mascarenhas M, Garasia S, Berthiaume P, Corrin T, Greig J, Ng V, et al. A scoping review of published literature on Chikungunya virus. PLoS One 2018;13(11)e0207554.

Ministério da Saúde. Boletim Epidemiológico — Monitoramento dos casos de dengue, febre de chikungunya e febre pelo vírus Zika até a Semana Epidemiológica 52, 2017, vol. 49. Secretaria de Vigilância em Saúde; 2018. https://antigo.saude.gov.br/images/pdf/2018/janeiro/23/Boletim-2018-001-Dengue.pdf.

Ministério da Saúde. Boletim Epidemiológico. Monitoramento dos casos de arboviroses urbanas transmitidas pelo Aedes (Dengue, Chikungunya e Zika), Semanas Epidemiológicas 01 a 52/2019, vol. 51. Secretaria de Vigilância em Saúde; 2020. https://antigo.saude.gov.br/images/pdf/2020/janeiro/20/Boletim-epidemiologico-SVS-02-1-.pdf.

Monteiro JD, Valverde JG, Morais IC, Souza CR de M, Fagundes Neto JC, de Melo MF, et al. Epidemiologic and clinical investigations during a Chikungunya outbreak in Rio Grande do Norte State, Brazil. PLoS One 2020;15(11) e0241799.
Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc 2017;12:1261–76, doi:http://dx.doi.org/10.1038/nprot.2017.066.
Secretaria da Saúde Pública. Boletim Epidemiológico. Atualização Da Situação Epidemiológica Das Arboviroses No Rio Grande Do Norte. Monitoramento Dos Casos De Dengue, Febre De Chikungunya E Febre Pelo Vírus Zika, Até A Semana Epidemiológica No 52/2016. 2017. http://adcon.rn.gov.br/ACERVO/sesap/DOC/DOC000000000140313.PDF.

APPENDIX

Chikungunya virus ECSA lineage reintroduction in the northeasternmost region of Brazil

Joilson Xavier[1], Vagner Fonseca[1,2,3], Joao Felipe Bezerra[4,5], Manoella do Monte Alves[6], Maria Angélica Mares-Guia[7], Ingra Morales Claro[8], Ronaldo de Jesus[1], Talita Adelino[1], Emerson Araújo[2], Karina Ribeiro Leite Jardim Cavalcante[2], Stephane Tosta[1], Themis Rocha de Souza[9], Flavia Emanuelle Moreira da Cruz[9], Allison de Araújo Fabri[7], Elaine Cristina de Oliveira[10], Noely Fabiana Oliveira de Moura[2], Rodrigo Fabiano do Carmo Said[11], Carlos Frederico Campelo de Albuquerque[11], Vasco Azevedo[1], Tulio de Oliveira[3], Ana Maria Bispo de Filippis[7], Rivaldo Venâncio da Cunha[12], Kleber Giovanni Luz[6], Marta Giovanetti[1,7], Luiz Carlos Junior Alcantara[1,7,*]

[1]Laboratório de Genética Celular e Molecular, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil;
[2]Secretaria de Vigilância em Saúde, Ministério da Saúde, Brasília, Distrito Federal, Brazil;
[3]KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), University of KwaZuluNatal, Durban, South Africa;
[4]Laboratório Central de Saúde Pública do Rio Grande do Norte Dr. Almino Fernandes, Natal, Brazil;
[5]Escola Técnica de Saúde, Universidade Federal da Paraíba, Brazil;
[6]Departamento de Infectologia, Universidade Federal do Rio Grande do Norte, Natal, Brazil;
[7]Laboratório de Flavivírus, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil;
[8]Instituto de Medicina Tropical e Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil;
[9]Secretaria de Saúde do Estado do Rio Grande do Norte, Brasil;
[10]Laboratório Central de Saúde Pública do Estado de Mato Grosso, Cuiabá, Brazil;
[11]Organização Pan-Americana da Saúde/Organização Mundial da Saúde, Brasília, Distrito Federal, Brazil;
[12]Fundação Oswaldo Cruz Pantanal, Campo Grande, Brazil;

***Corresponding author.** Luiz Carlos Junior Alcantara, Laboratório de Flavivírus, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil.
E-mail address: luiz.alcantara@ioc.fiocruz.br

Key words: chikungunya virus, ECSA lineage, Northeast Brazil, genomic epidemiology

METHODS

Sample collection and laboratory diagnosis

Serum samples from individuals presented with febrile disease associated with joint pain were collected for molecular diagnostics performed at the State Central Laboratory of Public Health Dr Almino Fernandes, Rio Grande do Norte state, northeast Brazil. Then, to confirm chikungunya infection, those samples were sent to the Brazilian national reference laboratory for arbovirus diagnosis at the Laboratory of Flavivirus in the Oswaldo Cruz Institute (IOC) from the Oswaldo Cruz Foundation (Fiocruz). Serum samples were submitted to nucleic acid purification using the MagMax Viral RNA Mini kit (Thermo Fischer Scientific), following manufacturer's recommendations. RT-qPCR tests were performed on 13 samples sent to Fiocruz, where chikungunya infection was confirmed (Lanciotti et al., 2007).

Sequencing experiments

Samples were submitted to cDNA synthesis and PCR, using a sequencing protocol based on multiplex PCR tiling amplicon approach design for MinION nanopore sequencing (Quick et al. 2017). DNA library preparation was performed on 8 selected samples (selection based on DNA concentration after clean-up being > 4ng/μL) using the Ligation Sequencing Kit (Oxford Nanopore Technologies). Individual samples were barcoded using the Native Barcoding Kit (NBD104, Oxford Nanopore Technologies, Oxford, UK). Sequencing library was loaded onto a R9.4 flow cell and data was collected for up to 48 sequencing hours.

Generation of consensus sequences

We used Guppy (https://github.com/nanoporetech) for basecalling of raw FAST5 files, while qcat (https://github.com/nanoporetech/qcat) was used for barcode demultiplexing. Consensus sequences were generated by *de novo* assembling using Genome Detective (https://www.genomedetective.com/) (Vilsker et al. 2019).

Phylogenetic and Bayesian analysis

The 8 new sequences reported in this study were initially submitted to a genotyping analysis using the phylogenetic arbovirus subtyping tool from Genome Detective (http://genomedetective.com/app/typingtool/chikungunya) (Fonseca et al. 2019). We used MAFFT (Katoh et al. 2005) to align these 8 new sequences to other 224 closely related CHIKV sequences publicly available in NCBI. We assessed phylogenetic signal on this dataset (n=232) using the likelihood mapping analysis implemented in IQ-TREE 2.1.1 software (Nguyen et al. 2015). We used IQ-TREE software to infer a Maximum Likelihood (ML) phylogeny under the GTR+F+I+G4 best fit substitution model identified by ModelFinder (Kalyaanamoorthy et al. 2017).

From the ML tree inferred from the dataset we constructed a small sub-dataset (n= 117) containing mainly CHIKV ECSA Brazilian sequences according to the Brazilian clade observed in the ML tree. We used TempEst (Rambaut et al. 2016) to assess the presence of temporal signal. To infer a time-scaled phylogeny from the sub-dataset, we used BEAST v1.10.4 (Suchard et al. 2018) to perform Bayesian phylogenetic analysis under GTR+G4 codon partition (CP)1+2,3 model. We combined sextuplicate independent runs of 50 million states each. For these runs, we used the uncorrelated relaxed molecular clock and the Bayesian skygrid population model (Gill et al. 2013). For comparison, we also used the strict molecular clock with the Bayesian skygrid

population model. Maximum clade trees were summarized from the MCMC samples using TreeAnnotator after discarding 10% as burn-in.

Epidemic curves from Chikungunya cases reported in Brazil

We used chikungunya notified cases data from the Brazil Ministry of Health National Reporting System (SINAN-Net) (Ministério da Saúde, 2020) to calculate incidence and to plot time series charts.

SUPPLEMENTARY FIGURES AND TABLES

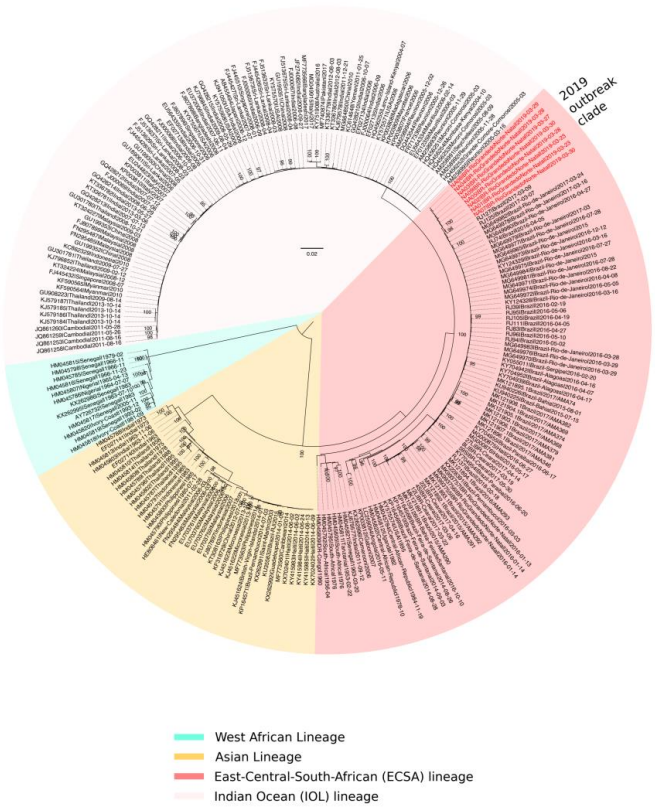


Figure S1. Global chikungunya phylogeny. A Maximum Likelihood phylogeny comprising 8 new sequences (coloured in red) generated in this study plus 224 sequences retrieved from NCBI. Chikungunya virus lineages clades are coloured according to the legend.

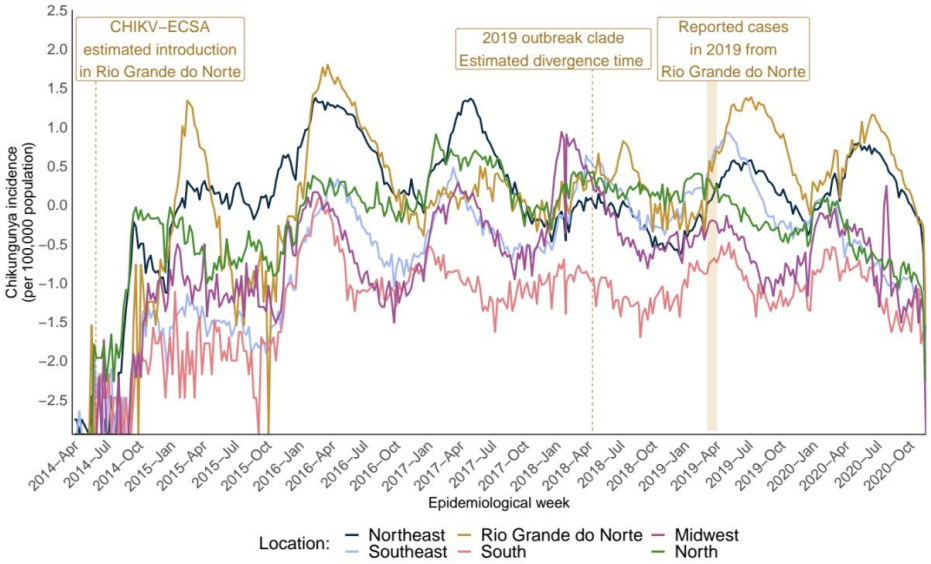


Figure S2. Time series of chikungunya outbreaks reported in Rio Grande do Norte state and all regions of Brazil. Epidemic curve of chikungunya virus incidence per 100,000 inhabitants calculated from notified cases from Rio Grande do Norte state and all regions of Brazil. Y-axis values were log-transformed.

Table S1. Summary statistics of epidemiological and sequencing data.

Sample ID	Isolate ID	NCBI Accession number	Sample type	Symptoms onset date	Collection date	Delay (days) onset and collection date	Year of age	Sex	State	Municipality	Ct value	Reads	Coverage (%)	Depth of coverage
2746	NA01	MW260512	Serum	2019-03-21	2019-03-23	2	32	M	RN	Natal	21	159071	93.6	9740.8
2747	NA02	MW260513	Serum	2019-03-23	2019-03-23	0	19	M	RN	Natal	22	105093	94.5	7614.8
2748	NA03	MW260515	Serum	2019-03-25	2019-03-27	2	53	M	RN	Natal	26	280612	93.1	15545.7
2749	NA04	MW260514	Serum	2019-03-26	2019-03-28	2	23	M	RN	Natal	34	208775	92	10591.3
2750	NA05	MW260516	Serum	2019-03-27	2019-03-28	1	22	F	RN	Natal	28	147517	93.2	8447.5
2751	NA06	MW260517	Serum	2019-03-24	2019-03-29	5	78	F	RN	Natal	29	188258	92.2	10414.6
2753	NA07	MW260518	Serum	2019-03-28	2019-03-30	2	78	F	RN	Natal	17	177192	95.2	15484.6
2755	NA09	MW260519	Serum	2019-03-29	2019-03-30	1	57	M	RN	Natal	19	223411	94	11299.2
2754/2019	-	-	Serum	2019-03-27	2019-04-01	5	57	F	RN	Natal	34	-	-	-
2743/2019	-	-	Serum	2019-03-20	2019-03-23	3	25	M	RN	Natal	37	-	-	-
2744/2019	-	-	Serum	2019-03-19	2019-03-23	4	55	M	RN	Natal	36	-	-	-
2745/2019	-	-	Serum	2019-03-21	2019-03-23	2	25	M	RN	Natal	33	-	-	-
2752/2019	-	-	Serum	2019-03-28	2019-03-30	2	26	M	RN	Natal	30	-	-	-

Table S2. Unique nucleotide substitutions shared by CHIKV isolates from the 2019 outbreak in Natal, Rio Grande do Norte, Brazil.

Nucleotide position*	Synonymous substitution	Genes
1675	G→A	nsP1
2326	A→ G	nsP2
3077	A→ C	nsP2
3529	C→ T	nsP2
5080	T →C	nsP3
7869	C →T	capsid

*Compared to the NCBI CHIKV reference genome (NC_004162.2) and to isolate KP164568 and MK244641, from Bahia and Rio de Janeiro Brazilian states, respectively.

REFERENCES

Fonseca, Vagner, Pieter J. K. Libin, Kristof Theys, Nuno R. Faria, Marcio R. T. Nunes, Maria I. Restovic, Murilo Freire, et al. 2019. “A Computational Method for the Identification of Dengue, Zika and Chikungunya Virus Species and Genotypes.” *PLoS Neglected Tropical Diseases* 13 (5): e0007231.

Gill, Mandev S., Philippe Lemey, Nuno R. Faria, Andrew Rambaut, Beth Shapiro, and Marc A. Suchard. 2013. “Improving Bayesian Population Dynamics Inference: A Coalescent-Based Model for Multiple Loci.” *Molecular Biology and Evolution* 30 (3): 713–24.

Kalyaanamoorthy, Subha, Bui Quang Minh, Thomas K. F. Wong, Arndt von Haeseler, and Lars S. Jermiin. 2017. “ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates.” *Nature Methods* 14 (6): 587–89.

Katoh, Kazutaka, Kei-Ichi Kuma, Hiroyuki Toh, and Takashi Miyata. 2005. “MAFFT Version 5: Improvement in Accuracy of Multiple Sequence Alignment.” *Nucleic Acids Research* 33 (2): 511–18.

Lanciotti, R. S., Kosoy, O. L., Laven, J. J., Panella, A. J., Velez, J. O., Lambert, A. J., & Campbell, G. L. (2007). Chikungunya virus in US travelers returning from India, 2006. *Emerging Infectious Diseases*, *13*(5), 764–767.

Ministério da Saúde. (2020). Sistema de Informação de Agravos de Notificação - Sinan/Dengue/Chikungunya. http://portalsinan.saude.gov.br/sinan-dengue-chikungunya

Nguyen, Lam-Tung, Heiko A. Schmidt, Arndt von Haeseler, and Bui Quang Minh. 2015. “IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood

Phylogenies." *Molecular Biology and Evolution* 32 (1): 268–74.
Quick, Joshua, Nathan D. Grubaugh, Steven T. Pullan, Ingra M. Claro, Andrew D. Smith, Karthik Gangavarapu, Glenn Oliveira, et al. 2017. "Multiplex PCR Method for MinION and Illumina Sequencing of Zika and Other Virus Genomes Directly from Clinical Samples." *Nature Protocols* 12 (6): 1261–76.
Rambaut, Andrew, Tommy T. Lam, Luiz Max Carvalho, and Oliver G. Pybus. 2016. "Exploring the Temporal Structure of Heterochronous Sequences Using TempEst (formerly Path-O-Gen)." *Virus Evolution* 2 (1): vew007.
Suchard, Marc A., Philippe Lemey, Guy Baele, Daniel L. Ayres, Alexei J. Drummond, and Andrew Rambaut. 2018. "Bayesian Phylogenetic and Phylodynamic Data Integration Using BEAST 1.10." *Virus Evolution* 4 (1): 170.
Vilsker, Michael, Yumna Moosa, Sam Nooij, Vagner Fonseca, Yoika Ghysens, Korneel Dumon, Raf Pauwels, et al. 2019. "Genome Detective: An Automated System for Virus Identification from High-Throughput Sequencing Data." *Bioinformatics* 35 (5): 871–73.

PLOS NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

Short report: Introduction of chikungunya virus ECSA genotype into the Brazilian Midwest and its dispersion through the Americas

Elaine Cristina de Oliveira[1☯], Vagner Fonseca[2,3,4☯], Joilson Xavier[2☯], Talita Adelino[5], Ingra Morales Claro[6], Allison Fabri[7], Eduardo Marques Macario[8], Ana Elisa Viniski[1], Claudio Luis Campos Souza[1], Evanil Sebastiana Gomes da Costa[9], Claudia Soares de Sousa[9], Flávia Guimarães Dias Duarte[9], Arnaldo Correia de Medeiros[8], Carlos F. Campelo de Albuquerque[10], Rivaldo Venancio Cunha[11], Noely Fabiana Oliveira De Moura[12], Ana Maria Bispo de Filippis[7], Tulio de Oliveira[3], José Lourenço[13], André Luiz de Abreu[4], Luiz Carlos Junior Alcantara[2,7]*, Marta Giovanetti[2,7]*

1 Laboratório Central de Saúde Pública do Estado de Mato Grosso, Cuiabá, Brazil, **2** Laboratório de Genética Celular e Molecular, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, **3** KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), College of Health Sciences, University of KwaZuluNatal, Durban 4001, South Africa, **4** Coordenação Geral dos Laboratórios de Saúde Pública/Secretaria de Vigilância em Saúde, Ministério da Saúde, Brasília, Distrito Federal, Brazil, **5** Laboratório Central de Saúde Pública, Fundação Ezequiel Dias, Belo Horizonte, Brazil, **6** Instituto de Medicina Tropical e Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil, **7** Laboratório de Flavivírus, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil, **8** Secretaria de Vigilância em Saúde/Ministério da Saúde, Brasília, Distrito Federal, Brazil, **9** Secretaria de Saúde de Cuiaba, Mato Grosso, Brazil, **10** Organização Pan-Americana da Saúde/Organização Mundial da Saúde, Brasília-DF, Brazil, **11** Fundação Oswaldo Cruz, Bio-Manguinhos, Rio de Janeiro, Brasil, **12** Coordenacao Geral das Arboviroses, Secretaria de Vigilância em Saúde/Ministério da Saúde, Brasília, Distrito Federal, Brazil, **13** Department of Zoology, University of Oxford, Oxford, OX1 3PS, United Kingdom

☯ These authors contributed equally to this work.
* luiz.alcantara@ioc.fiocruz.br (LCJA); giovanetti.marta@gmail.com (GM)

OPEN ACCESS

Citation: de Oliveira EC, Fonseca V, Xavier J, Adelino T, Morales Claro I, Fabri A, et al. (2021) Short report: Introduction of chikungunya virus ECSA genotype into the Brazilian Midwest and its dispersion through the Americas. PLoS Negl Trop Dis 15(4): e0009290. https://doi.org/10.1371/journal.pntd.0009290

Editor: Brett M. Forshey, DoD - AFHSB, UNITED STATES

Received: October 12, 2020

Accepted: March 4, 2021

Published: April 16, 2021

Data Availability Statement: Newly generated CHIKV sequences have been deposited in GenBank under accession numbers MN428504-MN428527. https://www.ncbi.nlm.nih.gov/nuccore/MN428504 https://www.ncbi.nlm.nih.gov/nuccore/MN428506 https://www.ncbi.nlm.nih.gov/nuccore/MN428507 https://www.ncbi.nlm.nih.gov/nuccore/MN428508 https://www.ncbi.nlm.nih.gov/nuccore/MN428511 https://www.ncbi.nlm.nih.gov/nuccore/MN428509 https://www.ncbi.nlm.nih.gov/nuccore/MN428510 https://www.ncbi.nlm.nih.gov/nuccore/MN428513

Abstract

Since introduction into Brazil in 2014, chikungunya virus (CHIKV) has presented sustained transmission, although much is unknown about its circulation in the midwestern states. Here, we analyze 24 novel partial and near complete CHIKV genomes from Cuiaba, an urban metropolis located in the Brazilian midwestern state of Mato Grosso (MT).

Nanopore technology was used for sequencing CHIKV complete genomes. Phylogenetic and epidemiological approaches were used to explore the recent spatio-temporal evolution and spread of the CHIKV-ECSA genotype in Midwest Brazil as well as in the Americas.

Epidemiological data revealed a reduction in the number of reported cases over 2018–2020, likely as a consequence of a gradual accumulation of herd-immunity. Phylogeographic reconstructions revealed that at least two independent introductions of the ECSA lineage occurred in MT from a dispersion event originating in the northeastern region and suggest that the midwestern Brazilian region appears to have acted as a source of virus transmission towards Paraguay, a bordering South American country.

https://www.ncbi.nlm.nih.gov/nuccore/MN428512
https://www.ncbi.nlm.nih.gov/nuccore/MN428514
https://www.ncbi.nlm.nih.gov/nuccore/MN428515
https://www.ncbi.nlm.nih.gov/nuccore/MN428516
https://www.ncbi.nlm.nih.gov/nuccore/MN428517
https://www.ncbi.nlm.nih.gov/nuccore/MN428518
https://www.ncbi.nlm.nih.gov/nuccore/MN428519
https://www.ncbi.nlm.nih.gov/nuccore/MN428522
https://www.ncbi.nlm.nih.gov/nuccore/MN428520
https://www.ncbi.nlm.nih.gov/nuccore/MN428521
https://www.ncbi.nlm.nih.gov/nuccore/MN428523
https://www.ncbi.nlm.nih.gov/nuccore/MN428524
https://www.ncbi.nlm.nih.gov/nuccore/MN428525
https://www.ncbi.nlm.nih.gov/nuccore/MN428526
https://www.ncbi.nlm.nih.gov/nuccore/MN428527.

Funding: This work was supported by Decit/SCTIE/BrMoH/CNPq (440685/2016-8 - 421598/2018-2), by CAPES (88887.130716/2016-00), by the European Union's Horizon 2020 under grant agreements ZIKACTION (734857) and ZIKAPLAN (734548), by the Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro – FAPERJ (E-26/2002.930/2016), the International Development Research Centre (IDRC) Canada (108411-001). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Our results show a complex dynamic of transmission between epidemic seasons and suggest a possible role of Brazil as a source for international dispersion of the CHIKV-ECSA genotype to other countries in the Americas.

Author summary

Since its introduction into Brazil in 2014, chikungunya virus (CHIKV) has presented sustained transmission, although much is unknown about its circulation in the midwestern states. Here, using a combined strategy of a mobile sequencing mission through this region, with genomic and epidemiological analysis we generated 24 novel partial and near complete CHIKV genome sequences by means of portable nanopore sequencing of chikungunya virus isolates obtained directly from clinical samples. Our findings reinforce that continued genomic surveillance strategies are needed to assist in the monitoring and understanding of arbovirus epidemics, which might help to attenuate public health impact of infectious diseases.

Introduction

After the introduction of the Asian lineage of chikungunya virus (CHIKV) into the Americas in 2013, and subsequent detection of the East/Central/South African (ECSA) lineage in 2014 in Bahia state Northeast Brazil, more than 2.9 million infections have been reported in Brazil up to 2019 [1]. Common clinical manifestations of the disease include fever, muscle pain, rash and severe joint pain, which may last for months to years [2]. It is argued that 82% of infections caused by the ECSA lineage are symptomatic, and there are suggestions of lineage-specific infection outcomes [3]. The ECSA lineage seems to be the main genotype currently circulating in Brazil since its introduction in the northeastern region in 2014, despite detection of the Asian lineage in that same year in northern Brazil [4]. From previous studies addressing inter-regional dispersion of the virus [4], CHIKV outbreaks have been registered in northern, northeastern, southeastern and midwestern states of Brazil between 2016 and 2019 [5–8]. The number of probable cases reached 87,687 in 2018 and 132,205 in 2019, with the midwestern region accounting for the second largest number of cases (15.8%) registered in 2018 in Brazil [9]. Despite the large number of cases, much is unknown about the genomic diversity and evolution of CHIKV lineages currently circulating in the midwestern region of Brazil, as well as their dispersion dynamics in South American countries that border this Brazilian region. Thus, to investigate the genomic diversity and evolution of CHIKV, we analyzed 24 new CHIKV genomes generated by next generation sequencing, providing additional information on the introduction and spread of the ECSA lineage in Midwest Brazil as well as in the Americas.

Materials and methods

Ethics statement

This research was reviewed and approved by the Ethical Committee of the Pan American World Health Organization (No. PAHO-2016-08-0029) and the Brazilian Ministry of Health (MoH) as part of the arbovirus genomic surveillance efforts within the terms of Resolution 510/2016 of CONEP (Comissão Nacional de Ética em Pesquisa, Ministério da Saúde; National

Ethical Committee for Research, Ministry of Health). Residual anonymized clinical diagnostic samples, with no or minimal risk to patients, were provided for research and surveillance purposes within the terms of Resolution 510/2016 of CONEP. Processing of human samples was approved and the need for participants consent was waived by the Institutional Review Board from the Fundação Oswaldo Cruz/Instituto Oswaldo Cruz (CEP/CAAE: 90249218.6.1001.5248; approval number 2.998.362).

Sample collection

A total number of 24 serum samples from individuals presenting with symptoms compatible with arbovirus infection, with availability of epidemiological metadata, such as date of symptom onset, date of sample collection, sex, age and municipality of residence, were collected for molecular diagnostics by the Central Public Health Laboratory of Mato Grosso state, Midwest Brazil, during genomic surveillance activities of the project ZIBRA2 (https://www.zibra2project.org/).

Viral RNA isolation and quantitative real-time RT-PCR

Serum samples were submitted to nucleic acid purification using the QIAmp Viral RNA Mini kit (Qiagen), following manufacturer's recommendations. The CHIKV RNA detection by RT-qPCR was performed using a Real Time RT-qPCR protocol adapted from [10], using Promega GoTaq Probe 1-Step RT-PCR System Kit in Bioer LineGene 9660 equipment. Samples were selected for sequencing based on Ct-value <32 to maximize genome coverage of clinical samples by nanopore sequencing [11] (**Table 1**).

cDNA synthesis and multiplex tiling PCR

Samples were submitted to a cDNA synthesis protocol [11] using ProtoScript II First Strand cDNA Synthesis Kit. Then, a multiplex tiling PCR was conducted using Q5 High Fidelity Hot-Start DNA Polymerase (New England Biolabs) and a CHIKV sequencing primers scheme (divided into two separated pools) designed using Primal Scheme (http://primal.zibraproject.org) [12]. The thermocycling conditions involved 40 cycles, and reaction conditions was previously reported in [12].

Library preparation and nanopore sequencing

Amplicons were purified using 1x AMPure XP Beads and cleaned-up PCR products concentrations were measured using Qubit dsDNA HS Assay Kit on a Qubit 3.0 fluorimeter (ThermoFisher). DNA library preparation was carried out using the Ligation Sequencing Kit and the Native Barcoding Kit (NBD104, Oxford Nanopore Technologies, Oxford, UK) [12]. Purified PCR products of each sample were quantified and DNA concentration were normalised before barcoding reactions. One barcode was used per sample in order to maximize the number of samples per flow cell. Sequencing library was loaded onto a R9.4 flow cell, and data was collected for up to 48 hours, but generally less.

Generation of consensus sequences

Raw files were basecalled using Guppy and barcode demultiplexing was performed using qcat. Consensus sequences were generated by *de novo* assembling using Genome Detective (https://www.genomedetective.com/) [13].

Table 1. Epidemiological data for the sequenced samples.

ID	Sample type	Date onset symptoms	Collection date	Age	Sex	State	Municipality	Ct value	Reads	Coverage (%)	Acession Number
CB08	Serum	2018-02-23	2018-02-23	33	F	MT	Cuiaba	18,59	900	43,70	MN428504
CB04	Serum	2018-05-18	2018-05-21	22	F	MT	Cuiaba	30,98	4318	52,60	MN428505
CB09	Serum	2018-01-29	2018-02-01	23	F	MT	Cuiaba	19,66	291904	93,60	MN428506
CB10	Serum	2018-07-03	2018-07-05	29	F	MT	Cuiaba	26,88	620828	82,70	MN428507
CB12	Serum	2018-04-25	2018-04-27	22	F	MT	Cuiaba	30,65	175360	78,60	MN428508
CB17	Serum	2018-03-11	2018-03-13	27	M	MT	Cuiaba	22,16	99926	83,90	MN428509
CB18	Serum	2018-04-24	2018-04-25	43	M	MT	Cuiaba	27,38	554552	79,90	MN428510
CB16	Serum	2018-04-12	2018-04-18	31	F	MT	Cuiaba	19,87	577642	85,10	MN428511
CB22	Serum	2018-03-10	2018-03-13	26	M	MT	Cuiaba	28,33	147381	88,50	MN428512
CB19	Serum	2018-04-05	2018-04-26	37	F	MT	Cuiaba	24,93	514308	82,70	MN428513
CB23	Serum	2018-03-13	2018-03-13	30	F	MT	Cuiaba	23,62	278973	53,00	MN428514
CB24	Serum	2018-03-18	2018-03-20	62	F	MT	Cuiaba	20,31	235010	93,50	MN428515
CB26	Serum	2018-03-21	2018-03-23	22	F	MT	Cuiaba	24,08	4167	86,50	MN428516
CB27	Serum	2018-03-15	2018-03-17	40	F	MT	Cuiaba	29,33	6855	84,10	MN428517
CB28	Serum	2018-03-07	2018-03-08	19	F	MT	Cuiaba	20,07	4391	89,40	MN428518
CB29	Serum	2018-03-15	2018-03-16	20	F	MT	Cuiaba	20,96	6384	88,20	MN428519
CB31	Serum	2018-03-08	2018-03-09	31	F	MT	Cuiaba	29,22	7537	85,20	MN428520
CB32	Serum	2018-03-16	2018-03-16	29	F	MT	Cuiaba	21,22	6010	87,10	MN428521
CB30	Serum	2018-03-07	2018-03-08	24	F	MT	Cuiaba	31,21	9104	83,50	MN428522
CB33	Serum	2018-03-08	2018-03-08	31	F	MT	Cuiaba	22,16	4753	89,40	MN428523
CB34	Serum	2018-03-05	2018-03-06	34	F	MT	Cuiaba	22,25	6554	86,50	MN428524
CB35	Serum	2018-03-05	2018-03-06	24	F	MT	Cuiaba	25,24	4454	89,40	MN428525
CB36	Serum	2018-03-06	2018-03-06	16	F	MT	Cuiaba	21,42	8795	86,40	MN428526
CB37	Serum	2018-03-05	2018-03-06	23	F	MT	Cuiaba	26,95	3957	89,40	MN428527

Ct = RT-qPCR quantification cycle threshold value.

https://doi.org/10.1371/journal.pntd.0009290.t001

Phylogenetic and Bayesian analysis

The 24 new sequences reported in this study were initially submitted to a genotyping analysis using the phylogenetic arbovirus subtyping tool, available at http://genomedetective.com/app/typingtool/chikungunya [14]. After excluding low-quality genomes, > 10% of ambiguous positions, the newly genomic data generated in this study were aligned with 89 CHIKV-ECSA genome sequences from Brazil plus all available CHIKV-ECSA genome sequences from Americas (currently available only from Paraguay n = 5 and Haiti n = 2). Full details of the sequences used are provided in (**S1 Table**). Sequences were aligned using MAFFT [15] and edited using AliView [16]. The dataset was assessed for presence of phylogenetic signal by applying the likelihood mapping analysis implemented in the IQ-TREE 1.6.8 software [17]. A Maximum likelihood phylogeny was reconstructed using IQ-TREE 1.6.8 software under the HKY+G4 substitution model [17]. In order to investigate the temporal signal in our CHIKV-ECSA dataset we regressed root-to-tip genetic distances from this ML tree against sample collection dates using TempEst v 1.5.1 [18]. The ML phylogeny was used as a starting tree for Bayesian time-scaled phylogenetic analysis using BEAST 1.10.4 [19]. We employed a stringent model selection analysis using both path-sampling and steppingstone models to estimate the most appropriate model combination for Bayesian phylogenetic analysis [20]. The best fitting model was the HKY+G4 substitution model with a Bayesian skyline coalescent model [21, 22]. A discrete phylogeographical model *[23]* was also used to reconstruct the spatial diffusion of

the virus across the compiled dataset sampling locations. Phylogeographic analyses were performed by applying an asymmetric model of location transitioning, coupled with the Bayesian stochastic search variable selection (BSSVS) procedure. Monte Carlo Markov chains (MCMC) were run for sufficiently long to ensure stationarity and an adequate effective sample size (ESS) of >200.

Epidemiological data assembly

Data of weekly notified CHIKV cases in Brazil, availables at the Sistema de Informação de agravos de notificação (SINAN) (https://portalsinan.saude.gov.br/dados-epidemiologicos-sinan), were supplied by Brazilian Ministry of Health and were plotted using the R software version 3.5.1.

Results

We obtained CHIKV RT-qPCR positive clinical samples as part of the genomic surveillance project called ZIBRA 2 (https://www.zibra2project.org/), which aimed to perform, from a lab on wheels, genome sequencing of arboviruses circulating in Midwest Brazil, in the Mato Grosso state. Supported by the Brazilian Ministry of Health and PAHO/WHO, ZIBRA 2 has carried out genomic surveillance of arboviruses in the northern, northeastern and southeastern regions of Brazil [5, 6, 24]. We used the portable MinION sequencer and an amplicon approach [12] to generate 24 partial and near complete CHIKV genomes from serum samples provided by the Central Public Health Laboratory and collected during the 2018 outbreak in the midwestern state of Mato Grosso (MT). These samples, most of which were collected in March 2018, had an average Ct value of 24.48 (ranging from 18.59 to 31.21) and were from patients with average 29 years of age, and in their majority were female (87.5%) living in the city of Cuiaba, MT (**Table 1 and Fig 1A**).

The 24 new CHIKV genome sequences from MT had a mean genome coverage of 81.79% (coverage range 43.7%-93.6%). This genome coverage obtained is considered sufficient to perform phylogenetic inferences, according to a study that showed the occurrence of a decrease in phylogenetic accuracy when genome coverage is reduced from 40% to 20% [25].

Genomic data obtained in this study belonged to the ECSA lineage, as confirmed by the chikungunya virus typing tool (https://www.genomedetective.com/) [14], and clustered with other Brazilian sequences from previous outbreaks reported in other geographic regions (**Fig 1C**). These new genome sequences were submitted to GenBank under the accession numbers MN428504-MN428527.

Fig 1B shows the CHIKV weekly cases normalized per 100K individuals notified between 2015 and 2020 (until epidemiological week–EW 06) in five Brazilian regions: Southeast (SE), Northeast (NE), Midwest (MW), North (N), and South (S). Five CHIKV epidemic waves were found (2016–2020), characterized by a reduction in total cases per year from 2018 to 2020. Although without further information it is difficult to assert the drivers of this reduction in incidence, it is likely that significant herd-immunity has accumulated since CHIKV's introduction, as suggested in other studies [26, 27]. Weekly reported incidence reveals that chikungunya was mostly reported in the Northeast region between 2015 and 2018, the South region had the lowest incidence in the entire time period, and that contrary to all other regions, the Southeast has presented an increasing incidence trend over the years (**Fig 1B**).

To explore in more detail the evolutionary relationship of these new MT sequences in the Brazilian context and to infer the ancestral location of CHIKV strains circulating in MT, we used a Bayesian discrete phylogeographic approach, employing the uncorrelated relaxed molecular clock and a Bayesian Skyline coalescent model (and a linear regression of root-to-

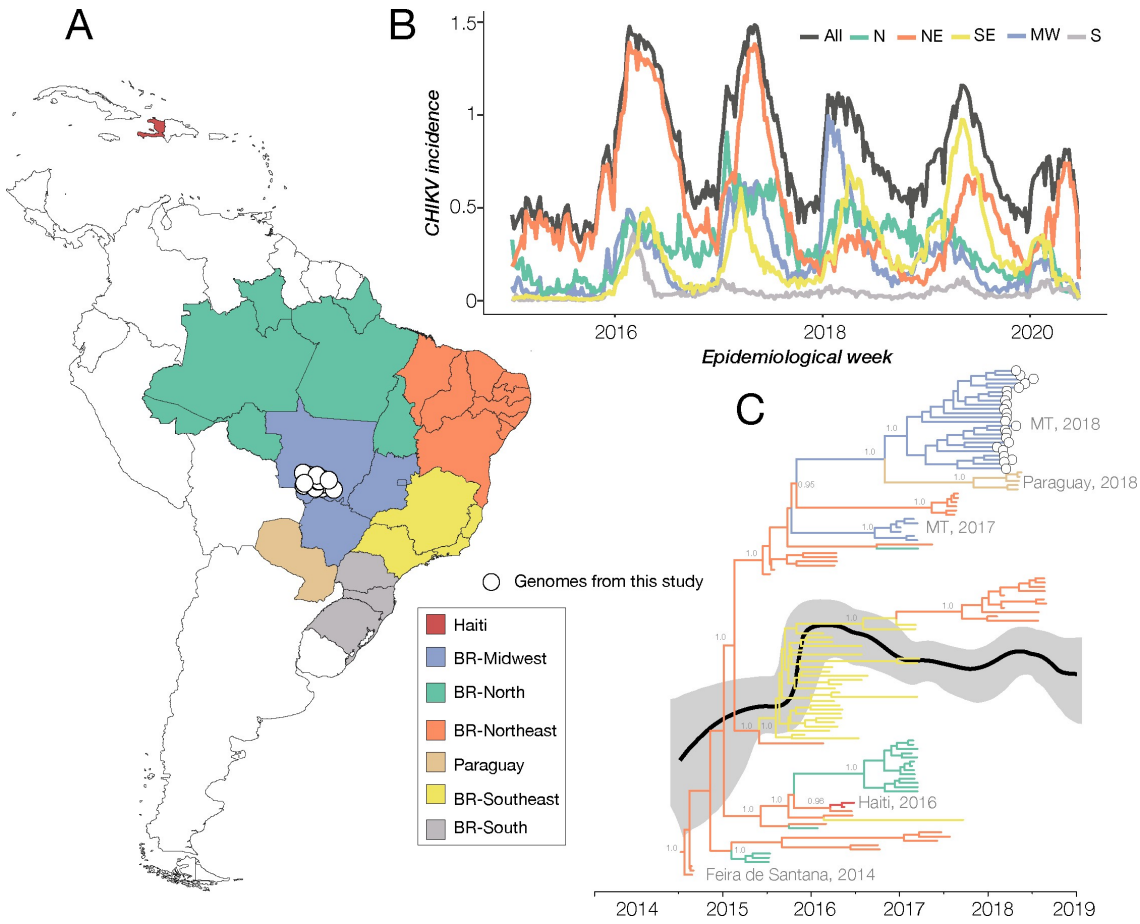


Fig 1. CHIKV transmission dynamics in Brazil (2014–2019). (A) Map of Brazil and Americas showing sampling location of the CHIKV genomes from this study (white circles). (B) Weekly notified chikungunya cases normalized per 100K individuals per Brazilian region in 2015–2020 (until EW06). Incidence (cases per 100K population) is presented in log10 for visual purposes. Epidemic curves are coloured according to geographical macro region: SE = Southeast, NE = Northeast, MW = Midwest, N = North, S = South. (C) Time-scaled phylogeographic tree of 120 complete and near-complete CHIKV genome sequences from the ECSA genotype sampled in Brazil and Americas. Colours represent different sampling locations according to the legend on the left of the tree. Tip circles (white) represent the genome sequences generated in this study. Skyline plot (black and gray lines) is superimposed. Relative genetic diversity is represented here as a surrogate for the product of effective population size and generation time. The solid black line represents the mean relative genetic diversity and the light gray area around the line represents the 95% HPD interval of the estimates.

https://doi.org/10.1371/journal.pntd.0009290.g001

tip genetic distance against sampling date which revealed sufficient temporal signal, r2 = 0.70; S1 Fig) in a dataset comprising the new MT sequences described in this study plus 96 published CHIKV genomes (Fig 1C). The Bayesian demographic reconstructions (Fig 1C, superimposed to the tree) provided evidence of seasonal oscillations, although with wide credible intervals, highlighting a slight but gradual decline of the median Effective population size (Ne) estimated over 2018–2020 which appears to be in agreement with decreasing incidence in time (Fig 1B).

Our Maximum clade credibility (MCC) tree showed that the new 2018 MT isolates formed a single well-supported clade, with posterior support of 1.0 (Fig 1C). Interestingly, these new isolates did not group with the other six previously published sequences sampled in 2017 also from Cuiabá-MT [28]. As these 2017 MT sequences also formed a well-supported clade (posterior support = 1.0), our MCC tree topology suggests that at least two independent introductions of the ECSA lineage occurred in MT. A previous study reported the circulation of CHIKV ECSA lineage in MT and indicated, by RT-qPCR, that the earliest human case of the disease recorded in that state dates to July 2015 [8]. From our time-measured tree, we estimated the time of the most recent common ancestor (TMRCA) of the two independent introduction events in MT to be between late September 2016 (May 2016 to December 2016, 95% HPD) for the first introduction event, and early February 2017 (September 2016 to June 2017, 95% HPD) for the second event. This interval does not include the earliest CHIKV positive case previously reported in MT and this divergence might display an absence of sufficient data because of non-sampling of earliest isolates [8].

The phylogeographic analysis showed that CHIKV was most likely introduced in MT from a dispersion event originating in the northeastern region of Brazil (location probability 0.98). Clades comprising other isolates from 2017–2018 and sampled in the northeastern region illustrate the persistence and re-emergence of the ECSA lineage in the northeastern region of Brazil since its introduction in 2014. In addition, our tree shows the 2018 MT outbreak clade is closely related to the clade containing isolates from Paraguay sampled in 2018, when 1,237 CHIKV cases were reported in that country [1]. We estimate that the TMRCA of the isolates from Paraguay dates back to December 2017 (June 2017 to May 2018, 95% HPD) and probably originated in the midwestern region of Brazil (location probability 0.90). This would be the second event of cross border transmission from Brazil, as our tree also shows isolates from Haiti sampled in 2016 clustered with an isolate from Northeast Brazil also sampled in 2016 (posterior support = 0.97, location probability 0.97). These results indicate a possible role of Brazil as a source for cross-border dispersion of the CHIKV ECSA lineage to other countries in the Americas since its introduction into the country.

Discussion

More than 930,000 cases have been notified since CHIKV was first detected in Brazil in 2014 (4). Despite this large burden of disease, much is unknown about the origins of the virus responsible for the Brazilian outbreaks. To get more insight regarding CHIKV dispersion through different Brazilian regions and South American countries we generated 24 partial and near complete genome sequences from the 2018-CHIKV-ECSA epidemic registered in the state of Mato Grosso (MT), Midwest of Brazil, using a combined strategy of a mobile sequencing mission through this region, genomic, and epidemiological analysis.

Epidemiological data revealed yearly patterns of CHIKV transmission with a reduction in the number of reported cases over 2018–2020, likely a consequence of an expected, gradual accumulation of herd-immunity over the 7 years since its introduction in 2014. Phylogeographic reconstructions suggest that at least two independent introductions of the ECSA lineage occurred in MT from a dispersion event originating in the northeastern region of Brazil and estimated the time of the most recent common ancestor (TMRCA) to be between late September 2016 for the first introduction event, and early February 2017 for the second event. Furthermore, our analysis suggests that the midwestern Brazilian region appears to have acted as a source of virus transmission towards Paraguay, a bordering South American country to the Brazilian Midwest.

In summary, our data reveals a complex pattern of CHIKV transmission between epidemic seasons and sampled locations and suggests that Brazil has played a role as source for

international dispersion (enhanced by cross-border transmission to other Americas countries such as Paraguay and Haiti) of the CHIKV-ECSA genotype to other American countries. Those results highlight the utility of combining genomic, epidemiological and evolutionary methods to understand ongoing mosquito-borne epidemics. Our analyses further indicate that additional data is required to better identify routes of CHIKV-ECSA genotype transmission into Brazil, and to understand its transmission dynamics through other American countries.

Supporting information

S1 Fig. Analysis of temporal structure in CHIKV-ECSA genotype. Root-to-tip genetic divergence against time of sampling. Colours represent different sampling locations.
(PDF)

S1 Table. Full details of the 96 complete and near-complete CHIKV genome sequences from the ECSA genotype samples in Brazil and Americas used in this study.
(DOCX)

Acknowledgments

We thank all personnel from Health Surveillance System from the Mato Grosso and Brazilian Ministry of Health that helped with samples, sources, epidemiological data collection and transportation of reagents.

Author Contributions

Conceptualization: Elaine Cristina de Oliveira, Luiz Carlos Junior Alcantara, Marta Giovanetti.

Data curation: Elaine Cristina de Oliveira, Joilson Xavier, Eduardo Marques Macario, Ana Elisa Viniski, Claudio Luis Campos Souza, Evanil Sebastiana Gomes da Costa, Claudia Soares de Sousa, Flávia Guimarães Dias Duarte, Arnaldo Correia de Medeiros, Rivaldo Venancio Cunha, Noely Fabiana Oliveira De Moura, Tulio de Oliveira, José Lourenço, André Luiz de Abreu, Luiz Carlos Junior Alcantara, Marta Giovanetti.

Formal analysis: Elaine Cristina de Oliveira, Vagner Fonseca, Joilson Xavier, Talita Adelino, Ingra Morales Claro, Allison Fabri, José Lourenço, André Luiz de Abreu, Marta Giovanetti.

Funding acquisition: Luiz Carlos Junior Alcantara.

Investigation: Elaine Cristina de Oliveira, Vagner Fonseca, Joilson Xavier, Talita Adelino, Ingra Morales Claro, Allison Fabri, Claudio Luis Campos Souza, Evanil Sebastiana Gomes da Costa, Claudia Soares de Sousa, Flávia Guimarães Dias Duarte, Arnaldo Correia de Medeiros, Carlos F. Campelo de Albuquerque, Rivaldo Venancio Cunha, Ana Maria Bispo de Filippis, Tulio de Oliveira, José Lourenço, Marta Giovanetti.

Methodology: Elaine Cristina de Oliveira, Vagner Fonseca, Talita Adelino, Ingra Morales Claro, Allison Fabri, Rivaldo Venancio Cunha, José Lourenço, Marta Giovanetti.

Resources: Elaine Cristina de Oliveira, Eduardo Marques Macario, Claudia Soares de Sousa, Flávia Guimarães Dias Duarte, Arnaldo Correia de Medeiros, Carlos F. Campelo de Albuquerque, Noely Fabiana Oliveira De Moura, André Luiz de Abreu, Luiz Carlos Junior Alcantara.

Validation: Vagner Fonseca, Ana Elisa Viniski, José Lourenço, André Luiz de Abreu, Marta Giovanetti.

Writing – original draft: Luiz Carlos Junior Alcantara.

Writing – review & editing: Elaine Cristina de Oliveira, Vagner Fonseca, Joilson Xavier, Eduardo Marques Macario, Carlos F. Campelo de Albuquerque, Rivaldo Venancio Cunha, Ana Maria Bispo de Filippis, Tulio de Oliveira, José Lourenço, Luiz Carlos Junior Alcantara, Marta Giovanetti.

References

1. Pan American Health Organization / World Health Organization. PAHO/WHO Data—Weekly Report | PAHO/WHO [Internet]. Pan American Health Organization / World Health Organization. 2019 [cited 2020 Jul 26]. Available from: https://www.paho.org/data/index.php/en/mnu-topics/chikv-en/550-chikv-weekly-en.html
2. Cunha MS, Costa PAG, Correa IA, de Souza MRM, Calil PT, da Silva GPD., et al. 2020. Chikungunya Virus: An Emergent Arbovirus to the South American Continent and a Continuous Threat to the World. Front Microbiol 2020; 10:33–89. https://doi.org/10.3389/fmicb.2020.01297 PMID: 32670231
3. Bustos Carrillo F, Collado D, Sanchez N, Ojeda S, Lopez Mercado B, Burger-Calderon R, et al. Epidemiological Evidence for Lineage-Specific Differences in the Risk of Inapparent Chikungunya Virus Infection. J Virol 2019; 15:9–4. https://doi.org/10.1128/JVI.01622-18 PMID: 30463967
4. Nunes MRT, Faria NR, de Vasconcelos JM, Golding N, Kraemer MUG, de Oliveira LF, et al. Emergence and potential for spread of Chikungunya virus in Brazil. BMC Med 2015; 30:13–102. https://doi.org/10.1186/s12916-015-0348-x PMID: 25976325
5. Xavier J, Giovanetti M, Fonseca V, Thézé J, Gräf T, Fabri A, et al. Circulation of chikungunya virus East/Central/South African lineage in Rio de Janeiro, Brazil. PLOS ONE. 2019; 2:17–71. https://doi.org/10.1371/journal.pone.0217871 PMID: 31185030
6. Naveca FG, Claro I, Giovanetti M, de Jesus JG, Xavier J, Iani FC de M., et al. Genomic, epidemiological and digital surveillance of Chikungunya virus in the Brazilian Amazon. PLoS Negl Trop Dis 2019; 13:70–79. https://doi.org/10.1371/journal.pntd.0007065 PMID: 30845267
7. Pereira Gusmão Maia Z., Mota Pereira F, do Carmo Said RF, et al. 2019. Return of the founder Chikungunya virus to its place of introduction into Brazil is revealed by genomic characterization of exanthematic disease cases. Emerg Microbes Infect. 9:53–57. https://doi.org/10.1080/22221751.2019.1701954 PMID: 31880218
8. de Souza Costa MC, Siqueira Maia LM, de Souza VC, Gonzaga AM, de Azevedo VC, Ramos Martins L, et al. Arbovirus investigation in patients from Mato Grosso during Zika and Chikungunya virus introduction in Brazil, 2015–2016. Acta Trop 2019; 190:395–402. https://doi.org/10.1016/j.actatropica.2018.12.019 PMID: 30552880
9. Secretaria de Vigilância em Saúde. Boletim Epidemiológico—Monitoramento dos casos de dengue, febre de chikungunya e doença aguda pelo vírus Zika até a Semana Epidemiológica 52 de 2018 [Internet]. Vol. 50. Ministério da Saúde; 2019 Jan [cited 2020 Jul 26]. Report No.: 4.
10. Lanciotti RS, Kosoy OL, Laven JJ, Panella AJ, Velez JO, Lambert AJ, et al. Chikungunya virus in US travelers returning from India, 2006. Emerg Infect Dis 2007; 13:764–7. https://doi.org/10.3201/eid1305.070015 PMID: 17553261
11. Faria NR, Quick J, Claro IM, Theze J, de Jesus JG, Giovanetti M, et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature 2017; 546:406–10. https://doi.org/10.1038/nature22401 PMID: 28538727
12. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc 2017; 12:1261–76. https://doi.org/10.1038/nprot.2017.066 PMID: 28538739
13. Vilsker M, Moosa Y, Nooij S, Fonseca V, et al. Genome Detective: an automated system for virus identification from high-throughput sequencing data. Bioinformatics 2019; 35: 871–3. https://doi.org/10.1093/bioinformatics/bty695 PMID: 30124794
14. Fonseca V, Libin PJK, Theys K, et al. A computational method for the identification of dengue, zika and chikungunya virus species and genotypes. PLoS Negl Trop Dis 2019; 13:1–15.
15. Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 2005; 33:511–8. https://doi.org/10.1093/nar/gki198 PMID: 15661851
16. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large data sets. Bioinformatics 2014; 30:22–2.

17. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 2015; 32:268–74. https://doi.org/10.1093/molbev/msu300 PMID: 25371430

18. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol 2016; 2:1–44. https://doi.org/10.1093/ve/vew007 PMID: 27774300

19. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 2018; 4:1–5. https://doi.org/10.1093/ve/vey016 PMID: 29942656

20. Baele G., et al., Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. Mol Biol Evol 2013; 30:2–43. https://doi.org/10.1093/molbev/mss243 PMID: 23090976

21. Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. Mol Biol Evol 2013; 30:713–24. https://doi.org/10.1093/molbev/mss265 PMID: 23180580

22. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 2012; 29(7):8–73. https://doi.org/10.1093/molbev/mss075 PMID: 22367748

23. Lemey P, Rambaut A, Drummond AJ. Bayesian phylogeography finds its roots. PLoS Comput Biol 2009; 5(1):12–87. https://doi.org/10.1371/journal.pcbi.1000520 PMID: 19779555

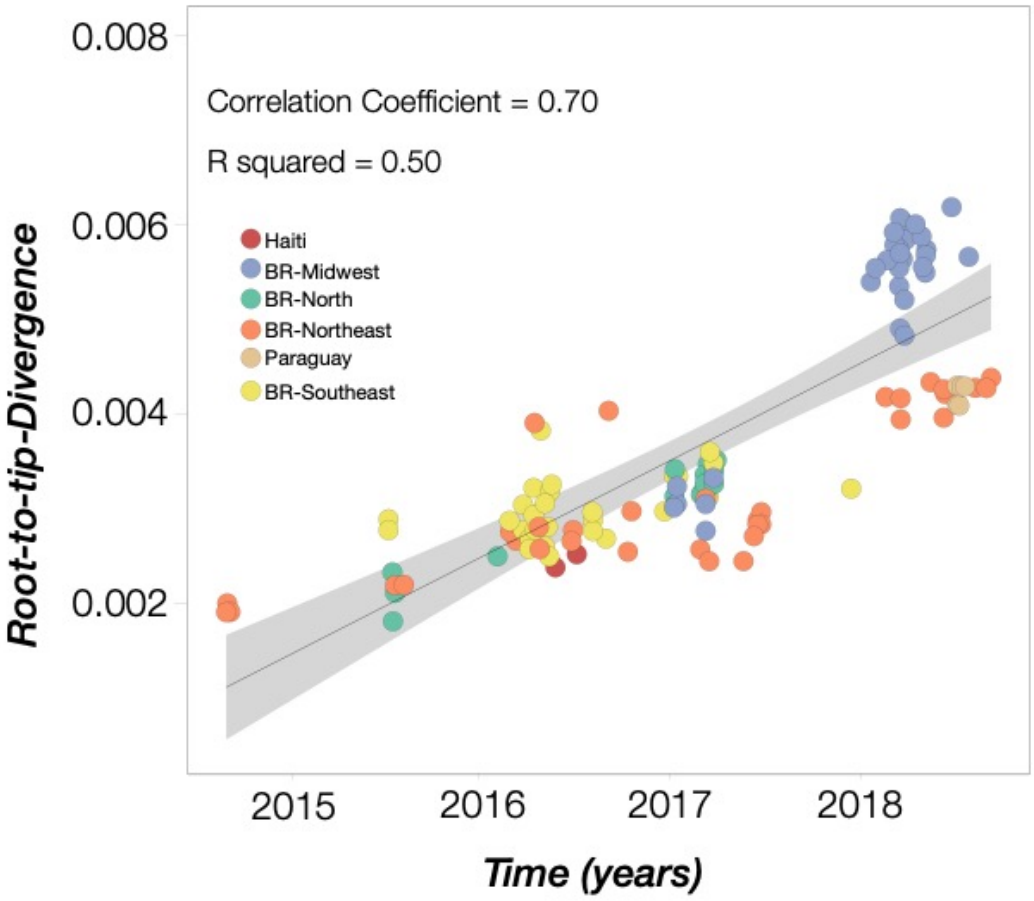
24. Faria NR, Quick J, Claro IM, Thézé J, de Jesus JG, Giovanetti M, et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature 2017; 15(2):546–606. https://doi.org/10.1038/nature22401 PMID: 28538727

25. Thézé J, Li T, du Plessis L, Bouquet J, Kraemer MUG, Somasekar S, et al. Genomic Epidemiology Reconstructs the Introduction and Spread of Zika Virus in Central America and Mexico. Cell Host Microbe 2018; 23(6):855–64. https://doi.org/10.1016/j.chom.2018.04.017 PMID: 29805095

26. Ribeiro GS, Hamer GL, Diallo M, Kitron U, Ko AI, Weaver SC. Influence of herd immunity in the cyclical nature of arboviruses. Curr Opin Virol 2020; 40(2):1–10. https://doi.org/10.1016/j.coviro.2020.02.004 PMID: 32193135

27. Lourenço J, Maia de Lima M, Faria NR, et al. Epidemiological and ecological determinants of Zika virus transmission in an urban setting. Elife 2017; 9(1):6–29.

28. Vasconcellos AF, Silva JMF, de Oliveira AS, Prado PS, Nagata T, Resende RO. Genome sequences of chikungunya virus isolates circulating in midwestern Brazil. Arch Virol 2019; 4(1):1205–8. https://doi.org/10.1007/s00705-019-04174-4 PMID: 30729309



S1 Fig. Analysis of temporal structure in CHIKV-ECSA genotype. Root-to-tip genetic divergence against time of sampling. Colours represent different sampling locations.

nature communications

Article

https://doi.org/10.1038/s41467-023-40099-y

Increased interregional virus exchange and nucleotide diversity outline the expansion of chikungunya virus in Brazil

Received: 20 April 2023

Accepted: 12 July 2023

Published online: 21 July 2023

Check for updates

A list of authors and their affiliations appears at the end of the paper

The emergence and reemergence of mosquito-borne diseases in Brazil such as yellow fever, zika, chikungunya, and dengue have had serious impacts on public health. Concerns have been raised due to the rapid dissemination of the chikungunya virus across the country since its first detection in 2014 in Northeast Brazil. In this work, we carried out on-site training activities in genomic surveillance in partnership with the National Network of Public Health Laboratories that have led to the generation of 422 chikungunya virus genomes from 12 Brazilian states over the past two years (2021–2022), a period that has seen more than 312 thousand chikungunya fever cases reported in the country. These genomes increased the amount of available data and allowed a more comprehensive characterization of the dispersal dynamics of the chikungunya virus East-Central-South-African lineage in Brazil. Tree branching patterns revealed the emergence and expansion of two distinct subclades. Phylogeographic analysis indicated that the northeast region has been the leading hub of virus spread towards other regions. Increased frequency of C > T transitions among the new genomes suggested that host restriction factors from the immune system such as ADAR and AID/APOBEC deaminases might be driving the genetic diversity of the chikungunya virus in Brazil.

Mosquito-borne viral diseases have impacted the lives of millions of people across several populations in the tropics and subtropics[1,2]. This scenario prompted the World Health Organization (WHO) to issue a guideline in 2017 with strategies for global vector control response aiming to reduce the burden and threat of vector-borne diseases, such as Dengue, Zika, and chikungunya fever, by 2030[3]. Five years later, considering the public health implications brought by the coronavirus disease (COVID-19) pandemic, the WHO outlined ten proposals to strengthen health emergency preparedness, response, and resilience[4]. One of the proposals calls for the development and establishment of a collaborative surveillance system with improved laboratory capacity for pathogen and genomic surveillance that would guide the public health response.

Several genomic surveillance initiatives have been carried out in the last few years to build knowledge regarding the genetic diversity and transmission dynamics of arboviruses in Brazil, such as chikungunya virus (CHIKV)[5–12]. CHIKV infection can cause long-lasting effects such as debilitating arthritis and arthralgia, and there are no effective treatments available[13]. Vaccine candidates in development have reached phase 2 and 3 clinical trials using attenuated virus derived from the Indian Ocean lineage (IOL) or virus-like particle containing recombinant structural proteins derived from a Senegalese viral strain[14,15]. Genomic sequencing has revealed that the first case of chikungunya fever (CHIKF) reported in Brazil was an infection by the Asian lineage introduced in a northern state in 2014, while another case reported seven days later in a north-eastern state represented the first known introduction of the East-Central-South-African (ECSA) lineage in the country[12]. The establishment of the CHIKV in the Brazilian territory was followed by several outbreaks reported across the country, accounting for more than 200 thousand confirmed cases in only the last two years[16]. Viral genomic data from Brazilian cases has revealed that the ECSA lineage is widespread throughout the country[17] and has

e-mail: luiz.alcantara@ioc.fiocruz.br; ana.bispo@ioc.fiocruz.br; giovanetti.marta@gmail.com

been linked to fatal cases observed in both risk and non-risk groups (young adults and people without comorbidities)[18].

Viral genomic surveillance activities have been driven by the rapid development of DNA sequencing technology and bioinformatics tools for genomic data analysis[19]. Such tools have allowed the characterization of the genome and dispersal patterns of emerging and reemerging pathogens[6,9,20]. The use of such tools during the COVID-19 pandemic allowed, for example, the rapid identification of emerging mutations likely associated with increased transmissibility and immune escape[21]. Despite technological advances and the high number of CHIKF cases reported in recent years in Brazil, the amount of genomic data available in public databases has consisted of genomes from localized outbreaks that could be limited in terms of representativeness across different states and outbreak events. By generating CHIKV sequence data, extensive genomic surveillance efforts in Brazil can provide updated information on the genetic pool of the viral population circulating in the country, fostering new studies on virus evolution and disease treatment.

Recurring outbreaks demonstrate that CHIKV is currently endemic in Brazil. The existence of abundant vectors, together with adequate climatic conditions for vector survival in areas of high population density, create conditions that can modify the adaptive landscape, allowing the continued expansion and evolutionary adaptation of CHIKV[22–24]. Faced with a scenario of limited availability of genetic information on potential strains causing a rapid increase in the number of CHIKF cases over the past two years in Brazil, in this work, we carried out on-site training activities in genomic surveillance in 12 Brazilian states covering four geographic regions to increase the number of available viral genomic sequences. This has allowed comprehensive monitoring of the expansion of the ECSA lineage and its variants circulating in different states, in addition to the characterization of the most up-to-date structured phylogeny of the CHIKV-ECSA lineage in the country.

Results

East, Central, and South African (ECSA) lineage monitoring through countrywide genomic surveillance

Nanopore sequencing was performed on selected CHIKV-positive samples provided by state public health laboratories from 12 states across 4 geographic regions (Northeast, Midwest, Southeast, and South) of Brazil (Fig. 1a) during the years 2021–2022, which saw a significant increase in the number of CHIKF cases reported across different Brazilian regions, with a peak incidence rate of more than 20 cases per 100,000 population, and a total of more than 312,000 suspected cases reported nationally in this 2-year period (Fig. 1b). Due to the portability and easy setup of the nanopore sequencing protocol, which allows data generation in less than 24 h, the collaborative work with the public health laboratories was able to not only generate genomic data but also promote on-site genomic surveillance training activities for the local laboratory staff. This approach combining wet lab and data analysis training allowed local teams to understand how genomic data can be linked to demographic information in order to produce comprehensive and relevant inferences regarding the epidemiology and evolution of CHIKV circulating in Brazil.

A total of 425 CHIKV-positive samples were subjected to nanopore whole-genome sequencing, with 84.94% ($n = 361$) of these samples originating in the Northeast region, consisting largely of samples collected from the state of Bahia ($n = 102$) (Fig. 1a, Table 1, and Supplementary Fig. 1). These samples presented a mean RT-qPCR cycle threshold (Ct) value of 24.04 (ranging from 11 to 35.90) (Table 2). Patients' mean age upon sample collection was similar for both females and males (39 years of age), with 57.88% ($n = 246$) of the participants identified as female (Table 2). The clinical status of patients at the time of sample collection, and travel history data were not available for these samples.

Multiplex PCR-tiling amplicon sequencing on MinION allowed the recovery of 425 genomic sequences from CHIKV with a mean genome coverage and sequencing depth of 90.98% (range 31.80 to 96.19%) and 3290x (range 58 to 129,706), respectively (Supplementary Data 1). Of the 425 sequences, 14 were recovered from old CHIKV-PCR samples collected in Bahia state during July–August 2015 and stored since. These isolates had a mean genome coverage of 70% (range 31.80 to 92.9%). The remaining genomes have an associated collection date ranging from April 2019 to June 2022. To better capture the phylogenetic signal, only the sequences with genome coverage over 60% ($n = 422$) were considered for further analysis (discarded sequences are listed in the methods).

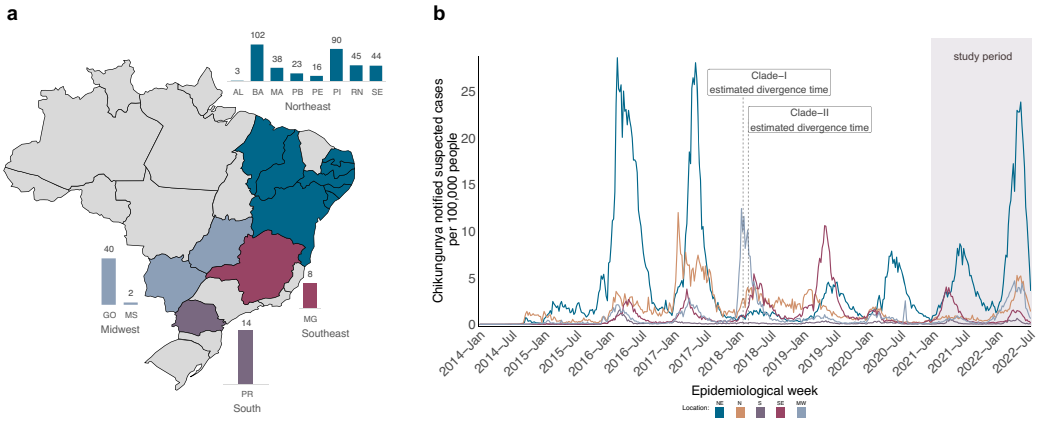


Fig. 1 | Spatiotemporal distribution of chikungunya fever (CHIKF) in Brazil. **a** Map of Brazil displays states colored according to the location where samples were collected and sequenced. Bar plots indicate the number of isolates obtained from each state. States abbreviations: AL = Alagoas, BA = Bahia, MA = Maranhão, PB = Paraíba, PE = Pernambuco, PI = Piauí, RN = Rio Grande do Norte, SE = Sergipe, MG = Minas Gerais, PR = Paraná, GO = Goiás, MS = Mato Grosso do Sul. **b** Time series of monthly reported CHIKF suspected cases normalized per 100 K individuals in five Brazilian macroregions over 2014–2022 (until epidemiological week 28). Epidemic curves are colored according to geographical macroregion: N = North, NE = Northeast, MW = Midwest, SE = Southeast, S = South. The shaded rectangle indicates the period in which samples were collected for this study. See "Data availability" for Source data.

 https://doi.org/10.1038/s41467-023-40099-y

Table 1 | Number of chikungunya virus-positive samples sequenced during the genomic surveillance in Brazil, 2021–2022, by geographical origin

Region and State	# Samples (total *n* = 425)
Northeast	(*n* = 361)
Alagoas (AL)	3 (0.71%)
Bahia (BA)	102 (24.00%)
Maranhão (MA)	38 (8.94%)
Paraíba (PB)	23 (5.41%)
Pernambuco (PE)	16 (3.77%)
Piauí (PI)	90 (21.18%)
Rio Grande do Norte (RN)	45 (10.59%)
Sergipe (SE)	44 (10.35%)
Midwest	(*n* = 42)
Goiás (GO)	40 (9.41%)
Mato Grosso do Sul (MS)	2 (0.47%)
Southeast	
Minas Gerais (MG)	8 (1.88%)
South	
Paraná (PR)	14 (3.29%)

Table 2 | Demographic and laboratory characteristics of chikungunya virus-infected patients

	Patients (*n* = 425)
Sex	
Female	246 (57.88%)
Male	179 (42.12%)
Mean age on sample collection (SD)	
Female	39.10 ± 18.83
Male	39.37 ± 22.03
Median days of symptoms prior to sampling (SD)	3.47 ± 8.52
RT-qPCR mean Ct (cycle threshold) value (SD)	24.04 ± 4.15

SD standard deviation.

All the newly recovered genomes were assembled using Genome Detective software which also classified all of them as belonging to the East, Central, and South African (ECSA) lineage. To investigate the phylogenetic relationship of the new sequences with other Brazilian and non-Brazilian sequences available in public databases, we built a global dataset (*n* = 1987) composed of 1565 CHIKV genomes retrieved from GenBank NCBI in addition to 422 sequences from this study. It is noteworthy that the two years of genomic surveillance activities of this study contributed to a more than doubling of the number of CHIKV genomes from Brazil available in the NCBI (by then there were 332 complete sequences) since the virus emerged in the country in 2014 (Fig. 2c).

Updated CHIKV phylogeny reveals two distinct emerging subclades

A preliminary Maximum Likelihood (ML) tree was reconstructed using the global dataset that showed all Brazilian sequences grouped in the ECSA clade (Supplementary Fig. 2). It can also be observed from the ML tree that most of the new sequences formed two well-distinct derived clades, where it can be noticed that sequences collected in different geographical regions were closely related (see ML tree in Supplementary Data 2). To investigate in more detail the phylogenetic features of these clades within a time-aware evolutionary framework, which can benefit from the increased amount of genomic data obtained in this study, we performed a Bayesian phylogenetic analysis using a downsampled dataset (*n* = 713) mostly composed of Brazilian sequences.

Root-to-tip genetic distance regression indicated that the downsampled dataset presented sufficient temporal signal ($R^2 = 0.70$ and correlation coefficient = 0.83) to infer a time-measured phylogeny (Fig. 2b). Consistent with the ML tree, the inferred Maximum Clade Credibility (MCC) tree also revealed two distinct more derived clades (henceforth clade I and II) formed mainly by 2021–2022 sequences (Fig. 2a). The Bayesian evolutionary analysis estimated the time of the most recent common ancestor (tMRCA) of clade I to be late January 2018 (95% highest posterior density (HPD): December 2017 and March 2018), while clade II presented a slightly late tMRCA estimated to be early February 2018 (95% HPD: January 2018 and March 2018).

Some composition differences can be noticed between these clades. Clade I comprises sequences from 14 distinct states mostly collected from 2021 to 2022 (*n* = 304) and from northeastern Brazil (62.38%, *n* = 204), with sequences from Sergipe (13.5%, Northeast), Minas Gerais (1.8%), São Paulo (16.2%, Southeast), Goiás (11.6%), Mato Grosso do Sul (0.6%, Midwest), and Paraná (4.3%, South) states uniquely present in this clade (Fig. 2a). Meanwhile, clade II is mostly composed of sequences collected in 2022 from northeastern states (87.5%, *n* = 147), with a total of 8 states sampled and Piaui state being the most represented (46.42%, *n* = 78) (Fig. 2a).

A closer look at sequence distribution inside these clades reveals recurrent virus movement between states and regions, with midwestern isolates closely related to isolates from the Northeast and from the state of Minas Gerais (Southeast) in clade I. It can also be noticed in clade I that several distinct CHIKV introductions occurred in the state of Goiás (Midwest) and in northeastern states (Bahia, Rio Grande do Norte, Sergipe, Paraiba, Pernambuco, and Piaui) (Fig. 2a). Contrarily, the clade I sequence distribution reveals that apparently, only one viral introduction event has happened in the southern state of Paraná, sharing a most recent common ancestor (dated from Jun. 2019 to Jun. 2020, 95% HPD, posterior probability = 1.0) with isolates from the state of São Paulo (Southeast cluster). Similarly, clade II displays viral exchange between states, especially from the Northeast, as indicated by a single well-supported subclade (posterior = 1.0), dated from Jan. 2020 to Dec. 2020 (95% HPD) and dominated by sequences from northeastern states (Piauí, Bahia, Rio Grande do Norte, Paraíba, Pernambuco, Maranhão, and Alagoas) (Fig. 2a).

CHIKV dispersal in Brazil has been mainly seeded by the Northeast

In face of the recurrent virus movement observed across the country, as indicated by our MCC phylogeny, we employed a Bayesian phylogeographic approach to reconstruct the spatial dispersal dynamic of CHIKV in Brazil (closely related sequences from Haiti and Paraguay were included) and to estimate the ancestral locations of clades I and II. The resulting phylogeny kept the topology from the MCC tree shown in Fig. 2a and revealed the Northeast as a leading source of CHIKV transmission in Brazil, seeding the network of frequent virus exchange among states mainly from the Northeast, Southeast, and Midwest, as indicated by the location probability of the 5 early branching events inferred by the discrete phylogeography (Fig. 2d and Supplementary Fig. 3). Moreover, the ECSA lineage circulating in Brazil extended its transmission network by reaching other countries in the region such as Paraguay and Haiti, likely via the Midwest and Northeast of Brazil, respectively. An alternative approach, using a transmission network generated from transition states summarized from the Bayesian phylogeography and centrality metrics, also indicated the Northeast as a source (Source Hub Ratio of 0.66) in the network for the CHIKV spread in the country, where intense interactions are displayed between the Northeast and Southeast (Supplementary Fig. 4 and Supplementary Table 1). Moreover, the discrete state ancestral reconstruction

 https://doi.org/10.1038/s41467-023-40099-y

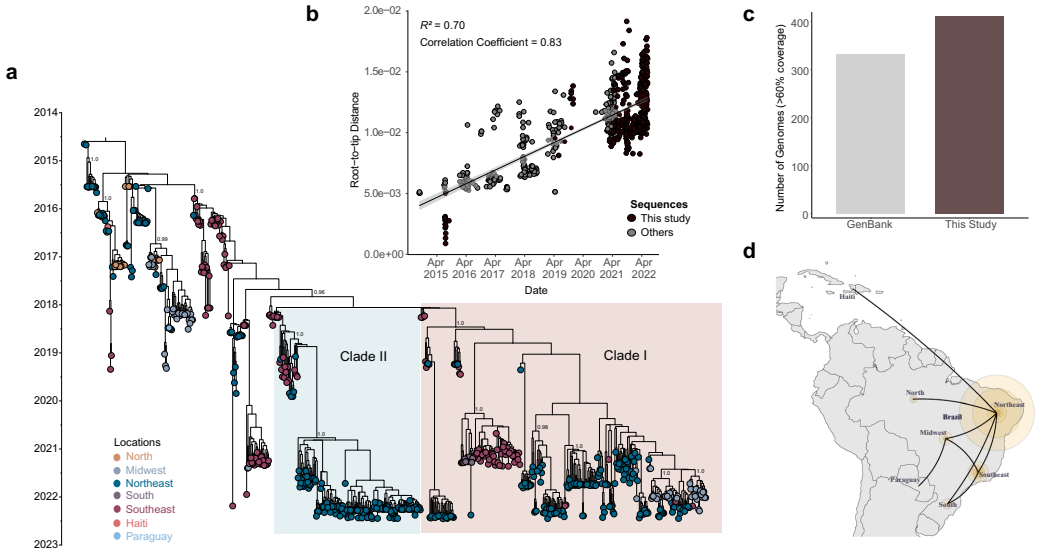


Fig. 2 | Time-measured phylogeny of chikungunya virus ECSA lineage in Brazil. **a** Maximum Clade Credibility tree reconstructed using 706 sequences from Brazil (in addition to 5 sequences from Paraguay and 2 from Haiti) and a molecular clock approach. Numbers in black show clade posterior probabilities of main nodes. Some posterior probability values were omitted for clarity. Tip colors represent the sampling location. **b** Root-to-tip genetic distance regression in a maximum likelihood phylogeny of the CHIKV ECSA lineage ($n = 713$). New sequences are colored in black. The black line represents the medium values of the linear regression while the light gray bands around the line represent the standard error. R^2 indicates the coefficient of determination. **c** Number of genomes generated in this study (with >60% genome coverage) compared to the number of Brazilian CHIKV ECSA sequences available on the GenBank up to 27th Jan. 2023. **d** The spatial spread of CHIKV in Brazil estimated, using SPREAD4 software, under a discrete diffusion model employed in the Bayesian Phylogeographic approach using a dataset with 471 sequences. Size of colored circles was scaled by location posterior support. See Data Availability for Source data.

employed in our Bayesian analysis estimated that both clades I and II might have emerged in the Southeast region with a location probability of 1.0 and with estimated divergence times (clade I: 95% HPD Nov. 2017–Feb. 2018; clade II: Dec. 2017–Mar. 2018) (Supplementary Fig. 3) comparable to those obtained in the MCC tree from Fig. 2a inferred using a comprehensive dataset.

These estimates place the divergence time of clades I and II in the period that marks the return of CHIKV-increased transmissions after two main epidemic seasons registered from 2016 to 2017 when a total of more than 565 thousand disease cases were notified in the country (Fig. 1b). From the time series graph of CHIKF cases, we can see an increase in the incidence rate around early 2018 for the Midwest and North regions. In that same period, clades I and II were estimated to emerge in the Southeast region, which also presented an increased incidence rate. Since then, a seasonal epidemic pattern has been observed in the CHIKV transmission dynamic with incidence peaks being displayed in the first months of the following years (Fig. 1b). Although the Northeast displayed a slightly later peak in its incidence curve in 2020 and 2021, but not in 2022, when compared to the other regions. These changes in the incidence curve patterns could be explained by differences in local climatic conditions within regions associated with differences in population susceptibility and immunity to the virus, leading to the differential spread of the virus in other parts of Brazil. Alternatively, it has been suggested that spatial heterogeneity associated to virus spread could determine CHIKV epidemics in urban environments[25]. Moreover, surveillance and reporting biases cannot be ruled out as the 2020–2021 period comprises the time when the states were heavily impacted by rapid increase in the number of SARS-CoV-2 infections, which subsided in 2022.

Consistent with the phylogeography and the transmission network analyses, the CHIKF case time-series plot presents the Northeast region as a major source of virus transmission in the country, as shown by the consecutive incidence peaks registered for that region in the last three years.

Transition changes outline the evolutionary expansion of the Brazilian ECSA lineage

Despite having close divergence times, clades I and II have different sequence compositions that likely explain their apparent branching distance. While sequences from clades I and II were obtained from patients with similar median age, a significant difference ($p < 0.05$) was observed between the median RT-qPCR Ct value of the samples of each clade (Fig. 3a). This difference in the median Ct values, however, might have been caused by an imbalance present in the dataset used for comparison since clade I has almost twice as many sequences compared to clade II. Alternatively, the between-clades inconsistency observed in the time between symptoms onset and sample collection could have affected sample Ct values.

We also assessed these clades as to which selective regime they are likely to be subject to. The results from the BUSTED analysis provided evidence that the envelope polyprotein coding region experienced positive diversifying selection in both clades I ($p = 1.774e{-}8$) and II ($p = 1.840e{-}11$) (Supplementary Table 2). For comparison, we employed a second method, MEME, that identified 12 sites under positive selection for clade I, whereas 13 sites were identified in clade II. Of these sites, 7 are shared by both clades, while exclusive positive selected sites were reported for each clade (Supplementary Table 3). These results suggest the active status of these clades and consequently provide evidence of the continuous evolutionary expansion of the Brazilian ECSA lineage.

Separate sequence alignments representative of each clade revealed a significant difference between the median frequency of 3

https://doi.org/10.1038/s41467-023-40099-y

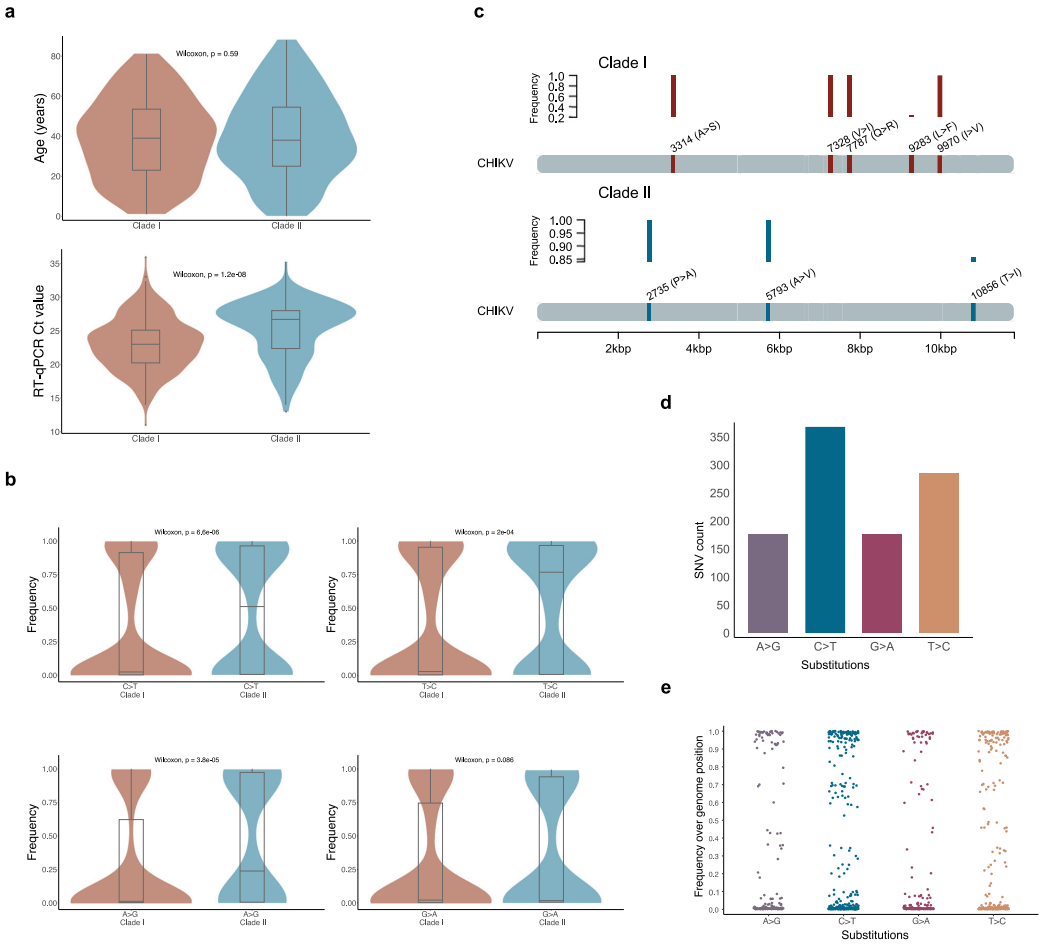


Fig. 3 | Genetic and demographic aspects of the newly generated chikungunya virus (CHIKV) sequences. a Violin plot and boxplot of the patient's age and samples' cycle threshold (Ct) value distributions by clades I ($n = 255$ biologically independent samples) and II ($n = 147$ biologically independent samples). Mann–Whitney U (Wilcoxon) two-sided test with a significance level alpha = 0.05. **b** Violin plots and boxplots of the frequency distributions for each class of single nucleotide variation (SNV) identified across genome positions of sequences from clades I ($n = 327$ biologically independent samples) and II ($n = 168$ biologically independent samples). Mann–Whitney U (Wilcoxon) two-sided test with a significance level alpha = 0.05. **a**, **b** Data visualized as violin plots (outer shape representing the data kernel density) and box-and-whisker plots (box represents the median value and 25 and 75% quartiles, the whiskers display the minimum and maximum values, 1.5x interquartile range; points, outliers). **c** Genomic positions and frequency (represented by plotted bars) of the non-synonymous substitutions uniquely identified in the sequences from clades I and II. Numbers represent the genomic position of each substitution. **d** Bar plot of the absolute number of CHIKV genome mutated positions for each class of single nucleotide variation (SNV) identified in the alignment of the new CHIKV sequences ($n = 422$ biologically independent samples). **e** Scatterplot of the frequency distributions for each class of single nucleotide variation (SNV) identified in the alignment of the new CHIKV sequences ($n = 422$ biologically independent samples). Each dot represents the frequency of a distinct mutated genome position. **b**, **d**, **e** Letters in the x-axis represent nucleotides: A—adenine, G—guanine, C—cytosine, T—thymine. See "Data availability" for Source data.

classes of single nucleotide variation (SNV) identified in the clades. Clade II presented a higher median frequency of SNV of type C > T, A > G, and T > C transitions (Fig. 3b). By comparing the mutational profile of the two clades, we identified 27 SNVs exclusively present and shared by sequences from clade II, with three of them being non-synonymous substitutions leading putatively to amino acid change such as a T288I substitution in the E1 protein (Table 3 and Fig. 3c). The other two non-synonymous substitutions were identified in two non-structural proteins (nsP2-P352A and nsP4-A43V).

Sequences from Clade I, on the other hand, presented 13 exclusive SNVs, of which six are non-synonymous mutations present in two nonstructural proteins (nsP2 and nsP4) and three structural proteins (capsid, 6k and E2) (Table 3 and Fig. 3c). The capsid sequence from this clade presented two contiguous transition substitutions in the same codon (7787-7788) resulting in a Q74R change. Moreover, both clades contain genomes harboring the E1-T98A mutation (121/327 sequences in clade I, and 31/167 sequences in clade II), while only clade II reported 113 sequences (67%) presenting another mutation, the V264A in the E2. These two mutations have been associated with increased virus infectivity in *Aedes ssp.* when present in distinct genetic backgrounds (E1-226V with E1-T98A or E1-226A with E2-V264A)[26,27]. However, the E1-226V mutation, as well as other E1 (K211E) and E2 (D60G, R198Q,

 https://doi.org/10.1038/s41467-023-40099-y

Table 3 | List of single nucleotide variations (SNV) uniquely identified for clades I and II

Genomic position[a]	Reference base	Altered base	Substitution type	aa[b] change	Clade	Genomic region	aa[c] position	Frequency[d]
3208	C	T	S	D > D	1	nsP2	509	1.00
3217	T	C	S	Y > Y	1	nsP2	512	1.00
3277	C	T	S	S > S	1	nsP2	532	1.00
3295	G	A	S	P > P	1	nsP2	538	1.00
3314	**G**	**T**	**NS**	**A > S**	**1**	**nsP2**	**544**	**1.00**
7328	**G**	**A**	**NS**	**V > I**	**1**	**nsP4**	**555**	**1.00**
7787	**A**	**G**	**NS**	**Q > R**	**1**	**capsid**	**74**	**1.00**
7788	**G**	**A**	**NS**	**Q > R**	**1**	**capsid**	**74**	**1.00**
8641	C	T	S	L > L	1	E2	34	0.68
9283	**C**	**T**	**NS**	**L > F**	**1**	**E2**	**248**	**0.22**
9285	C	T	S	L > L	1	E2	248	0.73
9936	T	C	S	C > C	1	6K	42	1.00
9970	**A**	**G**	**NS**	**I > V**	**1**	**6K**	**54**	**1.00**
1754	T	C	S	L > L	2	NSP2	25	1.00
1759	T	A	S	V > V	2	NSP2	26	1.00
1783	T	G	S	R > R	2	NSP2	34	1.00
1789	A	G	S	Q > Q	2	NSP2	36	1.00
1795	T	C	S	L > L	2	NSP2	38	1.00
2735	**C**	**G**	**NS**	**P > A**	**2**	**NSP2**	**352**	**1.00**
2767	A	G	S	G > G	2	NSP2	362	1.00
2887	T	C	S	L > L	2	NSP2	402	1.00
3403	T	C	S	F > F	2	NSP2	574	1.00
3901	A	G	S	V > V	2	NSP2	740	1.00
3916	T	C	S	F > F	2	NSP2	745	1.00
4711	T	C	S	D > D	2	NSP3	212	1.00
4837	A	G	S	S > S	2	NSP3	254	1.00
4843	C	T	S	P > P	2	NSP3	256	1.00
4939	T	C	S	S > S	2	NSP3	288	1.00
5783	C	T	S	L > L	2	NSP4	40	1.00
5793	**C**	**T**	**NS**	**A > V**	**2**	**NSP4**	**43**	**1.00**
5807	C	T	S	L > L	2	NSP4	48	1.00
5857	T	C	S	Y > Y	2	NSP4	64	1.00
5921	T	C	S	L > L	2	NSP4	86	1.00
5956	C	T	S	T > T	2	NSP4	97	1.00
6994	T	C	S	N > N	2	NSP4	443	1.00
7213	A	G	S	T > T	2	NSP4	516	1.00
9153	T	C	S	G > G	2	E2	204	0.89
9642	T	A	S	T > T	2	E2	367	0.80
10856	**C**	**T**	**NS**	**T > I**	**2**	**E1**	**288**	**0.86**
11241	T	C	S	G > G	2	E1	416	0.81

Non-synonymous substitutions are highlighted in bold.
[a]Genomic position according to the reference sequence NC_004162.2.
[b]Amino acid changes.
[c]Amino acid position in the respective peptide.
[d]Frequency of the substitution in the dataset from the respective clade.

L210Q, I211T, K233E, K252Q) adaptative mutations have not been detected in the Brazilian isolates.

In an alignment analysis of all 422 new sequences, we found that among the total of CHIKV genome mutated positions the majority consisted of type C > T transitions (29.5%, $n = 368$) followed by T > C substitutions (22.9%, $n = 285$) (Fig. 3d). These transitions along with A > G and G > A transitions displayed a comparable frequency distribution across the sequence dataset, where mutations at several positions were identified in all new sequences (Fig. 3e). Due to the possibility of these substitutions resulting from sequencing errors, we performed the same comparative analysis on a different dataset containing sequences that form a subclade (from the South region) in clade I and that were generated by a different sequencing technology (Illumina) and a different research group (data in Supplementary files). We observed the same pattern of increased frequency (26.8%) of C > T transitions among those southern sequences, which suggested the observed increased transition changes in clades I and II were likely derived from mutation rather than sequencing artifacts.

Discussion

Advances in sequencing technology and bioinformatics tools have contributed to an increased understanding of the genetic diversity and

transmission dynamics of emerging viruses causing epidemics[28]. In this study, we increased the number of CHIKV genomes deposited in the NCBI by generating 422 genomes that cover 12 Brazilian states over the years 2021–2022. By combining genomic and demographic data this effort has provided a comprehensive overview of CHIKV-ECSA phylogeny in Brazil which summarizes and updates previous phylogenies from other studies of localized outbreaks[10,12,17,29,30]. The updated and metadata-enriched dataset allowed us not only to describe the genetic diversity of circulating CHIKV variants but also to describe branching patterns observed in the updated phylogeny and infer the divergence time and likely location of emerging distinct subclades from the Brazilian lineage.

Following the arrival of ECSA lineage in the Northeast region, a number of other studies have predicted and reported the establishment and spread of this lineage to the rest of the country[10–12,17,20,29–32]. The Bayesian phylogeographic approach employed in this study has revealed the main routes of dispersion of CHIKV in the country. After being introduced in the Northeast region, the ECSA lineage dispersed towards several states from all five regions in Brazil. In line with previous findings[10,17,29,33–35], both the Bayesian phylogeographic and transmission network approaches indicated the Northeast region has been acting as the leading hub of virus spread towards other regions in the country, including by forming an intense virus exchange network with the Southeast region. These interactions are likely driven by human mobility between these areas, as both Southeast and Northeast regions are densely populated, housing transportation infrastructure and major urban centers that attract a substantial number of visitors. From the Southeast, the virus dispersed to the southern state of Paraná (posterior probability of 0.96) with the time of the most recent common ancestor shared with samples from São Paulo estimated to be January 2020. In addition to spreading in Brazil, CHIKV has also extended its circulation to other countries such as Paraguay and Haiti. CHIKV circulation in Paraguay and Haiti has been previously associated with transmission events originating from Brazil through international viral exchange likely mediated by human movement[36–38]. These results evidence the lineage's potential expansion across Latin America, despite limited information available about this lineage in the region as evidenced by the lack of CHIKV-ECSA sequences from other Latin American countries.

Our time-measured phylogeny corroborated previous studies and revealed emerging branching patterns, mainly represented by two well-supported distinct subclades named clade I and II. Both these clades were estimated to emerge in the Southeast region around the first months of 2018. At the end of that year, the Southeast region accounted for more than 65 thousand reported cases. Moreover, differences in sequence composition were observed between these clades, with clade I being more diverse as it comprises sequences from four different regions (Northeast, Midwest, Southeast, and South) collected in the years 2021–2022, while clade II contains mostly sequences from northeastern states collected in 2022. This difference in the clades sequence composition profile might change as the lineage continues to expand into the country. The observed differences in geographic diversity between clades might reflect distinct transmission networks underlying the divergence and expansion of these subclades.

Different viral lineages might be under distinct evolutionary pressure that together with the emergence and selection of mutations can drive viral adaptation to a particular environment[39,40]. It has been argued that two different CHIKV lineages, IOL and Asian, have undergone different evolutionary trajectories leading to different vector adaptative potentials[41]. Here, we used the ratio of non-synonymous (*dN*) to synonymous (*dS*) nucleotide substitutions in the CHIKV envelope gene to assess the selective regime to which clades I and II might be subjected. Our analysis revealed that both clades have experienced positive diversifying selection. The CHIKV *envelope* gene encodes three viral glycoproteins associated with membrane fusion and receptor binding during infection, and these proteins are a target of neutralizing antibodies[42,43]. This gene constitutes a variable region in the virus genome where several adaptive mutations have been identified[44], for instance, mutations in the envelope proteins such as E1-226V and E2-L210Q have been implicated in the increased adaptation and transmission of the CHIKV IOL lineage in *Aedes albopictus*[22,45], leading to the epidemics reported in the Indian Ocean region between 2004 and 2007[46,47].

A higher ratio of non-synonymous substitutions is observed under a positive selection regime promoted by virus-host interactions[48]. Since SNVs continue to arise in RNA virus populations mainly driven by errors made by the virus replication complex that lead to genetic diversity, we compared the mutational profiles of clades I and II[49]. We identified in both clades isolates harboring mutations (E1-T98A and E2-V264A) previously associated with increased fitness in mosquitoes *Aedes ssp*. when these mutations are present in the background with either E1-226A[26] or E1-226V[27] mutations, although the latter was not observed in the genomes from this study. We identified several SNVs across non-structural and structural protein genes that were exclusive to each clade. Clade II presented more SNVs ($n = 27$) than clade I ($n = 13$), of which three are non-synonymous substitutions (E1-T288I; nsP2-P352A, and nsP4-A43V). Literature research revealed that E1-T288I change was previously identified in a 2017 sequence from Iran and also in CHIKV sequences collected in 2016 from infected cancer patients in Rio de Janeiro, Brazil[50,51]. The nsP2-P352A substitution was also present in sequences collected between 2016 and 2017 in Rio de Janeiro[17,52]. In turn, clade I sequences contain six non-synonymous mutations across nonstructural (nsP2 and nsP4) and structural protein sequences (capsid, 6k, and E2). The nsP4-V555I change was previously detected in sequences from Thailand in 2008–2009[53]. The E2-L248F from clade I has been reported in Asian lineage sequences from Colombia (2014–2015) and Philippines (2012)[54,55]. Isolates from Thailand, Indonesia, Lao PDR, Cameroon and India also presented the 6K-I54V mutation observed in clade I[53,56–59]. Despite the detection of these mutations in different countries (indicating homoplasy) by other previous studies, there is no information about the functional impact of such substitutions on CHIKV fitness, thus warranting further experimental studies to elucidate the potential effects of SNVs on lineage-specific evolutionary adaptation. It has been argued that not only non-synonymous mutations have the potential to promote adaptive changes but also synonymous mutations, including deletions and insertions in the 3′untranslated region[60,61], can lead to changes in the viral RNA that can drive differential viral gene expression and host adaptation[60–62].

Mutational analysis of the 422 sequences generated in this study revealed a higher amount of C > T and T > C transitions followed by A > G and G > A substitutions in the CHIKV genome and several genome positions presenting these transitions with higher frequency across all new sequences. Moreover, Clade II has a significantly higher median frequency of C > T, A > G, and T > C transitions compared to clade I. Although this study cannot experimentally establish the significance of these transitions for CHIKV evolutionary adaptation, other studies have associated this mutational pattern with the action of host antiviral immune response mediated by AID/APOBEC and ADAR families of deaminases[63,64]. These enzymes are part of the interferon-stimulated innate immune response and promote viral genome transitions mutations by catalyzing the deamination of adenosine to inosine to cause A > G/T > C (by ADAR) substitutions or deamination of cytosine to uracil that leads to C > T/G > A (by AID/APOBEC) mutations[65]. This RNA editing process has been experimentally observed targeting specific viral sequences of SARS-CoV-2 to produce C > T transitions and increasing viral replication in Caco-2 cells, thus promoting viral increased fitness and adaptative evolution[66]. However, specific information about the effect of these RNA editing processes on the CHIKV genome remains elusive, although the *APOBEC3A* gene

has been observed up-regulated in the expression profile of CHIKV-infected patients[67]. Despite arguments that these transitions happen mainly in phylogenetically uninformative sites, the available evidence indicates that RNA editing processes might act as a significant driver of viral sequence diversity and evolutionary adaptation through the introduction of nucleotide changes[64,68].

This study recognizes that some limitations should be noted. Although this study presented the results of a Bayesian phylogeographic and mutational profile analysis performed on 422 new CHIKV sequences collected from 12 Brazilian states, not all states were evenly represented in the dataset used, which might limit the estimates relative to divergence time and ancestral location reconstructions, prompting careful interpretation of the results presented here. Ongoing sequencing efforts across the country could reduce this disparity in the future. Moreover, although single nucleotide substitutions identified among the new sequences offer insights into the evolutionary dynamics of CHIKV in Brazil, further functional studies need to be undertaken to elucidate the actual adaptive effect of these mutations.

Recurring CHIKF epidemics, as indicated by the seasonal peak patterns displayed in the incidence time-series plot, are evidence that the virus is endemic in Brazil. Human mobility, population immunity, vector suitability, vegetation coverage, site socioeconomic status, and viral sequence variation are factors considered to mediate the dispersal of CHIKV in Brazil[29,69–71]. Although we did not find the E1:226V *Aedes albopictus*-adaptive mutation in the Brazilian sequences, the high abundance in the region of widely spread competent vectors, such as *Ae. aegypti* and *Ae. albopictus*, together with favorable climatic and social conditions in large urban centers create conditions that modify the adaptive landscape of CHIKV, which in turn can allow the continued expansion of the ECSA lineage in the country with a resultant increased impact on public health[23,72,73]. Therefore, public health measures should be undertaken to ensure continuous genomic surveillance of circulating CHIKV variants which can help to identify viral transmission routes where focused vector control strategies could be employed to reduce the risk of recurring CHIKF epidemics.

Methods

Ethical statement

The project was approved by the Pan American World Health Organization (PAHO) and the Ministry of Health of Brazil (MoH) as part of the arboviral genomic surveillance efforts within the terms of Resolution 510/2016 of CONEP (Comissão Nacional de Ética em Pesquisa, Ministério da Saúde; National Ethical Committee for Research, Ministry of Health). This authorizes the use of clinical samples collected in the Brazilian Central Public Health Laboratories to accelerate knowledge building and contribute to surveillance and outbreak response. The study protocol, including collection and publication of individual-level data, was reviewed and approved by the Research Ethics Committee of the Universidade Federal de Minas Gerais with approval No. 32912820.6.1001.5149. Personally identifiable information was de-identified in the datasets and tables in a way that minimizes the risk of unintended disclosure of identity of individuals and information about them.

Sample collection

Residual samples (serum or plasma) were obtained from the epidemiological surveillance routine of the Brazilian Central Public Health Laboratories (LACEN) from different states (Alagoas, Bahia, Goiás, Paraíba, Paraná, Pernambuco, Piauí, Maranhão, Minas Gerais, Mato Grosso do Sul, Rio Grande do Norte, and Sergipe). These samples, which were taken from patients that spontaneously seek medical care during epidemic season, were submitted to nucleic acid purification using the MagMax Viral RNA Mini kit (Thermo Fisher Scientific), following the manufacturer's recommendations, and were previously screened by each LACEN. CHIKV RT-qPCR positive samples were selected for sequencing based on the cycle threshold (Ct) value ≤ 30 and the availability of demographic metadata such as sex, age, and municipality of residency. These demographic patient data were provided by LACENs and were collected through a questionnaire filled out by patients and/or health professionals at local health care services.

cDNA synthesis and whole-genome sequencing

Extracted RNA from positive CHIKV samples were provided by collaborating LACENs and submitted to cDNA synthesis and PCR, using a sequencing protocol (primers sequences in Supplementary Data 1) based on multiplex PCR-tiling amplicon approach design for MinION nanopore sequencing[74]. All reactions were performed at biosafety level 2 facilities and using no template controls. PCR products were purified using 1x AMpure beads Beckman Coulter, UK) and quantified using Qubit 3.0 instrument (Life Technologies) and the Qubit dsDNA High Sensitivity assay. DNA library preparation was performed on all amplified samples using the Ligation Sequencing Kit (Oxford Nanopore Technologies). Individual samples were barcoded using the Native Barcoding Kit (NBD104, Oxford Nanopore Technologies, Oxford, UK). Sequencing library was loaded onto a R9.4 flow cell and data were collected for up to 48 sequencing hours.

Generation of consensus sequences

Basecalling of raw FAST5 files and demultiplexing of barcodes were performed using the software Guppy v6.0.1 (https://github.com/nanoporetech). Consensus sequences were generated by a hybrid assembling approach implemented on Genome Detective (https://www.genomedetective.com/)[75].

Phylogenetic reconstruction

We used MAFFT v.7 to align 422 new sequences (with coverage over >60% according to Thézé et al.[76], samples 736.22_RED, FS0116, and FS0132 were discarded) in addition to 1565 CHIKV whole-genome sequences publicly available in NCBI up to August 2022, forming a global dataset (n = 1987) that includes CHIKV all lineages. We used NCBI Entrez Utilities to retrieve worldwide CHIKV genomes according to the following inclusion criteria: chikungunya virus[title] AND 8000[SLEN]: 13000[SLEN] for minimum (60% genome coverage) and maximum sequence length. Alignment of the global dataset can be found on the repository https://doi.org/10.6084/m9.figshare.22335331.v2.

This global dataset was used to infer a Maximum Likelihood (ML) phylogeny using the IQ-TREE 2.1.1 software[77,78]. Statistical support for tree nodes was estimated using the ultrafast bootstrap (UFBoot) feature implemented in IQ-TREE with 1000 replicates. We then used the ML tree from the global dataset to extract the Brazilian ECSA clade and use it to form a second dataset (total n = 713, 706 Brazilian sequences, 2 from Haiti, and 5 from Paraguay; sequences with genome coverage >60% according to Thézé et al.[76]) which was used to infer a time-scaled phylogeny using BEAST v1.10.4. First, we investigated the temporal signal regressing root-to-tip genetic distances from this ML tree against sample collection dates using TempEst v.1.5.1[79]. Secondly, we employed a stringent model selection analysis using both path-sampling (PS) and stepping-stone (SS) procedures to estimate the most appropriate molecular clock model for the Bayesian phylogenetic analysis[80]. For the Bayesian analysis, the uncorrelated relaxed molecular clock was chosen as indicated by estimating marginal likelihoods, also employing the HKY + G4 nucleotide substitution model, and the nonparametric Bayesian Skyline coalescent model. We combined two independent runs of 200 million states each[81]. The convergence of MCMC chains was checked using Tracer[82]. Maximum clade credibility (MCC) trees were summarized using TreeAnnotator v1.10.4 after discarding 10% as burn-in.

CHIKV ECSA lineage movements across Brazil were investigated using the Bayesian phylogeographic approach with a discrete trait

phylogenetic model. A trait file was used to discretize sequences sampling location by five Brazilian regions (North, Northeast, Southeast, South, and Midwest). For this analysis, we downsampled our Brazilian ECSA clade to a dataset containing 471 sequences to maximize the temporal signal in the dataset. The final downsampled dataset was assembled by removing closely related sequences from clades with repetitive collection dates and locations, thus avoiding bias in the analysis due to those oversampled clades.

MCMC analyses were performed in BEAST v1.10.4, running in duplicate for 200 million interactions and sampling every 20,000 steps in the chain. Convergence for each run was assessed in Tracer. MCC trees for each run were summarized using TreeAnnotator after discarding the initial 10% as burn-in. Finally, we used the SPREAD 4 tool to map spatiotemporal information embedded in the MCC trees[83].

Transmission network analysis

A transmission network was reconstructed, using the StrainHub tool v1.1.2, from transition states summarized from the Bayesian phylogeography[84]. Centrality metrics on the tree nodes were also estimated for the network.

Comparative mutational analysis

For comparative mutational analysis, we assembled separate alignment datasets for each subclade (clade I = 327; clade II = 168) and for all new sequences ($n = 422$). The sequence datasets were compared against the NCBI reference strain NC_004162.2 using MALVIRUS v1.0.2[85]. We filtered and selected substitutions with an occurrence frequency of 100% across the whole dataset and substitutions with a frequency above 60% across the envelope genes E1 and E2. Results were summarized in boxplots using Rstudio 2022.12.0.

Selection pressure analysis

Since the ratio of non-synonymous (*dN*) to synonymous (*dS*) nucleotide substitutions can be used to study selection pressure on genomic sequences, we used the HYPHY software package that employs statistical methods that estimate the *dN/dS* to detect diversifying selection[86]. For that we performed an alignment of the envelope gene sequences of 720 Brazilian isolates and used BUSTED (restricting the analysis to each subclade I or II), an alignment-wide method implemented in HYPHY 2.5.42, that aims to detect evidence of episodic diversifying selection[87]. For comparison, we also used a different method, called MEME, a site-level approach, also implemented in HYPHY 2.5.42, that aims to detect evidence of both pervasive and episodic diversifying selection at individual sites (also restricting the analysis to each subclade I or II)[88].

Epidemic curves from chikungunya fever cases reported in Brazil

Data of weekly notified suspected cases of chikungunya fever in Brazil from 2014 to 2022 (up to epidemiological week 28 of the year 2022) were supplied by the Ministry of Health of Brazil (MoH) (Ministério da Saúde, 2022). These data were used to calculate incidence and to plot time series charts using Rstudio 2022.12.0. The MoH defines a suspected case as all human cases presenting symptoms compatible with CHIKV infection (sudden onset of fever or intense arthritis not explained by other conditions), a case that resides or has traveled to endemic or epidemic areas up to 14 days prior to symptom onset, or that has an epidemiological link to an imported confirmed case. The MoH also informs that during epidemic season most cases are confirmed by clinical criteria only.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The new sequences generated in this study have been deposited in NCBI GenBank under accession numbers OQ759652- OQ760076 listed in Supplementary Data 1. Input data used for the phylogenetic and mutational profile analyses are provided on the repository https://doi.org/10.6084/m9.figshare.22335331.v2.

Code availability

R codes used to create the charts are available on https://doi.org/10.6084/m9.figshare.22335331.v2.

References

1. Mourad, O., Makhani, L. & Chen, L. H. Chikungunya: an emerging public health concern. *Curr. Infect. Dis. Rep.* **24**, 217–228 (2022).
2. Chala, B. & Hamde, F. Emerging and re-emerging vector-borne infectious diseases and the challenges for control: a review. *Front. Public Health* **9**, 715759 (2021).
3. World Health Organization. *Global Vector Control Response 2017–2030* (World Health Organization, 2017).
4. World Health Organization. *10 Proposals to Build A Safer World Together – Strengthening the Global Architecture for Health Emergency Preparedness, Response and Resilience* (WHO Health Emergencies Programme, 2022).
5. Adelino, T. É. R. et al. Field and classroom initiatives for portable sequence-based monitoring of dengue virus in Brazil. *Nat. Commun.* **12**, 2296 (2021).
6. Faria, N. R. et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* **546**, 406–410 (2017).
7. Costa, É. A. et al. West Nile virus in Brazil. *Pathogens* **10**, 896 (2021).
8. Faria, N. R. et al. Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science* **361**, 894–899 (2018).
9. Giovanetti, M. et al. A retrospective overview of Zika virus evolution in the midwest of Brazil. *Microbiol. Spectr.* **10**, e0015522 (2022).
10. Naveca, F. G. et al. Genomic, epidemiological and digital surveillance of Chikungunya virus in the Brazilian Amazon. *PLoS Negl. Trop. Dis.* **13**, e0007065 (2019).
11. Xavier, J. et al. Circulation of chikungunya virus East/Central/South African lineage in Rio de Janeiro, Brazil. *PLoS ONE* **14**, e0217871 (2019).
12. Nunes, M. R. T. et al. Emergence and potential for spread of Chikungunya virus in Brazil. *BMC Med.* **13**, 102 (2015).
13. Powers, A. M. Vaccine and therapeutic options to control chikungunya virus. *Clin. Microbiol. Rev.* **31** e00104-16 (2018).
14. ClinicalTrials.gov. Pivotal study to evaluate safety and immunogenicity of a live-attenuated chikungunya virus vaccine candidate in Adults-NCT04546724. https://clinicaltrials.gov/study/NCT04546724 (2022).
15. Bennett, S. R. et al. Safety and immunogenicity of PXVX0317 an aluminium hydroxide-adjuvanted chikungunya virus-like particle vaccine: a randomised double-blind parallel-group phase 2 trial. *Lancet Infect. Dis.* **22**, 1343–1355 (2022).
16. Pan American Health Organization. Cases of chikungunya virus disease. *PAHO/WHO Data—Weekly Report*. https://www3.paho.org/data/index.php/en/mnu-topics/chikv-en/550-chikv-weekly-en.html (2023).
17. Souza, T. M. L. et al. Emergence of the East-Central-South-African genotype of Chikungunya virus in Brazil and the city of Rio de Janeiro may have occurred years before surveillance detection. *Sci. Rep.* **9**, 2760 (2019).
18. de Lima, S. T. S. et al. Fatal outcome of chikungunya virus infection in Brazil. *Clin. Infect. Dis.* **73**, e2436–e2443 (2021).
19. Robishaw, J. D. et al. Genomic surveillance to combat COVID-19: challenges and opportunities. *Lancet Microbe* **2**, e481–e484 (2021).

20. Xavier, J. et al. Chikungunya virus ECSA lineage reintroduction in the northeasternmost region of Brazil. *Int. J. Infect. Dis.* **105**, 120–123 (2021).
21. Garcia-Beltran, W. F. et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* **184**, 2372–2383.e9 (2021).
22. Tsetsarkin, K. A. et al. Multi-peaked adaptive landscape for chikungunya virus evolution predicts continued fitness optimization in Aedes albopictus mosquitoes. *Nat. Commun.* **5**, 4084 (2014).
23. Pancetti, F. G. M., Honório, N. A., Urbinatti, P. R. & Lima-Camara, T. N. Twenty-eight years of Aedes albopictus in Brazil: a rationale to maintain active entomological and epidemiological surveillance. *Rev. Soc. Bras. Med. Trop.* **48**, 87–89 (2015).
24. Carvalho, R. G., Lourenço-de-Oliveira, R. & Braga, I. A. Updating the geographical distribution and frequency of Aedes albopictus in Brazil with remarks regarding its range in the Americas. *Mem. Inst. Oswaldo Cruz.* **109**, 787–796 (2014).
25. de Souza, W. M. et al. Spatiotemporal dynamics and recurrence of chikungunya virus in Brazil: an epidemiological study. *Lancet Microbe* **4**, e319–e329 (2023).
26. Agarwal, A., Sharma, A. K., Sukumaran, D., Parida, M. & Dash, P. K. Two novel epistatic mutations (E1:K211E and E2:V264A) in structural proteins of chikungunya virus enhance fitness in *Aedes aegypti*. *Virology* **497**, 59–68 (2016).
27. Tsetsarkin, K. A. et al. Chikungunya virus emergence is constrained in Asia by lineage-specific adaptive landscapes. *Proc. Natl Acad. Sci. USA* **108**, 7872–7877 (2011).
28. Grubaugh, N. D. et al. Tracking virus outbreaks in the twenty-first century. *Nat. Microbiol.* **4**, 10–19 (2019).
29. Charlys da Costa, A. et al. Spread of chikungunya virus east/central/south African genotype in northeast Brazil. *Emerg. Infect. Dis.* **23**, 1742–1744 (2017).
30. Maeda, A. Y. et al. Circulation of Chikungunya virus East-Central-South African genotype during the 2020–21 outbreak in São Paulo State, Brazil. *J. Clin. Virol.* **2**, 100070 (2022).
31. de Castro Moreira, D. et al. Genetic characterization of Chikungunya virus circulating in individuals from Paraná, Brazil. *Braz. J. Microbiol.* **53**, 641–645 (2022).
32. Teixeira, M. G. et al. East/Central/South African genotype Chikungunya virus, Brazil, 2014. *Emerg. Infect. Dis.* **21**, 906–907 (2015).
33. Souza, T. M. A. et al. First report of the East-Central South African genotype of chikungunya virus in Rio de Janeiro, Brazil. *PLoS Curr.* **9**, 01–01 (2017).
34. Pereira Gusmão Maia, Z. et al. Return of the founder Chikungunya virus to its place of introduction into Brazil is revealed by genomic characterization of exanthematic disease cases. *Emerg. Microbes Infect.* **9**, 53–57 (2020).
35. Tanabe, E. L. et al. Report of East-Central South African Chikungunya virus genotype during the 2016 outbreak in the Alagoas State, Brazil. *Rev. Inst. Med. Trop. Sao Paulo* **60**, e19 (2018).
36. White, S. K. et al. A new "American" subgroup of African-lineage Chikungunya virus detected in and isolated from mosquitoes collected in Haiti, 2016. *PLoS ONE* **13**, e0196857 (2018).
37. Gräf, T. et al. Epidemiologic history and genetic diversity origins of chikungunya and dengue viruses, Paraguay. *Emerg. Infect. Dis.* **27**, 1393–1404 (2021).
38. de Oliveira, E. C. et al. Short report: Introduction of chikungunya virus ECSA genotype into the Brazilian Midwest and its dispersion through the Americas. *PLoS Negl. Trop. Dis.* **15**, e0009290 (2021).
39. Spielman, S. J. et al. Evolution of viral genomes: interplay between selection, recombination, and other forces. *Methods Mol. Biol.* **1910**, 427–468 (2019).
40. Yu, X. & Cheng, G. Adaptive evolution as a driving force of the emergence and re-emergence of mosquito-borne viral diseases. *Viruses* **14**, 435 (2022).
41. Chen, R. et al. Lineage divergence and vector-specific adaptation have driven Chikungunya virus onto multiple adaptive landscapes. *MBio* **12**, e0273821 (2021).
42. Quiroz, J. A. et al. Human monoclonal antibodies against Chikungunya virus target multiple distinct epitopes in the E1 and E2 glycoproteins. *PLoS Pathog.* **15**, e1008061 (2019).
43. Jin, J. et al. Neutralizing monoclonal antibodies block chikungunya virus entry and release by targeting an epitope critical to viral pathogenesis. *Cell Rep.* **13**, 2553–2564 (2015).
44. Bartholomeeusen, K. et al. Chikungunya fever. *Nat. Rev. Dis. Prim.* **9**, 17 (2023).
45. Tsetsarkin, K. A., Vanlandingham, D. L., McGee, C. E. & Higgs, S. A single mutation in chikungunya virus affects vector specificity and epidemic potential. *PLoS Pathog.* **3**, e201 (2007).
46. Tsetsarkin, K. A., Chen, R., Sherman, M. B. & Weaver, S. C. Chikungunya virus: evolution and genetic determinants of emergence. *Curr. Opin. Virol.* **1**, 310–317 (2011).
47. Mavalankar, D., Shastri, P., Bandyopadhyay, T., Parmar, J. & Ramani, K. V. Increased mortality rate associated with chikungunya epidemic, Ahmedabad, India. *Emerg. Infect. Dis.* **14**, 412–415 (2008).
48. Duggal, N. K. & Emerman, M. Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat. Rev. Immunol.* **12**, 687–695 (2012).
49. Collins, N. D. et al. Structural and nonstructural genes contribute to the genetic diversity of RNA viruses. *MBio* **9**, e01871–18 (2018).
50. Pouriayevali, M. H. et al. Imported cases of Chikungunya virus in Iran. *BMC Infect. Dis.* **19**, 1004 (2019).
51. Familiar-Macedo, D. et al. Molecular aspects of Chikungunya virus infections in cancer patients. *Mem. Inst. Oswaldo Cruz.* **117**, e210383 (2022).
52. Cunha, M. S. et al. Autochthonous transmission of east/central/south African genotype chikungunya virus, brazil. *Emerg. Infect. Dis.* **23**, 1737–1739 (2017).
53. Pongsiri, P. et al. Entire genome characterization of Chikungunya virus from the 2008-2009 outbreaks in Thailand. *Trop. Biomed.* **27**, 167–176 (2010).
54. Rodas, J. D. et al. Genetic characterization of Northwestern Colombian Chikungunya virus strains from the 2014-2015 epidemic. *Am. J. Trop. Med. Hyg.* **95**, 639–646 (2016).
55. Tan, K.-K. et al. Independent emergence of the cosmopolitan Asian chikungunya virus, Philippines 2012. *Sci. Rep.* **5**, 12279 (2015).
56. Hakim, M. S., Annisa, L. & Aman, A. T. The evolution of chikungunya virus circulating in Indonesia: sequence analysis of the orf2 gene encoding the viral structural proteins. *Int. Microbiol.* https://doi.org/10.1007/s10123-023-00337-1 (2023).
57. Khan, N. et al. Epidemiology and molecular characterization of chikungunya virus from human cases in North India, 2016. *Microbiol. Immunol.* **65**, 290–301 (2021).
58. Calvez, E. et al. Multiple chikungunya virus introductions in Lao PDR from 2014 to 2020. *PLoS ONE* **17**, e0271439 (2022).
59. Agbodzi, B. et al. Chikungunya viruses containing the A226V mutation detected retrospectively in Cameroon form a new geographical subclade. *Int. J. Infect. Dis.* **113**, 65–73 (2021).
60. Chen, R., Wang, E., Tsetsarkin, K. A. & Weaver, S. C. Chikungunya virus 3′ untranslated region: adaptation to mosquitoes and a population bottleneck as major evolutionary forces. *PLoS Pathog.* **9**, e1003591 (2013).
61. Morley, V. J. et al. Chikungunya virus evolution following a large 3′UTR deletion results in host-specific molecular changes in protein-coding regions. *Virus Evol.* **4**, vey012 (2018).
62. Faure, G., Ogurtsov, A. Y., Shabalina, S. A. & Koonin, E. V. Adaptation of mRNA structure to control protein folding. *RNA Biol.* **14**, 1649–1654 (2017).
63. Chaw, S.-M. et al. The origin and underlying driving forces of the SARS-CoV-2 outbreak. *J. Biomed. Sci.* **27**, 73 (2020).

64. Simmonds, P. & Ansari, M. A. Extensive C->U transition biases in the genomes of a wide range of mammalian RNA viruses; potential associations with transcriptional mutations, damage- or host-mediated editing of viral RNA. *PLoS Pathog.* **17**, e1009596 (2021).
65. Pecori, R., Di Giorgio, S., Paulo Lorenzo, J. & Nina Papavasiliou, F. Functions and consequences of AID/APOBEC-mediated DNA and RNA deamination. *Nat. Rev. Genet.* **23**, 505–518 (2022).
66. Kim, K. et al. The roles of APOBEC-mediated RNA editing in SARS-CoV-2 mutations, replication and fitness. *Sci. Rep.* **12**, 14972 (2022).
67. Soares-Schanoski, A. et al. Systems analysis of subjects acutely infected with the Chikungunya virus. *PLoS Pathog.* **15**, e1007880 (2019).
68. Piontkivska, H., Wales-McGrath, B., Miyamoto, M. & Wayne, M. L. ADAR editing in viruses: an evolutionary force to reckon with. *Genome Biol. Evol.* **13**, evab240 (2021).
69. Churakov, M., Villabona-Arenas, C. J., Kraemer, M. U. G., Salje, H. & Cauchemez, S. Spatio-temporal dynamics of dengue in Brazil: Seasonal travelling waves and determinants of regional synchrony. *PLoS Negl. Trop. Dis.* **13**, e0007012 (2019).
70. Freitas, L. P., Schmidt, A. M., Cossich, W., Cruz, O. G. & Carvalho, M. S. Spatio-temporal modelling of the first Chikungunya epidemic in an intra-urban setting: the role of socioeconomic status, environment and temperature. *PLoS Negl. Trop. Dis.* **15**, e0009537 (2021).
71. Power, G. M. et al. Socioeconomic risk markers of arthropod-borne virus (arbovirus) infections: a systematic literature review and meta-analysis. *BMJ Glob. Health* **7**, e007735 (2022).
72. de Melo Ximenes, M. et al. Arbovirus expansion: new species of culicids infected by the Chikungunya virus in an urban park of Brazil. *Acta Trop.* **209**, 105538 (2020).
73. Costa-da-Silva, A. L. et al. First report of naturally infected Aedes aegypti with chikungunya virus genotype ECSA in the Americas. *PLoS Negl. Trop. Dis.* **11**, e0005630 (2017).
74. Quick, J. et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **12**, 1261–1276 (2017).
75. Vilsker, M. et al. Genome Detective: an automated system for virus identification from high-throughput sequencing data. *Bioinformatics* **35**, 871–873 (2019).
76. Thézé, J. et al. Genomic epidemiology reconstructs the introduction and spread of Zika virus in Central America and Mexico. *Cell Host Microbe* **23**, 855–864.e7 (2018).
77. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
78. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005).
79. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).
80. Baele, G. et al. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol. Biol. Evol.* **29**, 2157–2167 (2012).
81. Suchard, M. A. et al. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, vey016 (2018).
82. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
83. Nahata, K. D. et al. SPREAD 4: online visualisation of pathogen phylogeographic reconstructions. *Virus Evol.* **8**, veac088 (2022).
84. de Bernardi Schneider, A. et al. StrainHub: a phylogenetic tool to construct pathogen transmission networks. *Bioinformatics* **36**, 945–947 (2020).
85. Ciccolella, S. et al. MALVIRUS: an integrated application for viral variant analysis. *BMC Bioinforma.* **22**, 625 (2022).
86. Kosakovsky Pond, S. L. et al. HyPhy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. *Mol. Biol. Evol.* **37**, 295–299 (2020).
87. Murrell, B. et al. Gene-wide identification of episodic selection. *Mol. Biol. Evol.* **32**, 1365–1371 (2015).
88. Murrell, B. et al. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8**, e1002764 (2012).

Acknowledgements

This work was supported by The Pan American Health Organization (PAHO/WHO), the Brazilian Ministry of Health grant SCON2021-00180 (Coordenação Geral de Laboratório de Saúde Pública-CGLAB and Coordenação Geral de Vigilância de Arboviroses-CGARB), the National Institutes of Health USA grant U01 AI151698 for the United World Arbovirus Research Network (UWARN). We thank the State Health Secretariats and the epidemiological surveillance services of Brazilian states and municipalities for all the support provided. J.X. was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES)-Finance Code 001. M.G. is funded by PON "Ricerca e Innovazione" 2014–2020. M.G. is supported in part by the CRP- ICGEB RESEARCH GRANT 2020 Project CRP/BRA20-03, Contract CRP/20/03. The authors would like to acknowledge the Global Consortium to Identify and Control Epidemics – CLIMADE (T.O., L.C.J.A., E.C.H., J.L., and M.G.) (https://climade.health/).

Author contributions

Conception and design: L.C.J.A., M.G., V.F., and J.X.; investigations: J.X., M.G., L.C.J.A., V.F., A.M.B.D.F., Mau.L., E.C., Heg.F., C.O., N.G., Tal.A., M.E., E.S.R, E.V.S, D.D.L.R., L.D.M., S.T., A.N., A.R., A.F.M., A.L., A.V., A.L.S.D.M., B.V., C.A.M., C.Z., C.F., C.F.C.D.A, C.N.D.D.S, C.S.S., C.A.D.S., C.C.M.G., D.T., D.F.L.N., D.C., E.C.D.O., E.L.N.M., F.M.P., F.I., F.P.D.C., G.A., G.B., G.G.D.C.L., G.C.P., H.B., H.C.F.F., Hivy.F., I.G., I.N.R., I.R., I.C.D.S., J.S., J.M.R., J.L., J.A., J.P.M.D.N., J.N.W., J.P., J.J.F.D.M., K.G.L., L.G.L.N., L.C.V.F., L.B.D.S., L.S., L.A.F.D.S., L.A.P., L.D., M.C.B.C., M.G.A., M.A., Mar.L., M.C.S.U.Z., M.M.P., M.B.F., M.G.J., N.F., N.M., N.F.O.D.M., P.E.A.D.S., P.R., R.V.D.C., K.S.T., R.F.D.C.S., R.K., R.S., R.D.J., R.H.S., S.K., S.N.S., Tam.A., T.R., T.C., V.N., V.D.S., W.G.C., W.C.V.V., and W.N.A. Data curation: J.X., M.G., V.F., L.C.J.A; formal analysis: J.X., M.G., V.F., and L.C.J.A.; writing—original draft preparation: J.X., M.G., and L.C.J.A.; revision: J.X., M.G., V.F., and L.C.J.A.; resources: J.M.R., M.A., W.N.A., A.R., R.F.D.C.S., C.F.C.D.A, W.C.V.V., P.R., M.G.J., J.N.W., R.S., E.L.N.M., P.E.A.D.S., H.C.F.F., C.F., N.F.O.D.M., L.C.V.F., A.M.B.D.F., and L.C.J.A.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-023-40099-y.

Correspondence and requests for materials should be addressed to Luiz Carlos Junior Alcantara, Ana M. B. de Filippis or Marta Giovanetti.

Peer review information *Nature Communications* thanks Rhys Inward and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

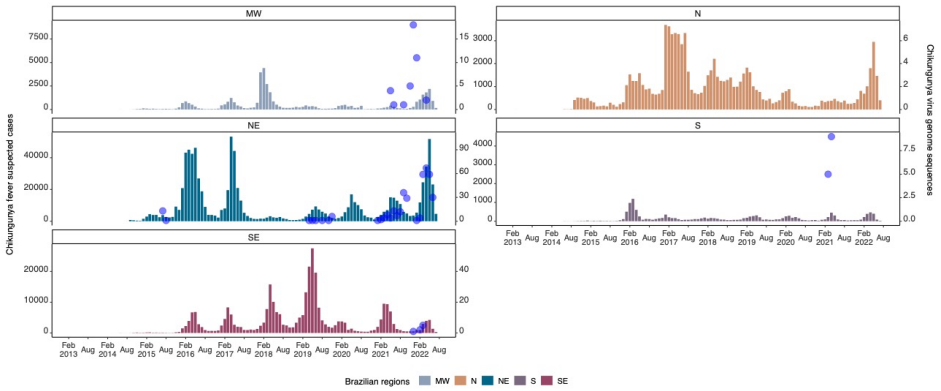
Article https://doi.org/10.1038/s41467-023-40099-y

Joilson Xavier[1,2,38], Luiz Carlos Junior Alcantara[1,2,38] ✉, Vagner Fonseca[3,38], Mauricio Lima[1,4], Emerson Castro[1,4], Hegger Fritsch[1,2], Carla Oliveira[5], Natalia Guimarães[4], Talita Adelino[4], Mariane Evaristo[6], Evandra S. Rodrigues[6], Elaine Vieira Santos[6], Debora de La-Roque[6], Laise de Moraes[7], Stephane Tosta[1,2], Adelino Neto[8], Alexander Rosewell[3], Ana Flavia Mendonça[9], Anderson Leite[10], Andreza Vasconcelos[11], Arabela L. Silva de Mello[12], Bergson Vasconcelos[13], Camila A. Montalbano[14], Camila Zanluca[15], Carla Freitas[16], Carlos F. C. de Albuquerque[3], Claudia Nunes Duarte dos Santos[15], Cleiton S. Santos[7], Cliomar Alves dos Santos[17], Crhistinne C. Maymone Gonçalves[18], Dalane Teixeira[13], Daniel F. L. Neto[16], Diego Cabral[11], Elaine C. de Oliveira[19], Ethel L. Noia Maciel[20], Felicidade Mota Pereira[12], Felipe Iani[4], Fernanda P. de Carvalho[11], Gabriela Andrade[12], Gabriela Bezerra[17], Gislene G. de Castro Lichs[21], Glauco Carvalho Pereira[4], Haline Barroso[13], Helena Cristina Ferreira Franz[16], Hivylla Ferreira[22], Iago Gomes[23], Irina N. Riediger[24], Isabela Rodrigues[17], Isadora C. de Siqueira[7], Jacilane Silva[11], Jairo Mendez Rico[25], Jaqueline Lima[12], Jayra Abrantes[23], Jean Phellipe M. do Nascimento[10], Judith N. Wasserheit[26], Julia Pastor[11], Jurandy J. F. de Magalhães[11,27], Kleber Giovanni Luz[28], Lidio G. Lima Neto[22], Livia C. V. Frutuoso[29], Luana Barbosa da Silva[19], Ludmila Sena[17], Luis Arthur F. de Sousa[22], Luiz Augusto Pereira[9], Luiz Demarchi[21], Magaly C. B. Câmara[23], Marcela G. Astete[12], Maria Almiron[25], Maricelia Lima[30], Marina C. S. Umaki Zardin[21], Mayra M. Presibella[24], Melissa B. Falcão[31], Michael Gale Jr.[32], Naishe Freire[11], Nelson Marques[24], Noely F. O. de Moura[29], Pedro E. Almeida Da Silva[20], Peter Rabinowitz[33], Rivaldo V. da Cunha[34], Karen S. Trinta[34], Rodrigo F. do Carmo Said[3], Rodrigo Kato[16], Rodrigo Stabeli[3], Ronaldo de Jesus[16], Roselene Hans Santos[11], Simone Kashima[6], Svetoslav N. Slavov[6,35], Tamires Andrade[13], Themis Rocha[23], Thiago Carneiro[13], Vanessa Nardy[12], Vinicius da Silva[9], Walterlene G. Carvalho[8], Wesley C. Van Voorhis[36], Wildo N. Araujo[25], Ana M. B. de Filippis[5] ✉ & Marta Giovanetti[1,2,37] ✉

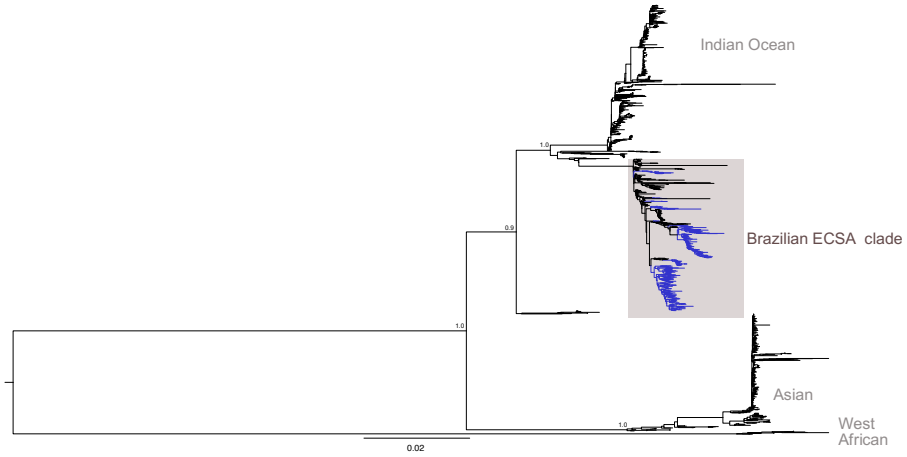
[1]Instituto René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Brazil. [2]Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil. [3]Organização Pan-Americana da Saúde, Organização Mundial da Saúde, Brasília, Brazil. [4]Laboratório Central de Saúde Pública de Minas Gerais, Fundação Ezequiel Dias, Belo Horizonte, Brazil. [5]Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil. [6]Fundação Hemocentro de Ribeirão Preto, Ribeirão Preto, Brazil. [7]Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Brazil. [8]Laboratório Central de Saúde Pública do Piaui, Piauí, Brazil. [9]Laboratório Central de Saúde Pública de Goias, Goiânia, Brazil. [10]Laboratório Central de Saúde Pública de Alagoas, Maceió, Brazil. [11]Laboratório Central de Saúde Pública de Pernambuco, Natal, Brazil. [12]Laboratório Central de Saúde Pública da Bahia, Salvador, Brazil. [13]Laboratório Central de Saúde Pública da Paraíba, João Pessoa, Brazil. [14]Universidade Federal de Mato Grosso do Sul, Campo Grande, Brazil. [15]Instituto Carlos Chagas, Fundação Oswaldo Cruz, Curitiba, Brazil. [16]Coordenação Geral dos Laboratórios de Saúde Pública, Ministério da Saúde, Brasília, Brazil. [17]Laboratório Central de Saúde Pública de Sergipe, Aracaju, Brazil. [18]Secretaria de Saúde do Estado do Mato Grosso do Sul, Campo Grande, Brazil. [19]Laboratório Central de Saúde Pública do Mato Grosso, Cuiabá, Brazil. [20]Secretaria de Vigilância em Saúde e Ambiente, Ministério da Saúde, Brasília, Brazil. [21]Laboratório Central de Saúde Pública do Mato Grosso do Sul, Campo Grande, Brazil. [22]Laboratório Central de Saúde Pública do Maranhão, São Luís, Brazil. [23]Laboratório Central de Saúde Pública do Rio Grande do Norte, Natal, Brazil. [24]Laboratório Central de Saúde Pública do Paraná, Paraná, Brazil. [25]Pan American Health Organization, Washington, USA. [26]Department of Global Health and Medicine, University of Washington, Washington, USA. [27]Universidade de Pernambuco, Serra Talhada, Brazil. [28]Universidade Federal do Rio Grande do Norte, Natal, Brazil. [29]Coordenação Geral das Arboviroses, Ministério da Saúde, Brasília, Brazil. [30]Universidade Estadual de Feira de Santana, Feira de Santana, Brazil. [31]Secretaria de Saúde de Feira de Santana, Feira de Santana, Brazil. [32]Department of Immunology, University of Washington, Washington, USA. [33]Department of Environmental and Occupational Health Sciences, University of Washington, Washington, USA. [34]Fundação Oswaldo Cruz, Instituto de Tecnologia em Imunobiológicos, Rio de Janeiro, Brazil. [35]Center for Research Development, CDC, Butantan Institute, São Paulo, Brazil. [36]Department of Medicine, University of Washington, Washington, USA. [37]Sciences and Technologies for Sustainable Development and One Health, University of Campus Bio-Medico, Rome, Italy. [38]These authors contributed equally: Joilson Xavier, Luiz Carlos Junior Alcantara, Vagner Fonseca. ✉e-mail: luiz.alcantara@ioc.fiocruz.br; ana.bispo@ioc.fiocruz.br; giovanetti.marta@gmail.com

Increased interregional virus exchange and nucleotide diversity outline the expansion of the chikungunya virus ECSA lineage in Brazil

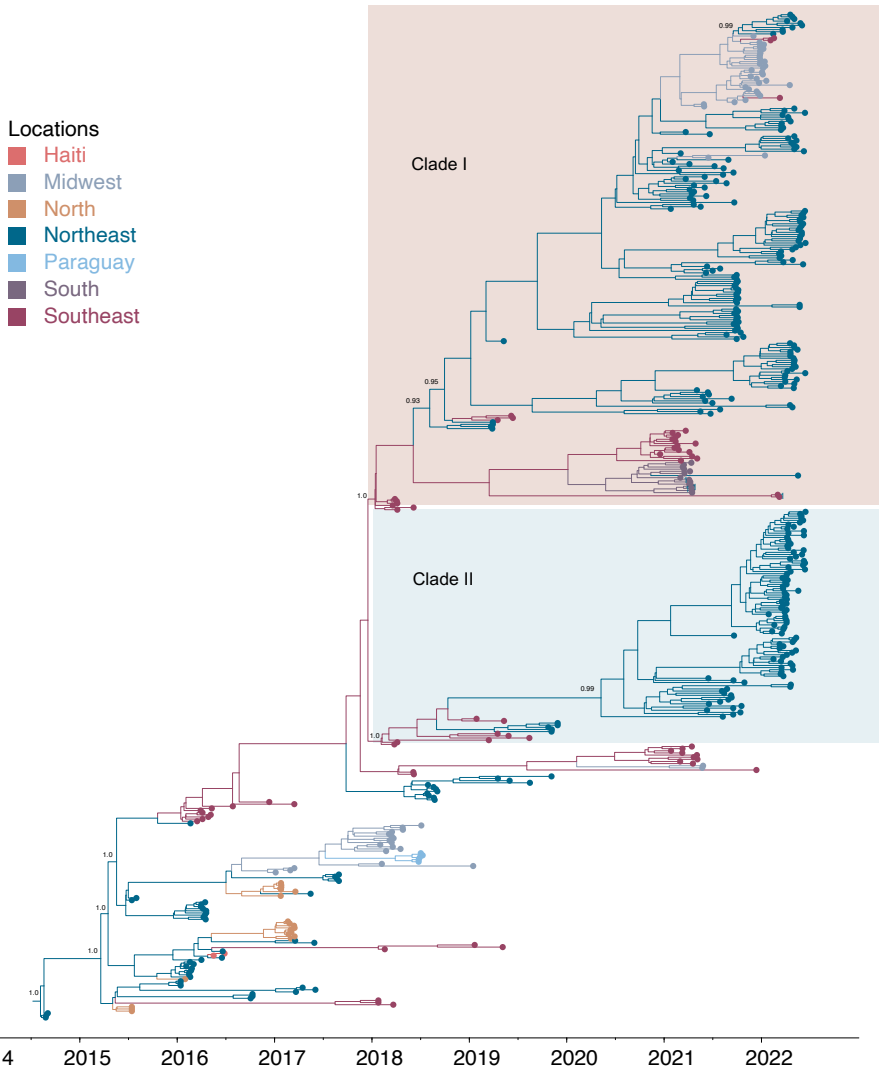
Supplementary Information



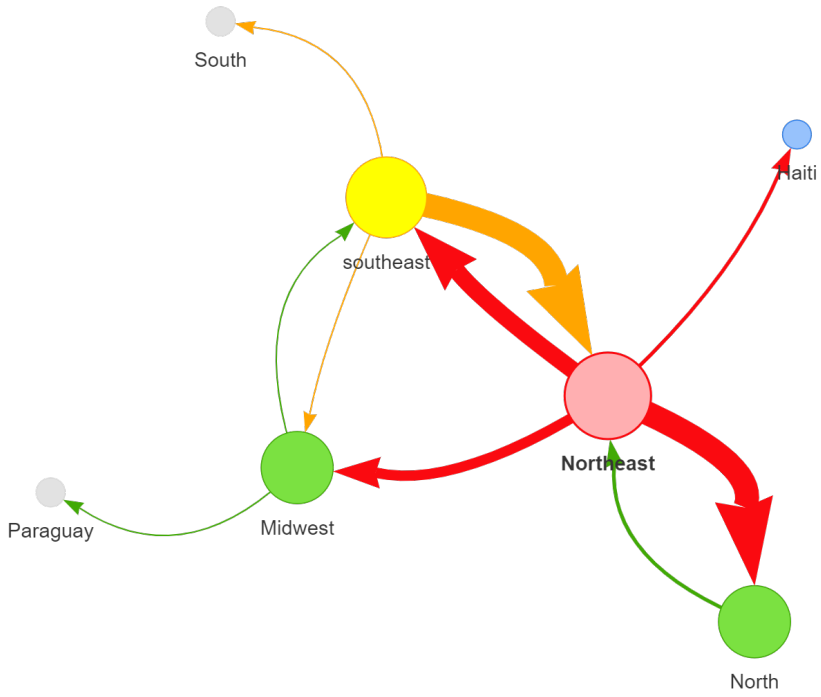
Supplementary Fig. 1. Number of chikungunya virus genomic sequences by number of cases. Bar plot of the number of notified suspected cases of chikungunya fever reported in different Brazilian geographic regions. The Brazilian chikungunya virus genomes generated in this study are indicated by light blue circles over the bars. Bar colors indicate different geographic regions in Brazil, indicated by the legend.



Supplementary Fig. 2. Global phylogeny of CHIKV lineages. Maximum Likelihood (ML) phylogeny was reconstructed using 1,987 CHIVK sequences, including the new genomes and others available on GenBank (see methods in the main text). New Brazilian genomes are indicated by blue branches in the Brazilian ECSA clade. Numbers are clade ultrafast bootstrap support values.



Supplementary Fig. 3. **Ancestral locations of viruses from the CHIKV ECSA lineage circulating in Brazil**. Maximum Clade Credibility tree reconstructed under a Bayesian phylogeography in discrete space using sequences (total n=471) from Brazil, Paraguay and Haiti (see methods in the main text)**.** Location probabilities of clades I and II as well as of main early branches estimated to be originated in the Northeast of Brazil are represented by numbers placed near nodes. Location probabilities of other clades were omitted for clarity. Branch colors represent the estimated ancestral location.



Supplementary Fig. 4. **Transmission network of CHIKV in Brazil**. The network was inferred from the transitions between geographic states from the Bayesian phylogeography under a discrete space model. Arrows indicate directionality and their thickness indicates transition frequencies. Node size was scaled by source hub ratio values estimated for each location using StrainHub tool v1.1.2. Colors are arbitrary.

Supplementary Table 1. Centrality Metrics employed to measure transmission networks generated by the StrainHub tool.

Metastates	Degree Centrality	Indegree Centrality	Outdegree Centrality	Betweenness Centrality	Closeness Centrality	Source Hub Ratio
Paraguay	1	1	0	0	0.07	0.00
Southeast	5	2	3	6	0.11	0.60
Northeast	6	2	4	9	0.13	0.67
Haiti	1	1	0	0	0.08	0.00
Midwest	4	2	2	3	0.11	0.50
North	2	1	1	0	0.08	0.50
South	1	1	0	0	0.07	0.00

Supplementary Table 2. Detailed site-by-site results from the BUSTED analysis for the envelope gene of clade I and II.

Codon	ER ($\omega>1$)[1]	LogL	Epost[α][2]	LR[3]
Clade I				
187	187,712.29	-78.412	0.294	12.721
372	742.301	-31.012	1.07	6.651
699	344.611	-92.661	0.294	3.369
60	211.964	-79.268	0.294	2.566
62	57.168	-17.817	3.408	4.991
Clade II				
187	213,975.17	-78.617	36.082	10.86
186	92,991.85	-62.607	36.082	7.669
371	726.104	-17.444	2.287	6.078
121	499.333	-15.376	0.6	7.502

184	105.634	-59.121	35.724	0.507
29	96.931	-36.048	33.526	4.941
60	19.358	-78.757	36.082	1.279
370	12.692	-42.632	36.074	1.966

[1]Evidence ratio for positive selection.
[2]Posterior mean of the synonymous rate, α.
[3]Site log-likelihood ratio contribution.

Supplementary Table 3. Sites under positive diversifying selection identified by HYPHY-MEME in the envelope gene of the clades I and II.

Site	alpha	beta+	LRT	p-value
Clade I				
33	0	5.58	4.31	0.05
60	3.62	364.35	7.88	0.01
62	0	284.76	8.45	0.01
138	0	5.28	4.11	0.06
184	0	58.06	7.23	0.01
187	1.09	1521.76	21.03	0
282	0	2.31	3.43	0.09
372	0	6084.51	18.38	0
441	0	5.35	6.41	0.02
539	0	9.23	4.23	0.06
695	0	556.34	9.8	0
699	0	131.88	11.93	0
Clade II				
29	0	315.97	9.55	0
53	0	3.7	3.71	0.07
60	4.18	469.33	4.83	0.04
121	0	181.66	8.77	0.01

138	0	4.53	3.31	0.09
184	0	133.16	8.43	0.01
186	4.7	988.28	12.58	0
187	2.1	1124.28	20.44	0
371	0	256.38	9.31	0
432	0	4.87	3.88	0.07
441	0	3.16	3.38	0.09
539	0	70.36	4.82	0.04
699	0	9.05	3.67	0.08

alpha: synonymous substitution rate at a site.

beta+: non-synonymous substitution rate at a site for the positive selection category.

LRT: likelihood ratio test statistic for episodic diversification.

p-value: asymptotic p-values with a significance level alpha = 0.01.

4.2 Geração de sequências genômicas para identificação e classificação viral

Artigo 1: A multiplex nanopore sequencing approach for the detection of multiple arboviral species. Submetido para publicação: Xavier *et al*., 2023.

A multiplex nanopore sequencing approach for the detection of multiple arboviral species

Joilson Xavier[1], Vagner Fonseca[2], Talita Adelino[3], Felipe Iani[3], Glauco C. Pereira[3], Myrian M. Duarte[3], Mauricio Lima[4], Emerson Castro[4], Carla Oliveira[5], Hegger Fritsch[1], Natalia Guimaraes[3], Ludmila O. Lamounier[3], Fernanda Khouri Barreto[6], Camilo M. M. Braga de Oliveira[7], Crhistinne C. Maymone Gonçalves[8], Danielle Malta Lima[9], Elaine C. de Oliveira[10], Gislene G. de Castro Lichs[11], Iago Gomes[12], Janaina Mazaro[13], Janete T. N. Rodrigues[7], Jayra Abrantes[12], Jeová K. B. Colares[9], Kleber G. Luz[14], Luana Barbosa da Silva[10], Luiz Demarchi[11], Magaly C. B. Câmara[12], Marina C. S. Umaki Zardin[11], Rafaela Sabatini Mello Pinheiro[15], Rutilene Barbosa Souza[7], Simone K. Haddad[16], Stephanni Figueiredo da Silva[10], Svetoslav N. Slavov[16], Themis Rocha[12], Noelia Morel[17], Hector Chiparelli[17], Analía Burgueño[17], Victoria Bórmida[17], María N. Cortinas[17], Rosario S. Martín[17], Allan C. Pereira[18], Marcelo F. dos Santos[18], Walter André Júnior[18], Jairo Mendez Rico[19], Leticia Franco[19], Alexander Rosewell[2], Rodrigo F. do Carmo Said[2], Carlos F. C. de Albuquerque[2], Ethel L. Noia Maciel[20], Marília Santini de Oliveira[21], Rivaldo Venâncio da Cunha[22], Livia C. Vinhal Frutuoso[23], Ana M.B. de Filippis[5], Marta Giovanetti[24], Luiz Carlos Junior Alcantara[4]

1Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Brazil;
2Organização Pan-Americana da Saúde, Organização Mundial da Saúde, Brazil;
3Laboratorio Central de Saúde Pública do Estado de Minas Gerais, Fundação Ezequiel Dias, Brazil;
4Instituto Rene Rachou, Fundação Oswaldo Cruz, Minas Gerais, Brazil;
5Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil;
6Instituto Multidisciplinar em Saúde, Universidade Federal da Bahia, Brazil;
7Laboratório Central de Saúde Pública do Acre, Brazil;
8Secretaria de Saúde do estado do Mato Grosso do Sul, Brazil;
9Universidade de Fortaleza, Brazil;
10Laboratorio Central de Saúde Pública do Mato Grosso, Brazil;
11Laboratorio Central de Saúde Pública do Mato Grosso do Sul, Brazil;
12Laboratório Central de Saúde Pública do Rio Grande do Norte, Brazil;
13Secretaria Estadual de Saúde do estado do Acre, Brazil
14Universidade Federal do Rio Grande do Norte, Brazil;
15Universidade Federal do Acre, Brazil;
16Fundação Hemocentro de Ribeirão Preto, Brazil;
17Departamento de Laboratorios de Salud Pública, Ministerio de Salud Pública, Uruguay;
18Laboratório de Fronteira Tabatinga, Brazil;
19Pan American Health Organization, Washington DC, USA;
20Secretária de Vigilância em Saúde e Ambiente, Ministério da Saúde, Brazil;
21Coordenação Geral dos Laboratórios de Saúde Pública, Ministério da Saúde, Brazil
22Fundação Oswaldo Cruz, Instituto de Tecnologia em Imunobiológicos, Rio de Janeiro, Brazil
23Coordenação Geral das Arboviroses, Ministério da Saúde, Brazil;
24Sciences and Technologies for Sustainable Development and One Health, University of Campus Bio-Medico, Italy;

Abstract

The emergence and continued geographic expansion of arboviruses and the growing number of infected people have highlighted the need to develop and improve multiplex methods for rapid and specific detection of pathogens. Sequencing technologies are promising tools that can help in the laboratory diagnosis of conditions that share common symptoms, such as pathologies caused by emerging arboviruses. In this study, we integrated nanopore sequencing and the advantages of reverse transcription polymerase chain reaction (RT-PCR) to develop a multiplex RT-PCR protocol for the detection of chikungunya virus (CHIKV) and several orthoflaviviruses (such as dengue, Zika, yellow fever, and West Nile viruses) in a single reaction, which provides data for sequence-based differentiation of arbovirus lineages.

Introduction

Infectious diseases caused by viruses transmitted by mosquitoes have been making world headlines since arbovirus outbreaks have appeared in large urban areas. An estimated 3.83 billion people are currently living in areas at risk of Dengue, and this number is predicted to increase to 6.1 billion people by 2080, which represents 60% of the global population (1). Other mosquito-borne diseases such as Zika and Chikungunya fevers also represent an important threat to the health and economics of populations around the world, mainly in the region of the Americas where a total of 3.1 million cases of arboviral diseases were reported in 2022, a relative increase of 118.5% from 2021 to 2022 (2).

Many arboviral diseases lead to clinically indistinguishable febrile syndromes, making correct diagnosis challenging. For instance, laboratory differentiation of members of the genus *Orthoflavivirus* (formerly named *Flavivirus*) by serological methods is limited due to extensive cross-reactivity (3). Nucleic acid tests such as quantitative reverse transcription polymerase chain reaction (RT-qPCR) have been extensively used for pathogen detection due to their high specificity and sensitivity (4), however, available

tests cannot provide information for differentiating arbovirus at the level of distinct lineages.

DNA sequencing technology has been proven useful for control efforts of recent infectious disease outbreaks, by providing relevant epidemiological aspects regarding the dynamics of an epidemic (5). The nanopore sequencing platform is a potential tool for diagnostic purposes due to its cost-effectiveness, rapid turnaround time, and portability, allowing it to be employed in the investigation of several viral outbreaks (6–8). Hence, in this study, we integrated nanopore sequencing and the advantages of reverse transcription polymerase chain reaction (RT-PCR) to develop a multiplex RT-PCR protocol for the detection of chikungunya virus (CHIKV) (genus *Alphavirus*) and several orthoflaviviruses in a single reaction, which provides data for sequence-based differentiation of arbovirus lineages.

Materials and Methods

Primers design and selection

We identified a set of primers previously designed for real-time quantitative reverse transcription PCR (RT-qPCR) (9) which originally amplified a fragment of 260 bp of a conserved region in the NS5 gene of orthoflaviviruses. To take advantage of the long-read sequencing technology of the MinION, we performed an alignment of 28 reference sequences of common orthoflaviviruses to design a new reverse primer. When used alongside the forward primers (Flavi-all-S and Flavi-all-S2) from a previously published scheme (9), this new reverse primer enables the amplification of a fragment of around 1000 bases from the NS5 region.

For the detection of chikungunya virus's nucleic acid, we selected a pair of primers previously published for Nanopore sequencing (10) and redirected these primers to compose the primer scheme of the optimized protocol described in this study.

We evaluated the specificity of chikungunya primers *in silico* using the NCBI primer-blast tool against the NCBI Nucleotide collection. However, assessment of orthoflavivirus primers was not feasible through the primer-blast tool due to its limitation in accepting

ambiguity letters beyond “N” within the primer sequences. Consequently, we employed an alternative tool, MFEprimer v.3.0 [11], to conduct *in silico* checks for primer-specific amplification against the NCBI virus database for both the CHIKV and orthoflavivirus primers.

Viruses and clinical specimens selection

We used virus stocks of chikungunya (CHIKV), dengue (DENV) (types 1-4), Zika (ZIKV), yellow fever (YFV) and West Nile (WNV) viruses passaged in C6/36 cells in L-15 medium kindly provided by the Public Health Laboratory of Minas Gerais state (Fundação Ezequiel Dias) and the Flavivirus Laboratory of the Oswaldo Cruz Foundation Rio de Janeiro. Clinical specimens (serum) were provided as residual samples from the epidemiological surveillance routine of the Brazilian Central Public Health Laboratories (LACEN) from Minas Gerais, Rio Grande do Norte, Acre, and Amazonas states. Clinical samples were also obtained from genomic surveillance activities carried out in the Public Health Laboratory of Uruguay and in Ceará state (Universidade de Fortaleza, Brazil). This study was approved by the Pan American World Health Organization (PAHO/WHO) and by the Research Ethics Committee of the Universidade Federal de Minas Gerais with approval No. 32912820.6.1001.5149. Personal information has been de-identified to minimize the risk of unintended disclosure of the identity of individuals.

Viral RNA from both culture supernatant and clinical specimens was extracted using the chemagic Viral DNA/RNA 300 Kit H96 (PerkinElmer) according to the manufacturer’s instructions. Extracted RNA was subjected to molecular diagnosis by RT-qPCR to confirm the presence of viral RNA in samples. Obtained cycle threshold (Ct) values were then used as a proxy indicator of the amount of viral genetic material in the screened samples. RT-qPCR tests were performed on Applied Biosystems 7500 (Thermo Fisher Scientific Inc.) Instrument following previously described protocols (11–14) for the detection of YFV, ZIKV, CHIKV and DENV.

Primers validation with cultured viruses and clinical specimens

Selected primers were pooled to a concentration of 10μM and were initially tested with culture isolates of DENV (types 1-4), ZIKV, CHIKV, YFV, and WNV. One Step RT-PCR was performed using the QIAGEN OneStep RT-PCR Kit according to manufacturer instructions, in a total reaction of 25μL containing 4.7μL of nuclease-free water, 5μL of 5x QIAGEN OneStep RT-PCR Buffer, 1μL of dNTP mix, 6.25μL of primers (from the pool at 10μM to a final concentration of 0.5μM of each primer), 1μL of QIAGEN OneStep RT-PCR Enzyme Mix, and 7μL of viral RNA as template. RT-PCR temperature cycling conditions were as follows: 50 °C for 30 min (reverse transcription), 95 °C for 15 min (Initial PCR activation), followed by 40 cycles of 94 °C for 30s, 55 °C for 30s, and 72 °C for 90s, and lastly 72 °C for 10 min (final extension). The same RT-PCR conditions were applied for the validation of the primers using clinical specimens. All reactions contained No Template Controls (NTC) to assess the extension of cross-contamination between neighboring barcoded samples. Electrophoresis in 1% agarose gel was performed to visualize PCR products alongside GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific). Amplicon concentrations were then quantified using the Qubit fluorimeter with the Qubit dsDNA HS Assay Kit (Invitrogen).

Nanopore sequencing

Amplicons were purified using 0.8x AMPure XP beads (Beckman Coulter) and sample concentrations were normalized to an initial input of 10ng each. After this point, nanopore library preparation (barcode and adapter ligation) followed the protocol previously described by (10) which used the Ligation Sequencing kit (SQKLSK109 - Oxford Nanopore Technologies) and Native Barcoding Expansion (NBD104 and EXPNBD114, Oxford Nanopore Technologies) for multiplexing of 24 samples. Prepared sequencing libraries were loaded on a R9.4 flow cell and data were collected for 2 hours.

Bioinformatic analysis for virus identification

The raw data (Fast5 files) from the MinION sequencing was basecalled (with the FAST model) and demultiplexed using Guppy v6.0 (Oxford Nanopore Technologies). Consensus sequence generation, as well as the identification and classification of viral

species and lineages (15), were carried out with the Genome Detective software which uses SPAdes for nanopore single-end reads (16).

Consensus sequences from clinical samples of DENV (types 1-4), CHIKV, ZIKV and YFV-positive cases were used to reconstruct the phylogeny of these viruses. Reference sequences were downloaded from NCBI to build separate datasets of partial DENV2 NS5 (n=36) and CHIKV envelope (n=25) genes. After alignment of the sequences generated in this study using MAFFT version 7, these datasets were used to infer Maximum Likelihood (ML) phylogenies using IQ-TREE 2.1.1 (17), which also implemented an ultrafast bootstrap (UFBoot) to estimate tree nodes statistical support with 10,000 replicates. Distribution charts of reads number and RT-qPCR Ct values were built using custom scripts on R studio 2023.06.1.

Results

The sequences of two previously published forward primers (9) were used alongside a new reverse primer (containing degenerate bases in three positions) designed to amplify a fragment of around 1000 bases from a conserved region in the NS5 gene of orthoflaviviruses (**Figure 1A**). Additionally, a pair of primers was redirected from a published CHIKV sequencing primers scheme (10) to amplify a ~1100 bases fragment from the CHIKV envelope gene (**Figure 1A**). We used an *in silico* approach to assess the specificity of these selected primers by aligning against the NCBI virus database, which resulted in 8,120 potential amplicons covering 20 different virus species when using the orthoflavivirus primers (**Table S2**). The same analysis using the CHIKV primers resulted in 584 potential amplicons all from CHIKV isolates. In addition, a Primer-BLAST analysis returned 6,084 Blast hits of chikungunya virus genomes (See data availability to access the output from the *in silico* specificity analysis).

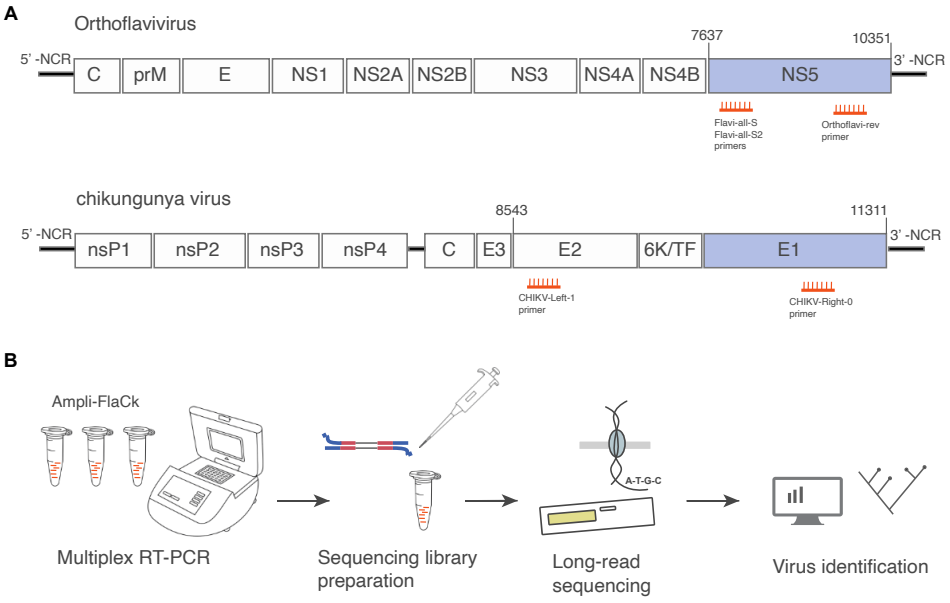


Figure 1. **A multiplex RT-PCR protocol for amplification and sequencing of nucleic acids of chikungunya virus and orthoflaviviruses (Ampli-FlaCk). A**) Organization of orthoflaviviruses and chikungunya virus (genus *Alphavirus*) genome. Primers (small orange lines) are depicted near their expected binding positions. Genome representations based on the following references for orthoflaviviruses and chikungunya virus, respectively: NC_002031 and KP164568.1. **B**) Schematic representation of the Ampli-FlaCk workflow.

These primers (a total of five) were pooled to achieve a multiplex RT-PCR protocol for the amplification of chikungunya's and orthoflaviviruse's nucleic acid (namely Ampli-FlaCk) in a single reaction. Ampli-Flack is able to amplify long amplicons, which when combined with nanopore sequencing can generate long sequences (~1000bp) from multiple distinct samples, allowing the identification and genotyping of chikungunya's and orthoflaviviruse's distinct lineages (**Figure 1B**). The nucleotide sequence and physical properties of the primers comprising the Ampli-FlaCk protocol are listed in **Table S1**.

To validate the protocol, the RNA from different cultured viruses (DENV, ZIKV, YFV, WNV) was extracted and used in RT-PCR reactions with the Ampli-FlaCk primers and the OneStep RT-PCR Kit which allows both reverse transcription and PCR amplification

to occur in the same reaction mix. Gel electrophoresis revealed clearly visible bands corresponding to the expected amplicons for each virus (**Figure 2A and Figure S1**). Then, the amplicons were barcoded and sequenced on a MinION (nanopore sequencing) for 2 hours. After basecalling, viral reads were analyzed using the Genome Detective software (16), an online tool of easy usage that performs read mapping, consensus creation, and genotyping of viruses such as ZIKV, DENV, CHIKV, and YFV (15).

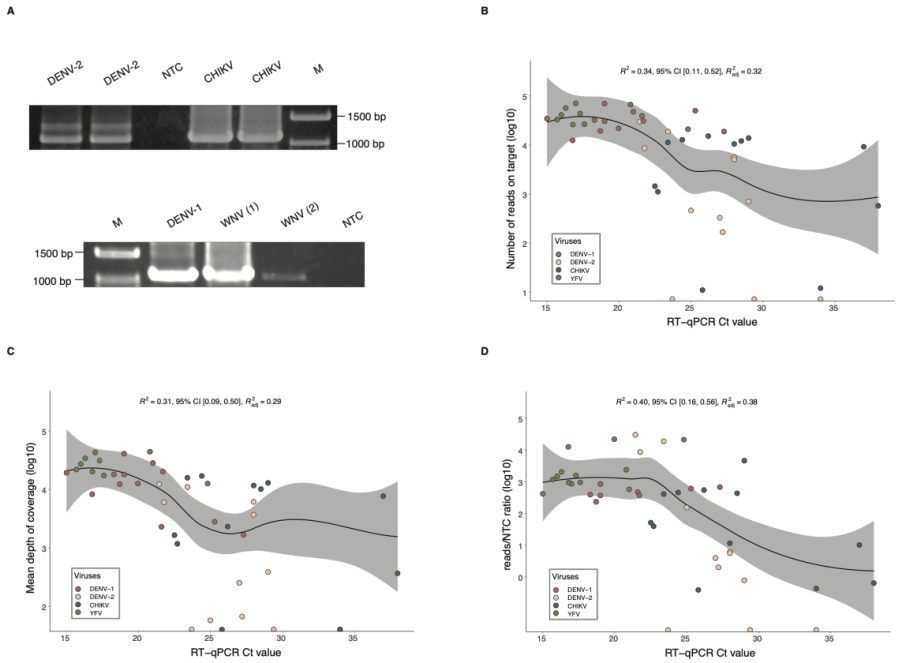


Figure 2. **Sequencing performance of the Ampli-FlaCk protocol. A**) Electrophoresis analysis of RT-PCR products from cultured dengue virus type 2 (DENV-2, Ct=22) and chikungunya virus (CHIKV, Ct=33), in duplicates (top image). The bottom image displays electrophoresis results for cultured dengue virus type 1 (DENV-1, Ct=25) and West Nile virus (WNV, numbers in parentheses represent samples with different Ct values, 1=21, and 2=33). M=Marker. Images were cropped and converted to grayscale for clarity; see Data Availability section for original unedited images. **B**) Distribution of the number of reads on target and the RT-qPCR Ct values from clinical specimens (n=45) used for testing the protocol. **C**) Distribution of the mean depth of coverage and the RT-qPCR Ct values from tested clinical specimens (n=45). **D**) Distribution of the reads/NTC ratio and the RT-qPCR Ct values from tested clinical specimens (n=45). **B-D**) The dark blue line

represents a smooth local regression line (method = 'loess') and the light grey area around the trend line represents the 95% confidence interval.

Sequencing results showed that the protocol can detect all eight viral species and generate nucleotide sequences for successful discrimination of viral lineages (**Table 1**). This nanopore library yielded an average of 5,682 reads for each sequenced virus, with a mean percentage of 1.75% of off-target reads (reads originated from random crosstalk between neighboring samples). An average 58.84 reads/NTC ratio was also estimated for this experiment. This ratio indicates the sample's number of reads on target normalized by the number of reads found in the no template control, and a ratio >10 was arbitrarily defined to indicate a reliable positive sample for the assigned virus, as previously suggested (18).

Table 1. Summary of sequencing statistics of cultured viruses used for evaluating the performance of the Ampli-FlaCk protocol.

Sample	Ct	Number of reads on target	off-target reads (%)	Reads/NTC ratio	Lineage
DENV-1	25	4095	0.7	67.1	Genotype I
DENV-2	22	4058	0.8	66.5	Genotype III
DENV-3	19	4720	1.0	77.4	Genotype V
DENV-4	28	3390	2.3	55.6	Genotype II
YFV	28	4823	1.2	79.1	South America I
ZIKV	25	2449	1.4	40.1	Asian
WNV	33	3967	3.9	65.0	Lineage 1A
CHIKV	33	17954	2.8	19.8*	ECSA
NTC	-	61	-	-	-
Mean (SD)		5682 (5015.23)	1.75 (1.14)	58.84 (20)	

SD= Standard Deviation
NTC= No Template Control
*The CHIKV sample was sequenced on a separate library/flowcell and in this library 905 reads were detected in the NTC.

Serial dilutions of these eight viruses were also sequenced following the protocol conditions described earlier and the results allowed to estimate the limit of detection of

the Ampli-FlaCk protocol using the dilution's Ct values as proxy indicators of viral load. Sequencing results showed that the highest Ct values that generated enough data for reliable detection ranged from 30 to 37 for dengue viruses, while CHIKV diluted samples could be positively detected up to a Ct value of 40 (full results are listed in **Table S3**).

Known positive clinical specimens were selected to validate the Ampli-FlaCk protocol under the same reaction conditions described earlier. Extracted RNA from 14 DENV-1, 13 DENV-2, 14 CHIKV, 1 ZIKV, and 7 YFV positive samples was subjected to multiplex RT-PCR and sequenced on a MinION for up to 6 hours to maximize sequencing yield per sample. Analysis of the reads showed that the protocol was able to reliably identify the viral lineage in 83.67% (41/49) of the clinical samples (**Table 2 and Table S4**). Although some viral reads were detected in 8 samples (16.33%), the reads/NTC ratio from these samples was lower than the threshold to consider a positive result. The median of 10,229 mapped reads per sample was calculated for the combined sequenced libraries, with long reads of ~1000bp being recovered from 89.80% (44/49) of the samples. Plotting the number of mapped reads against Ct values as a proxy indicator for viral load showed a downward trend in the number of reads as Ct values increased (**Figure 2B**). As expected, similar trends were also observed for depth of coverage and reads/NTC ratio (**Figure 2C and D**).

Table 2. Summary of sequencing statistics of clinical specimens used for evaluating the performance of the Ampli-FlaCk protocol.

Virus	# of samples	Ct	# of reads	% off-target reads	Reads/NTC ratio	Genotyping	Depth of coverage[1]	Sequence length
DENV-1	14	19.49 (3.5)	31754 (18341.4)	0.01 (0.43)	590.41 (7896.47)	Genotype V	15387.15 (12991.5)	984.5 (38.3)
DENV-2	13	27.1 (3.6)	705 (9158.4)	4.33 (18.17)	80.10 (10489.13)	Genotype III	2282.45 (4641.9)	1032.5 (109.5)
CHIKV	14	26.2 (4.6)	10229 (18977.3)	3.64 (31.15)	228.45 (6655.30)	ECSA	10946.6 (21163.4)	1051.5 (90.5)
YFV	7	16.8 (1.7)	41143 (15275.8)	0.14 (0.04)	1469.39 (545.57)	South America I	27291 (9394.7)	1044 (49.5)
ZIKV	1	ND	62	0	62	Asian	30.4	960

Median values (Standard deviation)
ND=Not Detected
[1]Mean depth of coverage was estimated by SPAdes implemented on Genome Detective using the following function: *read count * read length / contig length.*

Once established its efficacy to recover viral long reads from clinical samples, the Ampli-FlaCk protocol was employed for testing in field activities of an arbovirus genomic surveillance project carried out in Uruguay (early 2023) (19) and across Brazil over the years 2021 and 2022 (7). Sequencing of 20 clinical samples collected during those surveillance activities resulted in the recovery and identification of reads of ZIKV, DENV-2, DENV-3, and DENV-4 (**Table S5**). It is worth mentioning that the protocol was also able to accurately recover reads of a recently detected lineage of DENV-2 in Brazil (20), the genotype II (cosmopolitan). To demonstrate that consensus sequences generated with the Ampli-FlaCk protocol can also be used to reconstruct phylogenetic trees accurately, we performed phylogenic analyses using DENV-2 and CHIKV sequences from positive clinical samples used in this study. The maximum likelihood tree showed that CHIKV sequences grouped with other ECSA lineage isolates, while DENV-2 sequences were placed either in the genotype II – cosmopolitan clade or genotype III clade (**Figure 3**). Phylogenetic reconstructions were also performed for the other viruses sequenced in this study and can be seen in supplementary figures S2 and S3.

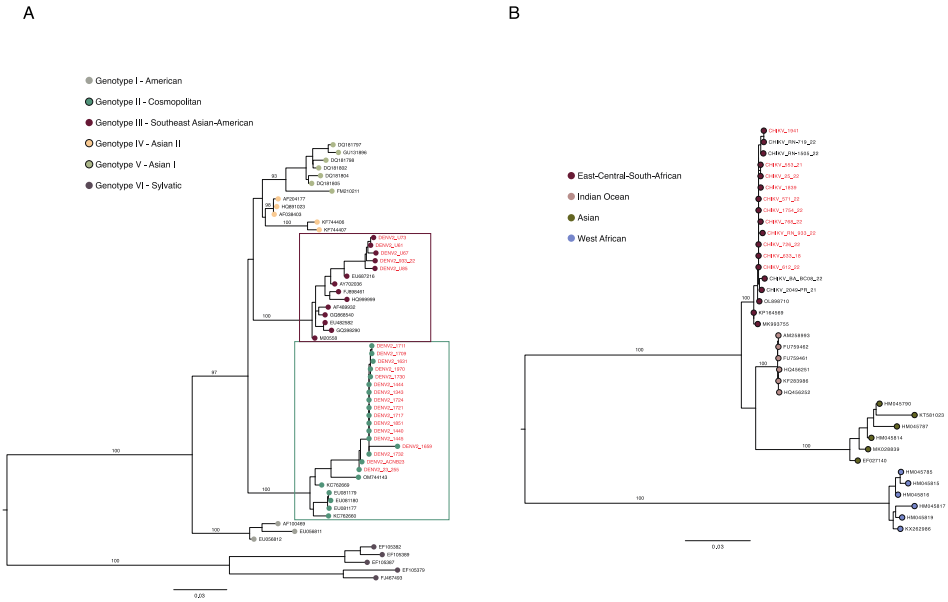


Figure 3. Reconstructed genotyping phylogeny of dengue virus type 2 and chikungunya virus. A) Maximum likelihood phylogenetic reconstruction of DENV-2 using partial sequences (n=58) of the NS5 gene, including sequences (n=22, red tips) generated in this study from clinical samples. **B**) Maximum likelihood phylogenetic reconstruction of CHIKV using partial sequences (n=36) of the envelope gene, including sequences (n=11, pink tips) generated in this study from clinical samples. Numbers along branches represent Ultrafast bootstrap values (only for main branches to maintain clarity).

Discussion

DNA sequencing technologies are promising tools that can help in the laboratory detection of emerging and reemerging pathogens, generating information that can contribute to the fight against epidemics (21, 22). In this study, we developed a protocol for the detection and genotyping of multiple viral species in a single reaction that combines the advantages of high sensitivity and specificity of multiplex PCR with the portability and speed of data generation provided by nanopore sequencing, which can perform simultaneous sequencing of up to 96 samples in a single experiment. The Ampli-FlaCk protocol also incorporates the benefits of the OneStep RT-PCR Kit that allows both reverse transcription and PCR amplification to take place in a single reaction (23).

Clinical samples were used for assessing the protocol performance, which allowed viral identification in 83% of the samples. After sequencing, it seems that the number of mapped reads followed a downward trend when Ct values increased. This relationship has been previously reported (24) for whole genome sequencing of arboviruses from clinical samples and upholds the importance of sampling during the acute period of infections when a high viral load is present. The Ct value of 37 is considered the limit of detection for dengue samples tested with the Trioplex Real-time RT-PCR Assay (CDC) (25). Comparatively, the Ampli-FlaCk protocol was able to detect DENV sample dilutions with the highest Ct values ranging from 30 to 37 according to the serotype, while CHIKV dilutions were detected up to Ct 40.

The diagnosis of orthoflavivirus infections by serological methods is challenging due to limitations imposed by the broad antigenic cross-reactivity among common orthoflaviviruses(26). The protocol described here was shown to contribute to the successful detection of all orthoflaviviruses tested here using long nucleotide sequences, which were also used for viral lineage assignments. Although this work has not tested other viruses, the protocol has the potential to detect other orthoflavivirus species due to the use of primers with degenerate bases that cover the nucleotide diversity of the NS5 gene, as suggested by the *in silico* analyses.

The protocol described in this study allowed us to obtain ~1K long reads from the NS5 and envelope genes of orthoflaviviruses and CHIKV, respectively. These reads can be used not only for identification but also for classification of viral lineages, with the reconstruction of phylogenetic trees and execution of evolutionary analysis that might help in understanding patterns of viral dispersion during an epidemic. The protocol practicality associated with the portability of the MinION device offers a possibility for using this approach in field studies in isolated areas. Furthermore, the ability to wash and reuse the flowcells allows for a reduction in costs for screening hundreds of samples considering the protocol's multiplex approach. Hence, the development of sequencing-based viral detection methods integrated into local epidemiological surveillance is

fostered in order to achieve a more anticipatory approach in public health to epidemic prevention and control (27).

Institutional Review Board Statement: This research was reviewed and approved by the Research Ethics Committee of the of the Universidade Federal de Minas Gerais with approval No. 32912820.6.1001.5149, by the Pan American World Health Organization (No. PAHO-2016-08-0029), the Oswaldo Cruz Foundation Ethics Committee (CAAE: 45279715.8.0000.0040), and the Brazilian Ministry of Health (MoH) as part of the arbovirus genomic surveillance efforts.

Data Availability

The new sequences generated in this study have been deposited in NCBI GenBank under accession numbers listed in Supplementary Table S6. Input data used for the phylogenetic analyses are provided on the repository https://doi.org/10.6084/m9.figshare.24270724.v2 .

Code Availability

All R scripts used in this study were made available on the repository https://doi.org/10.6084/m9.figshare.24270724.v2 .

Acknowledgements: This study was supported by The Pan American Health Organization (PAHO/WHO), the Brazilian Ministry of Health grant SCON2021-00180 (Coordenação Geral de Laboratório de Saúde Pública-CGLAB and Coordenação Geral de Vigilância de Arboviroses-CGARB), by the Programa Inova Fiocruz/ Fundação Oswaldo Cruz, the National Institutes of Health USA grant U01 AI151698 for the United World Arbovirus Research Network (UWARN), and in part by the CRP- ICGEB RESEARCH GRANT 2020 Project CRP/BRA20-03, Contract CRP/20/03. J.X. was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES)-Finance Code 001. MG is funded by PON “Ricerca e Innovazione” 2014-2020.

Author Contributions

Conceptualization: L.C.J.A., M.G., J.X.; methodology: L.C.J.A., M.G., J.X; formal analysis: J.X, M.G, V.F., M.L., C.O., L.C.J.A.; investigation: J.X., V.F., T.A., F.I., G.C.P., M.M.D., M.L., E.C., C.O., H.F., N.G., L.O.L., F.K.B., C.M.M.B.O., C.C.M.G., D.M.L., E.C.O., G.G.C.L., I.G., J.M., J.T.N.R., J.A., J.K.B.C., K.G.L., L.B.S., L.D., M.C.B.C., M.C.S.U.Z., R.S.M.P., R.B.S., S.K.H., S.F.S., S.N.S., T.R., N.M., H.C., A.B., V.B., M.N.C., R.S.M, A.C.P., M.F.S., W.A.J.; resources: J.M.R., L.F., A.R., R.F.C.S., C.F.C.A., E.L.N.M., M.S.O., R.V.C., L.C.V.F., A.M.B.F., L.C.J.A; writing-original draft: J.X; writing-review and editing: L.C.J.A., M.G., J.X; funding acquisition: L.C.J.A, J.M.R., A.R., R.F.C.S., C.F.C.A. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no conflict of interest.

References

1. Messina,J.P., Brady,O.J., Golding,N., Kraemer,M.U.G., Wint,G.R.W., Ray,S.E., Pigott,D.M., Shearer,F.M., Johnson,K., Earl,L., *et al.* (2019) The current and future global distribution and population at risk of dengue. *Nat. Microbiol.*, **4**, 1508–1515.

2. Pan American Health Organization (2023) Annual Arbovirus Bulletin 2022. Regional Office for the Americas of the World Health Organization https://www3.paho.org/data/index.php/en/mnu-topics/indicadores-dengue-en/annual-arbovirus-bulletin-2022.html (27 September 2023, date last accessed).

3. Wilder-Smith,A., Gubler,D.J., Weaver,S.C., Monath,T.P., Heymann,D.L. and Scott,T.W. (2017) Epidemic arboviral diseases: priorities for research and public health. *Lancet Infect. Dis.*, **17**, e101–e106.

4. Kralik,P. and Ricchi,M. (2017) A basic guide to real time PCR in microbial diagnostics: definitions, parameters, and everything. *Front. Microbiol.*, **8**, 108.

5. Maljkovic Berry,I., Melendrez,M.C., Bishop-Lilly,K.A., Rutvisuttinunt,W., Pollett,S., Talundzic,E., Morton,L. and Jarman,R.G. (2020) Next generation sequencing and bioinformatics methodologies for infectious disease research and public health: approaches, applications, and considerations for development of laboratory capacity. *J. Infect. Dis.*, **221**, S292–S307.

6. Adelino,T.É.R., Giovanetti,M., Fonseca,V., Xavier,J., de Abreu,Á.S., do Nascimento,V.A., Demarchi,L.H.F., Oliveira,M.A.A., da Silva,V.L., de Mello,A.L.E.S., *et al.* (2021) Field and classroom initiatives for portable sequence-based monitoring of dengue virus in Brazil. *Nat. Commun.*, **12**, 2296.

7. Xavier,J., Alcantara,L.C.J., Fonseca,V., Lima,M., Castro,E., Fritsch,H., Oliveira,C., Guimarães,N., Adelino,T., Evaristo,M., *et al.* (2023) Increased interregional virus exchange and nucleotide diversity outline the expansion of chikungunya virus in Brazil. *Nat. Commun.*, **14**, 4413.

8. Giovanetti,M., Slavov,S.N., Fonseca,V., Wilkinson,E., Tegally,H., Patané,J.S.L., Viala,V.L., San,E.J., Rodrigues,E.S., Santos,E.V., *et al.* (2022) Genomic epidemiology of the SARS-CoV-2 epidemic in Brazil. *Nat. Microbiol.*, **7**, 1490–1500.

9. Patel,P., Landt,O., Kaiser,M., Faye,O., Koppe,T., Lass,U., Sall,A.A. and Niedrig,M. (2013) Development of one-step quantitative reverse transcription PCR for the rapid detection of flaviviruses. *Virol. J.*, **10**, 58.

10. Quick,J., Grubaugh,N.D., Pullan,S.T., Claro,I.M., Smith,A.D., Gangavarapu,K., Oliveira,G., Robles-Sikisaka,R., Rogers,T.F., Beutler,N.A., *et al.* (2017) Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.*, **12**, 1261–1276.

11. Lanciotti,R.S., Kosoy,O.L., Laven,J.J., Velez,J.O., Lambert,A.J., Johnson,A.J., Stanfield,S.M. and Duffy,M.R. (2008) Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. *Emerging Infect. Dis.*, **14**, 1232–1239.

12. Lanciotti,R.S., Kosoy,O.L., Laven,J.J., Panella,A.J., Velez,J.O., Lambert,A.J. and Campbell,G.L. (2007) Chikungunya virus in US travelers returning from India, 2006. *Emerging Infect. Dis.*, **13**, 764–767.

13. Johnson,B.W., Russell,B.J. and Lanciotti,R.S. (2005) Serotype-specific detection of dengue viruses in a fourplex real-time reverse transcriptase PCR assay. *J. Clin. Microbiol.*, **43**, 4977–4983.

14. Domingo,C., Patel,P., Yillah,J., Weidmann,M., Méndez,J.A., Nakouné,E.R. and Niedrig,M. (2012) Advanced yellow fever virus genome detection in point-of-care facilities and reference laboratories. *J. Clin. Microbiol.*, **50**, 4054–4060.

15. Fonseca,V., Libin,P.J.K., Theys,K., Faria,N.R., Nunes,M.R.T., Restovic,M.I., Freire,M., Giovanetti,M., Cuypers,L., Nowé,A., *et al.* (2019) A computational method for the identification of Dengue, Zika and Chikungunya virus species and genotypes. *PLoS Negl. Trop. Dis.*, **13**, e0007231.

16. Vilsker,M., Moosa,Y., Nooij,S., Fonseca,V., Ghysens,Y., Dumon,K., Pauwels,R., Alcantara,L.C., Vanden Eynden,E., Vandamme,A.-M., *et al.* (2019) Genome Detective: an automated system for virus identification from high-throughput sequencing data. *Bioinformatics*, **35**, 871–873.

17. Nguyen,L.-T., Schmidt,H.A., von Haeseler,A. and Minh,B.Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.*, **32**, 268–274.

18. Wang,M., Fu,A., Hu,B., Tong,Y., Liu,R., Liu,Z., Gu,J., Xiang,B., Liu,J., Jiang,W., *et al.* (2020) Nanopore Targeted Sequencing for the Accurate and Comprehensive Detection of SARS-CoV-2 and Other Respiratory Viruses. *Small*, **16**, e2002169.

19. Burgueno,A., Giovanetti,M., Fonseca,V., Morel,M.N., Lima,M., Castro,E., Guimaraes,N.R., Iani,F.C.M., Bormida,V., Cortinas,M.N., *et al.* (2023) Genomic and eco-epidemiological investigations in Uruguay reveal local Chikungunya virus transmission dynamics during its expansion across the Americas in 2023. *medRxiv*, doi: 10.1101/2023.08.17.23294156.

20. Giovanetti,M., Pereira,L.A., Santiago,G.A., Fonseca,V., Mendoza,M.P.G., de Oliveira,C., de Moraes,L., Xavier,J., Tosta,S., Fristch,H., *et al.* (2022) Emergence of dengue virus serotype 2 cosmopolitan genotype, brazil. *Emerging Infect. Dis.*, **28**, 1725–1727.

21. Grubaugh,N.D., Ladner,J.T., Lemey,P., Pybus,O.G., Rambaut,A., Holmes,E.C. and Andersen,K.G. (2019) Tracking virus outbreaks in the twenty-first century. *Nat. Microbiol.*, **4**, 10–19.

22. Alcantara,L.C.J., Amenga-Etego,L., Andersson,R., Bhaumik,M., Choi,Y.K., Decaluwe,H., Geoghegan,J., Haagmans,B.L., López,S., Mukhtar,M.M., *et al.* (2022) Methods for fighting emerging pathogens. *Nat. Methods*, **19**, 395–397.

23. De Paula,S.O., de Melo Lima,C., Torres,M.P., Pereira,M.R.G. and Lopes da Fonseca,B.A. (2004) One-Step RT-PCR protocols improve the rate of dengue diagnosis compared to Two-Step RT-PCR approaches. *J. Clin. Virol.*, **30**, 297–301.

24. Naveca,F.G., Claro,I., Giovanetti,M., de Jesus,J.G., Xavier,J., Iani,F.C. de M., do Nascimento,V.A., de Souza,V.C., Silveira,P.P., Lourenço,J., *et al.* (2019) Genomic, epidemiological and digital surveillance of Chikungunya virus in the Brazilian Amazon. *PLoS Negl. Trop. Dis.*, **13**, e0007065.

25. Centers for Disease Control and Prevention-CDC (2020) CDC DENV-1-4 rRT-PCR Multiplex and Trioplex rRT-PCR Assays - Dengue - CDC. https://www.cdc.gov/dengue/healthcare-providers/testing/molecular-tests/assays.html (30 August 2023, date last accessed).

26. Chan,K.R., Ismail,A.A., Thergarajan,G., Raju,C.S., Yam,H.C., Rishya,M. and Sekaran,S.D. (2022) Serological cross-reactivity among common flaviviruses. *Front. Cell. Infect. Microbiol.*, **12**, 975398.

27. Gardy,J.L. and Loman,N.J. (2018) Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat. Rev. Genet.*, **19**, 9–20.

A multiplex nanopore sequencing approach for the detection of multiple arboviral species

Supplementary data

Table S1. Nucleotide sequence and physical properties of the primers comprising the Ampli-FlaCk protocol.

Name	Sequence	Length	Gene	Binding positions	GC content (%)	Melt temperature (mean)
Orthoflavi-rev	CATGTCDTCDGTNGTCATCCA	21	NS5	10054-10075	48.4	55.8 ºC
Flavi-all-S	TACAACATGATGGGGAARAGAGARAA	26	NS5	8993-9019	38.5	56.8 ºC
Flavi-all-S2	TACAACATGATGGGMAAACGYGARAA	26	NS5	8993-9018	40.4	59 ºC
CHIK_LEFT_1	CATGTACGCACCCATTTCACCA	22	E2	8913-8934	50	58.8 ºC
CHIK_RIGHT_0	CGGGATCACTGTTACGTGTTCG	22	E1	9997-10018	54.5	58.7 ºC

Primers physical properties were calculated using the OligoAnalyzer (Integrated DNA Technologies).
Binding positions refer to reference NC_002031 and KP164568.1 reference sequences for orthoflaviviruses (primes Orthoflavi-rev, Flavi-all-S, and Flavi-all-S2) and chikungunya virus (primers CHIK_LEFT_1 and CHIK_RIGHT_0), respectively.

Table S2. List of potential orthoflaviviruses identified during *in silico* specificity assessment using MFEprimer.

Virus name
dengue virus type 1
West Nile virus
dengue virus type 2
dengue virus type 3
Zika virus
Japanese encephalitis virus
Yellow fever virus
Tick-borne encephalitis virus
Louping ill virus
Alkhumra hemorrhagic fever virus
Bagaza virus
Powassan virus
Kyasanur Forest disease virus
Langat virus
Israel turkey meningoencephalomyelitis virus

	Omsk hemorrhagic fever virus
	Sepik virus
	dengue virus type 4
	Murray Valley encephalitis virus
	Wesselsbron virus

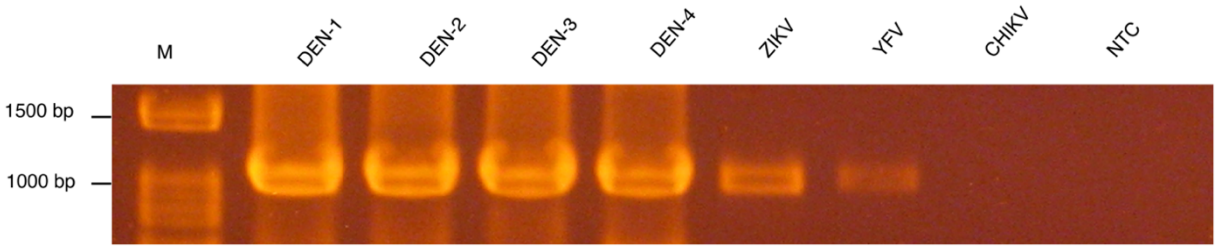


Figure S1. Validation of the Ampli-FlaCk primers on cultured viruses. RNA from cultured viruses (Chikungunya (CHIKV), dengue (DENV) 1 to 4, zika (ZIKV), yellow fever (YFV), and West Nile (WNV) viruses were used in RT-PCR with the three orthoflavivirus primers only (CHIKV primers were not included so specificity of the orthoflavivirus primers could be tested). Viruses presented the following RT-qPCR Ct values: DENV-1 (25), DENV-2 (22), DENV-3 (19), DENV-4 (28), CHIKV (33), YFV (28), and ZIKV (25). NTC= No Template Control.

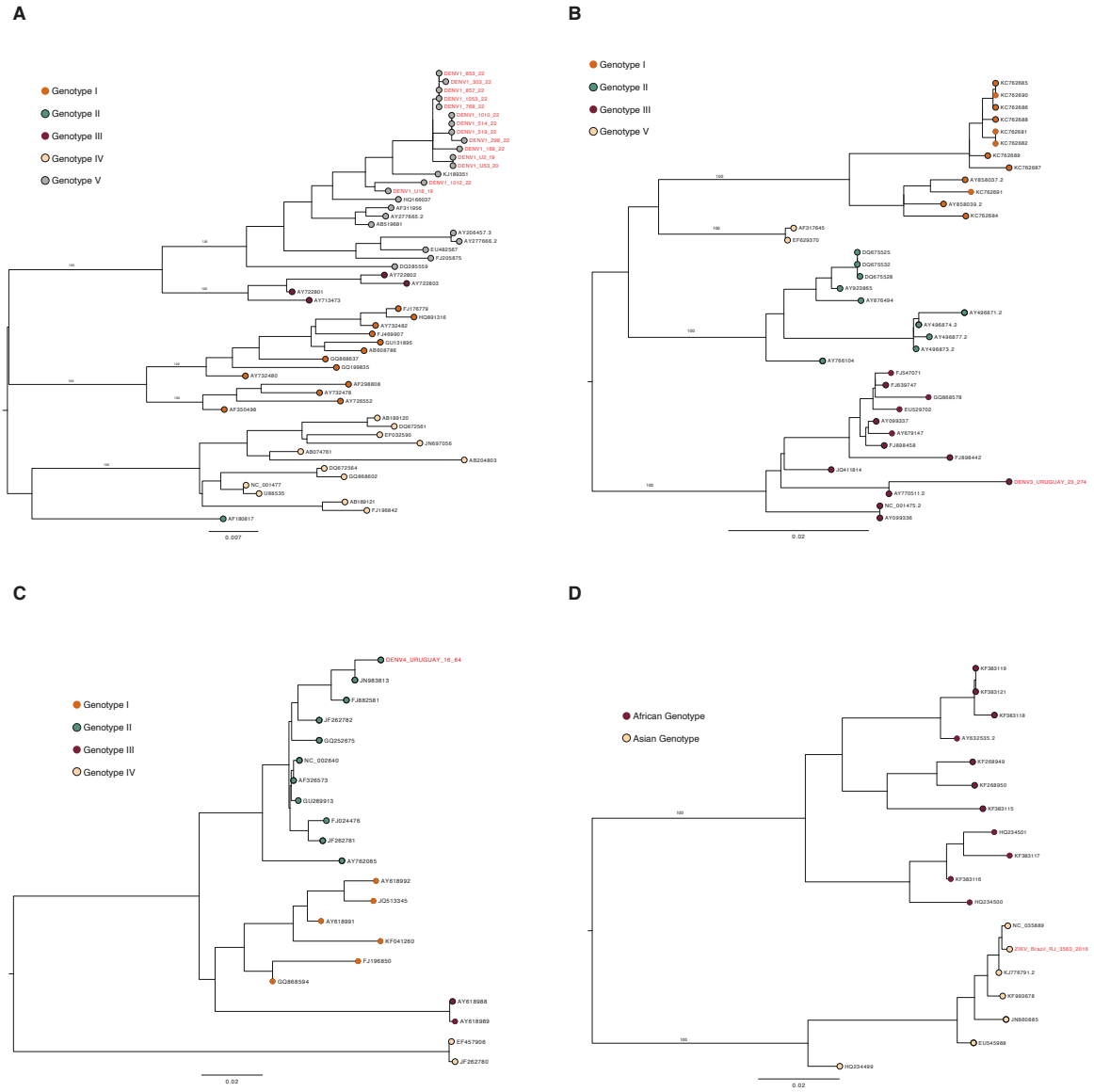


Figure S2. Maximum Likelihood phylogenies reconstructed using the new sequences generated in this study and lineages reference sequences for genotyping. A) Phylogeny of dengue virus serotype 1. New sequences' IDs (n=14) are colored in red. B) Phylogeny of dengue virus serotype 3. The new sequence's ID (n=1) is colored in red. C) Phylogeny of dengue virus serotype 4. New sequences' IDs (n=6) are colored in red. D) Phylogeny of zika virus. The new sequence's ID (n=1) is colored in red. A-D) Lineages are indicated by different colored circles at the tips of the branches.

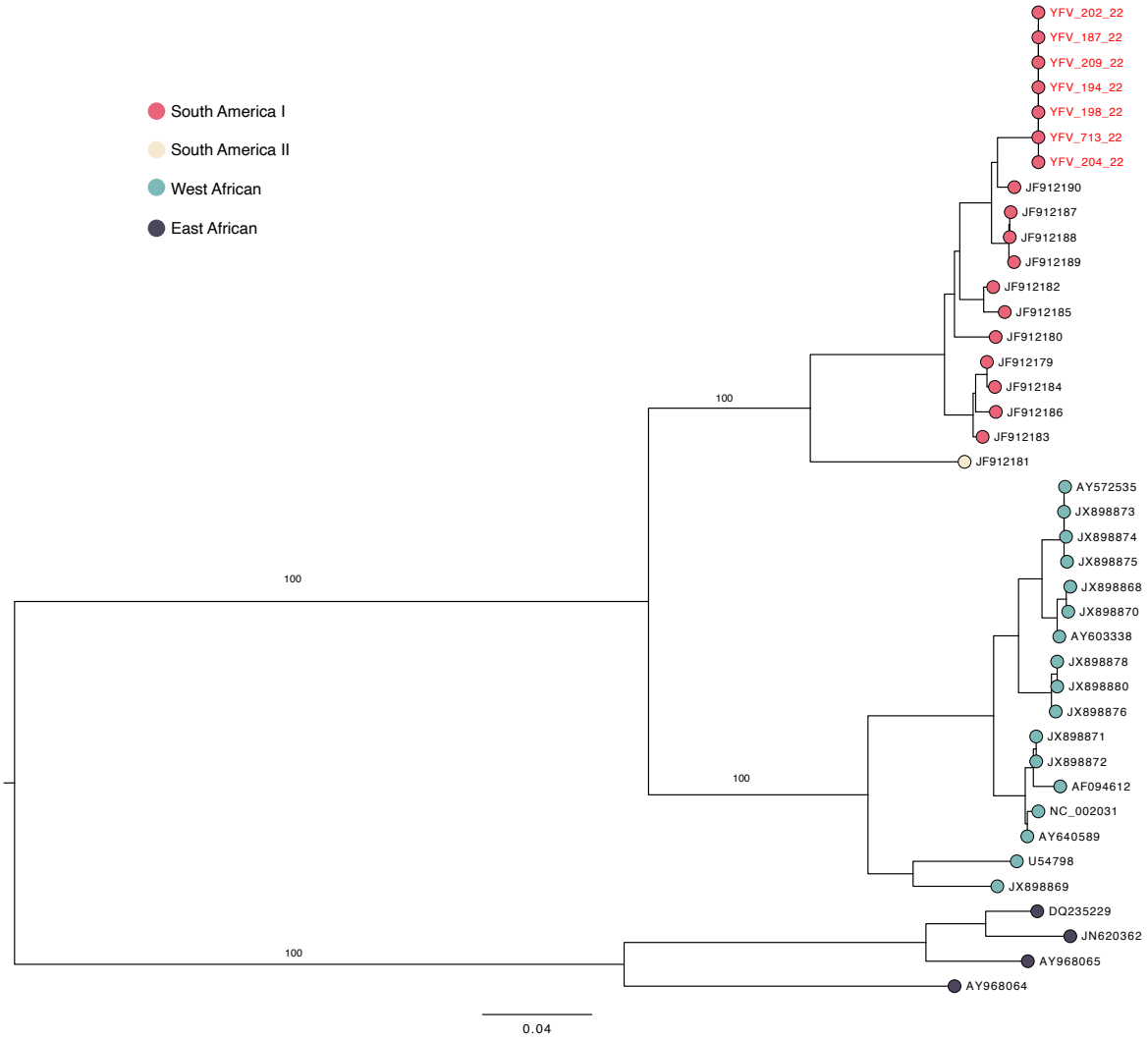


Figure S3. Maximum Likelihood phylogeny of yellow fever virus reconstructed using the new sequences generated in this study and lineages reference sequences for genotyping. New sequences' IDs (n=7) are colored in red. Lineages are indicated by different colored circles at the tips of the branches.

Table S3. Sequencing results of serially diluted virus isolates.

Virus	Ct	Dilution	Number of reads on target	reads/NTC ratio
DENV-1	27.1	10^{-3}	27651	141.8
	29.93	$2x10^{-4}$	13771	70.6
	37.07	***10^{-5}***	***546***	***91***
	40	10^{-6}	3	0.5
DENV-2	33.42	10^{-5}	6692	239
	33.39	***$2x10^{-6}$***	***1932***	***69***
	33.95	$13.3x10^{-7}$	13	0.5
DENV-3	30.98	10^{-4}	4198	149.9
	30.31	***$2x10^{-5}$***	***367***	***13.1***
	31.65	$13.33x10^{-6}$	57	2
DENV-4	30.05	$2x10^{-4}$	9954	51
	35.77	***10^{-5}***	***151***	***25.2***
	38.5	10^{-6}	1	0.2
ZIKV	***28.84***	***10^{-3}***	***3916***	***47.2***
	37.85	10^{-4}	250	3
	37.91	10^{-5}	512	6.2
CHIKV	35.9	10^{-1}	22110	172.7
	40.68	***10^{-2}***	***19525***	***152.5***
	ND	10^{-3}	78	0.6
YFV	***31.64***	***10^{-5}***	***9817***	***350.6***
	36.91	$2x10^{-7}$	18	0.6
	37.34	$4x10^{-8}$	19	0.7

ND=Not Detected

The detection limit for each virus is indicated by numbers highlighted in bold and was determined by a reads/NTC ratio > 10.

Table S4. Sequencing and genotyping results of clinical specimens used for validation of the Ampli-FlaCk protocol.

Library	Virus	Sample	Ct	Concentration [ng/µL][1]	Reads on target	off-target reads (%)	reads/NTC ratio	Genotyping	Bootstrap[2]	Depth[3]	Sequence length
1	DENV-2	854-18	28	0.53	5608	6.95	6.20	Genotype III	100	6162.4	971
	DENV-2	856-18	29	0.34	705	42.82	0.78	Genotype III	100	386.3	969
	DENV-2	1041-16	28	11.90	5081	6.56	5.61	Genotype III	100	3693.9	1044
	CHIKV	633-18	28	118.00	10504	4.72	11.61	ECSA	92.4	11715.7	1173
	CHIKV	553-21	37	5.42	9218	6.06	10.19	ECSA	88.2	7690.4	1056
	CHIKV	1092-18	ND	0.45	578	57.44	0.64	ECSA	88.2	366.8	1018
	NTC	NTC	-	0.00	905.0	-	-	-	-	-	-
2	DENV-1	U2	20	36.20	21792	1.59	21792.00	Genotype V	100	12706.5	993
	DENV-1	U18	16.8	18.90	12457	0.10	12457.00	Genotype V	100	8300.4	971
	DENV-1	U53	24.8	2.56	21072	0.56	21072.00	Genotype V	100	12630.2	976
	DENV-2	U61	NA	0.40	1120	2.10	1120.00	Genotype III	100	871	1021
	DENV-2	U67	23.4	5.56	18805	0.07	18805.00	Genotype III	100	11026.3	967
	DENV-2	U73	21.4	0.70	30119	0.07	30119.00	Genotype III	100	12228.6	1159
	DENV-2	U85	21.8	4.88	8621	0.47	8621.00	Genotype III	100	6057.8	1305
	NTC	NTC	-	0.20	0	-	-	-	-	-	-
3	CHIKV	1941	NA	3.14	75300	0.00	25100.00	ECSA	87.2	78885.7	1029
	CHIKV	1839	NA	7.20	9955	0.01	3318.33	ECSA	88	11739	1010
	CHIKV	**933**	29	13.60	13887	0.03	4629.00	ECSA	94.8	13017.8	1326
	DENV-2	**933**	25	13.60	462	0.86	154.00	Genotype III	100	57.4	1128
	NTC	NTC	-	0.352	3	-	-	-	-	-	-
4	DENV-1	169.22	18.3	86.20	32376	0.01	390.07	Genotype V	100	18142.3	994
	DENV-1	298.22	25.3	33.80	50287	0.01	605.87	Genotype V	98	2813.5	1072
	DENV-1	303.22	19	50.40	30634	0.03	369.08	Genotype V	100	18067.8	980
	DENV-1	514.22	18.7	63.20	19290	0.00	232.41	Genotype V	100	12325.3	986
	DENV-1	519.22	21.6	51.00	39341	0.00	473.99	Genotype V	100	2299.1	1053

Library	Virus	Sample	Ct	Concentration [ng/µL][1]	Reads on target	off-target reads (%)	reads/NTC ratio	Genotyping	Bootstrap[2]	Depth[3]	Sequence length
	DENV-1	1010.22	21	63.20	47721	0.00	574.95	Genotype V	100	28465.2	969
	DENV-1	1012.22	15	87.80	34577	0.00	416.59	Genotype V	100	19409.8	978
	DENV-1	1053.22	17	78.60	71161	0.00	857.36	Genotype V	100	43067.4	998
	DENV-1	853.22	21.7	95.40	31133	0.03	375.10	Genotype V	100	20268.2	980
	DENV-1	857.22	19	77.40	70155	0.00	845.24	Genotype V	99	41036.7	983
	DENV-2	745.22	29.4	0.00	0	-	-	-	-	-	-
	DENV-2	723.22	27.2	0.53	167	48.30	2.01	Genotype III	100	66.9	993
	DENV-2	1042.22	27	0.23	328	15.25	3.95	Genotype III	100	252.4	1115
	NTC	NTC	-	0.00	82	-	-	-	-	-	-
	DENV-2	1043.22	23.7	0.15	0	-	-	-	-	-	-
	DENV-2	1045.22	34	0.00	0	-	-	-	-	-	-
	CHIKV	25.22	22.7	90.60	1111	4.55	39.68	ECSA	85.1	1175	1042
	CHIKV	270.22	25.8	0.00	11	85.14	0.39	-	-	-	-
	CHIKV	571.22	22.5	82.60	1446	4.37	51.64	ECSA	86.1	1657.3	1047
	CHIKV	612.22	26.2	65.80	15233	2.93	544.04	ECSA	82	2325	1153
	CHIKV	726.22	23.4	108.00	11347	0.73	405.25	ECSA	83.4	15885.8	1085
	CHIKV	736.22	34	0.00	12	79.31	0.43	-	-	-	-
5	CHIKV	1754.22	24.4	116.00	12778	0.67	456.36	ECSA	88.1	17090.3	1046
	CHIKV	**768.22_C**	28.5	83.00	12103	0.22	432.25	ECSA	86.4	10177.5	1056
	DENV-1	**768.22_C**	27.3	83.00	18893	0.22	674.75	Genotype V	99	1677.7	1083
	YFV	187.22	16	78.80	41143	0.15	1469.39	South America I	100	27291	1096
	YFV	194.22	17.3	73.00	43711	0.08	1561.11	South America I	100	31319.8	1044
	YFV	198.22	15.7	85.00	33173	0.19	1184.75	South America I	100	22019.7	1041
	YFV	202.22	16.3	96.80	56720	0.11	2025.71	South America I	100	34273.5	1091
	YFV	204.22	20.8	39.00	66869	0.10	2388.18	South America I	100	44428.1	979
	YFV	209.22	17.6	104.00	26677	0.16	952.75	South America I	100	17354.4	1116

Library	Virus	Sample	Ct	Concentration [ng/µL][1]	Reads on target	off-target reads (%)	reads/NTC ratio	Genotyping	Bootstrap[2]	Depth[3]	Sequence length
	YFV	713.22	16.8	72.80	26232	0.19	936.86	South America I	100	20264.3	1011
	NTC	NTC	-	0.00	28	-	-	-	-	-	-

Samples of codetection cases are highlighted in bold.

ND=Not Detected

NA= Not Available

NTC= No Template Control

[1]Sample concentration after PCR measured by Qubit.

[2]Sequence bootstrap support values obtained by virus genotyping tool from Genome Detective.

[3]Mean depth of coverage calculated by Genome Detective.

Table S5. Sequencing and genotyping results of clinical specimens tested during viral genomic surveillance activities in Brazil and Uruguay.

Location	Virus	Sample type	Sample ID	Collection date	Municipality	Ct	Number of reads on target	Mean depth of coverage	Nucleotide Identity (%)	Lineage
Acre - Brazil	DENV-2	Serum	ACNB23	2022-04-11	Xapuri	34.2	675	356.50	91.90	Genotype II - Cosmopolitan
Uruguay (Imported case from Mexico)	DENV-3	Serum	23-274	2023-04-29	Montevideo	25	6817	5493.80	93.10	Genotype III
Uruguay (Imported case from Paraguay)	DENV-4	Serum	16-64	2016-01-31	Montevideo	24	17916	6926.90	80.20	Genotype II
Uruguay (Imported case from Argentina)	DENV-2	Serum	23-255	2023-04-26	Montevideo	28	57800	12455.50	81.90	Genotype II - Cosmopolitan
Rio de Janeiro - Brazil	ZIKV	Serum	3563-16	2016	NA	IND	62	30.40	99.58	Asian
Amazonas - Brazil	DENV-2	Serum	1343	2022-08-08	Benjamin Constant	18	20991	728.08	92.37	Genotype II - Cosmopolitan
Amazonas - Brazil	DENV-2	Serum	1721	2022-09-10	Atalaia do Norte	21	14306	356.03	92.37	Genotype II - Cosmopolitan
Amazonas - Brazil	DENV-2	Serum	1440	2022-08-16	Benjamin Constant	27	5699	223.43	92.37	Genotype II - Cosmopolitan

Location	Virus	Sample type	Sample ID	Collection date	Municipality	Ct	Number of reads on target	Mean depth of coverage	Nucleotide Identity (%)	Lineage
Amazonas - Brazil	DENV-2	Serum	1970	2022-11-16	Atalaia do Norte	NA	13012	280.59	92.27	Genotype II - Cosmopolitan
Amazonas - Brazil	DENV-2	Serum	1711	NA	NA	NA	20968	471.42	92.97	Genotype II - Cosmopolitan
Amazonas - Brazil	DENV-2	Serum	1717	2022-07-10	Atalaia do Norte	22	14441	335.44	92.37	Genotype II - Cosmopolitan
Amazonas - Brazil	DENV-2	Serum	1724	2022-11-10	Atalaia do Norte	22	20661	522.95	92.62	Genotype II - Cosmopolitan
Amazonas - Brazil	DENV-2	Serum	1631	NA	NA	NA	6666	146.62	92.66	Genotype II - Cosmopolitan
Amazonas - Brazil	DENV-2	Serum	1851	2022-10-24	Benjamin Constant	NA	2755	56.70	92.37	Genotype II - Cosmopolitan
Amazonas - Brazil	DENV-2	Serum	1444	2022-08-26	Benjamin Constant	26	2174	85.82	92.37	Genotype II - Cosmopolitan
Amazonas - Brazil	DENV-2	Serum	1730	2022-11-10	Atalaia do Norte	24	20421	468.81	92.26	Genotype II - Cosmopolitan
Amazonas - Brazil	DENV-2	Serum	1709	NA	NA	NA	9248	204.79	92.47	Genotype II - Cosmopolitan
Amazonas - Brazil	DENV-2	Serum	1732	2022-10-10	Atalaia do Norte	24	10989	4059.90	92.50	Genotype II - Cosmopolitan
Amazonas - Brazil	DENV-2	Serum	1445	NA	NA	NA	3490	1256.40	92.20	Genotype II - Cosmopolitan
Amazonas - Brazil	DENV-2	Serum	1659	NA	NA	NA	48	13.30	82.30	Genotype II - Cosmopolitan

ND=Not Detected

NA= Not Available

NTC= No Template Control

Mean depth of coverage calculated by Genome Detective.

Table S6. List of GenBank access number of the sequences generated from clinical samples in this study.

Sequence ID	Access Number
CHIKV_633_18	OR578962
CHIKV_553_21	OR578960
CHIKV_1941	OR578968
CHIKV_1839	OR578967
CHIKV_RioGrandedoNorte_933_202	OR578969
CHIKV_25_22	OR578959
CHIKV_571_22	OR578963
CHIKV_612_22	OR578961
CHIKV_726_22	OR578964
CHIKV_1754_22	OR578966
CHIKV_768_22_C	OR578965
YFV_713_22	OR578977
YFV_209_22	OR578976
YFV_204_22	OR578974
YFV_202_22	OR578975
YFV_198_22	OR578973
YFV_194_22	OR578972
YFV_187_22	OR578970
ZIKV_RJ_3563_2016	OR506342
DENV1_U2	OR500978
DENV1_857_22	OR500979
DENV1_853_22	OR500980
DENV1_1053_22	OR500981
DENV1_1012_22	OR500982
DENV1_1010_22	OR500983
DENV1_519_22	OR500984

Sequence ID	Access Number
DENV1_514_22	OR500985
DENV1_303_22	OR500986
DENV1_298_22	OR500987
DENV1_169_22	OR500988
DENV1_768_22_C	OR500989
DENV1_U53	OR500990
DENV1_U18	OR500991
DENV2_ACRE_ACNB23_2022	OR500992
DENV2_Amazonas_1659	OR500993
DENV2_Amazonas_1445	OR500994
DENV2_URUGUAY_Montevideo_23_255	OR500995
DENV2_Amazonas_1732_2022	OR500996
DENV2_Amazonas_1343_2022	OR500997
DENV2_Amazonas_1721_2022	OR500998
DENV2_Amazonas_1440_2022	OR500999
DENV2_Amazonas_1970_2022	OR501000
DENV2_Amazonas_1711	OR501001
DENV2_Amazonas_1717_2022	OR501002
DENV2_Amazonas_1724_2022	OR501003
DENV2_Amazonas_1631	OR501004
DENV2_Amazonas_1851_2022	OR501005
DENV2_Amazonas_1444_2022	OR501006
DENV2_Amazonas_1730_2022	OR501007
DENV2_Amazonas_1709	OR501008
DENV2_U85	OR501009
DENV2_U73	OR501010
DENV2_U67	OR501011

Sequence ID	Access Number
DENV2_U61	OR501012
DENV2_933_22	OR501013
DENV3_URUGUAY_Montevideo_23_274	OR501014
DENV4_URUGUAY_Montevideo_16_64	OR501020

5 DISCUSSÃO

A economia e a saúde da população mundial têm sido afetadas por doenças infecciosas virais cujos principais registros desde 1918 incluem doenças como por exemplo a “gripe espanhola”, febre do Nilo Ocidental, doença por vírus Ebola e HIV/AIDS (Excler *et al.*, 2021). Embora a infecção pelo DENV tenha afetado vários países dos trópicos e subtrópicos nos últimos 20 anos, a expansão dos vírus ZIKV e CHIKV, que causaram graves surtos entre 2014 e 2016 na América Latina, tem gerado preocupação aos países do hemisfério norte que até então não apresentavam registros de circulação desses vírus (Adelino *et al.*, 2021; Delisle *et al.*, 2015; Faria *et al.*, 2017; Giovanetti *et al.*, 2020, 2022; Grandadam *et al.*, 2011; Naveca *et al.*, 2019; Périssé *et al.*, 2020; World Health Organization, 2023a; Xavier *et al.*, 2019). A recente emergência e rápida disseminação internacional do SARS-CoV-2 resultou em mais de 770 milhões de casos da doença por coronavírus 2019 (COVID-19) em todo o mundo, além das disrupções no comércio, cadeia de mantimentos e circulação de pessoas, assim evidenciando o potencial de impacto social e econômico das doenças infeciosas (World Health Organization, 2023b).

No Brasil, as disrupções causadas pelo SARS-CoV-2 somam-se ao ônus de outras doenças infecciosas causadas por vírus endêmicos como DENV, ZIKV, CHIKV e YFV, além dos vírus respiratórios como o influenza, responsável por 11.724 casos e 1.480 óbitos por Síndrome Respiratória Aguda Grave em 2022 (Ministério da Saúde do Brasil, 2023b). Ainda outros arbovírus com potencial de disseminação representam uma ameaça real de circulação epidêmica no país. O vírus do Nilo Ocidental, que se expandiu rapidamente nas Américas, tem sido observado no Brasil a partir de evidências sorológicas da circulação desse vírus em equinos e aves das regiões do pantanal e Nordeste em 2009 e 2010 (Costa *et al.*, 2021; Melandri *et al.*, 2012; Pauvolid-Corrêa *et al.*, 2011; Silva *et al.*, 2013). Recentemente, genomas referentes a linhagem 1a do WNV foram recuperados em amostras de equinos com encefalite oriundas dos estados de Minas gerais, Piauí, Ceará, Bahia e Espírito Santo, além de amostras de mosquitos do gênero *Culex* spp. coletados no estado do Pará em 2017 (Costa *et al.*, 2021; Fritsch *et al.*, 2022; Martins *et al.*, 2019; Neto *et al.*, 2023). A detecção de outros arbovírus emergentes como o Oropouche e Mayaro, por sua vez, têm sido frequentemente relatada na região amazônica em pacientes que

apresentaram doença exantemática febril aguda, embora a detecção do vírus Oropouche tenha sido relatada também em amostras coletada na Bahia em 2016 e 2017 (Esposito and Fonseca, 2017; Fonseca *et al.*, 2020; Nascimento *et al.*, 2020; Vasconcelos *et al.*, 2011).

O combate às epidemias de doenças infecciosas causadas pelos vírus mencionados tem sido impulsionado pelas contribuições da epidemiologia genômica que combina dados do genoma de um patógeno com metadados a nível individual para rastrear e entender a transmissão de um doença através de populações em diferentes escalas temporais e espaciais (Hill *et al.*, 2021). O genoma de um patógeno contém informações acerca da sua história evolutiva que permite identificar a espécie e suas linhagens, assim como também determinar a origem de um surto e reconstruir a sua história de transmissão (Stockdale, Liu and Colijn, 2022). Por exemplo, neste trabalho foi apresentado uma revisão (ver Apêndice A) que descreve o uso de sequências de nucleotídeos para caracterizar a evolução das linhagens Asiática e Africana do ZIKV desde sua emergência no continente africano e para subsidiar discussões sobre a revisão da classificação viral, com a proposição de uma terceira linhagem do ZIKV. O uso de sequências genômicas associadas a localização e data de amostragem também permitiu caracterizar as origens da epidemia de Zika registrada nas Américas em 2015 e 2016 ao reconstruir as relações filogenéticas entre diversas sequências da América Latina e Caribe e estimar que a linhagem Asiática provavelmente emergiu no início do ano de 2014 na região Nordeste do Brasil (Faria *et al.*, 2017; Faria, Azevedo, *et al.*, 2016). Estima-se que o Nordeste atuou como fonte principal da disseminação continental da linhagem Asiática, caracterizada por múltiplos eventos de introdução do vírus em vários países como Colômbia, Porto Rico, Honduras, Republica Dominicana, Jamaica, Haiti e Estados Unidos da América (Faria *et al.*, 2017; Metsky *et al.*, 2017).

A geração de sequências genômicas também pode aprimorar a vigilância ao fornecer informações sobre diversidade genética de um vírus durante um surto, contribuindo com o desenvolvimento e atualização de vacinas e ensaios para o diagnóstico preciso (Pollett *et al.*, 2020). A identificação de linhagens e variantes virais emergentes possibilita que o monitoramento genômico atue como um sistema de alerta inicial para o aumento da gravidade de um surto devido a introdução de linhagens ou variantes mais virulentas ou com potencial de evasão imune (Hill *et al.*, 2023). O rastreamento da diversidade do SARS-CoV-2 ao redor do mundo, por

exemplo, mostrou-se essencial para a formulação de políticas de saúde pública para o controle e mitigação dos impactos ocasionados pela emergência de novas variantes de preocupação (Brito *et al.*, 2022). O sequenciamento da variante Omicron (B.1.1.529), que abriga 15 mutações no domínio de ligação ao receptor (RBD) da proteína Spike, forneceu, por exemplo, informações genéticas para a atualização de vacinas capazes de induzir uma reposta imune mais eficaz contra a infecção pelo SARS-CoV-2 (Chalkias *et al.*, 2022; Food and Drug Administration, 2023a; b; Muik *et al.*, 2022; Viana *et al.*, 2022; Winokur *et al.*, 2023).

Os impactos na saúde pública resultantes da emergência do SARS-CoV-2, causador da pandemia de COVID-19, tem impulsionado o desenvolvimento sem precedentes da capacidade de sequenciamento de patógenos virais no Brasil nos últimos anos (Leite *et al.*, 2022). Como consequência, por exemplo, o trabalho de diversos grupos de pesquisa e dos laboratórios centrais de saúde pública dos estados tem contribuído para a geração, sem precedentes, de mais de 237 mil sequências do SARS-CoV-2 disponíveis no banco de dados do GISAID Initiative (GISAID Initiative, 2023; Menezes *et al.*, 2022). Entre essas sequências estão aquelas geradas durante a vigilância genômica do SARS-CoV-2 no estado de Minas Gerais, conforme apresentado no Apêndice B deste trabalho, que revelou aspectos epidemiológicos preliminares sobre o início da epidemia de COVID-19 em expansão naquele estado, caracterizada por múltiplos eventos independentes de importação do vírus nos primeiros meses de 2020.

As possibilidades apresentadas pelo sequenciamento de genomas virais empregado como uma ferramenta para a saúde pública, tem levado pesquisadores a defender a integração das técnicas de sequenciamento a um sistema de vigilância de vírus contínua que possibilite uma rápida resposta para mitigação de surtos causados por vírus emergentes e reemergentes (Black *et al.*, 2020; Hill *et al.*, 2023). Diante desse contexto e objetivando fornecer ferramentas complementares na vigilância de vírus em diferentes laboratórios e em atividades de campo, este trabalho também apresenta os resultados do desenvolvimento de um protocolo para a detecção de múltiplas espécies virais em uma única reação de PCR. O protocolo Ampli-FlaCk utiliza os benefícios da tecnologia, que permite gerar longas sequências, presente no sequenciador portátil, MinION (tecnologia considerada o método do ano em 2023 (Marx, 2023)). A recuperação de longas sequências nucleotídicas (~1000 bases) contém mais sinal filogenético, resultando na detecção precisa das linhagens virais

em 83% das amostras clínicas testadas, compreendendo casos positivos de dengue, zika, chikungunya e febre amarela. Devido ao emprego de *primers* com bases degeneradas, o protocolo também tem o potencial de detectar outros ortoflavivírus além daqueles testados.

Os benefícios da tecnologia portátil de sequenciamento empregada no MinION têm sido utilizados para vigilância viral deste a emergência do ZIKV no Brasil, quando esse sequenciador foi empregado em um laboratório móvel que percorreu diversos estados da região Nordeste em 2016 (Faria *et al.*, 2017). Deste então, essa tecnologia tem contribuído para a caracterização genética de vários vírus endêmicos no país, como DENV, CHIKV e YFV (Adelino *et al.*, 2021; Faria, N R *et al.*, 2018; Giovanetti *et al.*, 2023; Naveca *et al.*, 2018; Xavier *et al.*, 2019). Nesse contexto, a reemergência e disseminação do vírus Chikungunya no Brasil foram caracterizadas com a ajuda de dados genômicos e epidemiológicos associados, conforme os resultados descritos em três publicações apresentadas neste trabalho.

Desde a emergência da linhagem ECSA no estado da Bahia, surto localizados de febre Chikungunya têm sido registrados em diferentes regiões do Brasil (Souza *et al.*, 2019). Um desses surtos aconteceu na cidade de Natal no estado do Rio Grande do Norte em 2019, quando mais de 13 mil casos de febre Chikungunya foram registrados naquele estado. Apesar do número amostral limitado, os resultados das análises das sequências do vírus circulante durante aquele surto contribuíram com observações iniciais sobre a reintrodução da linhagem ECSA no estado do Rio Grande do Norte a partir de eventos de dispersão provavelmente originados no estado do Rio de Janeiro, onde elevada incidência de CHIKV também foi registrada durante o período de 2018 e 2019 (Xavier *et al.*, 2019). Seguindo um padrão similar, os resultados descritos em outra publicação presente neste trabalho revelaram uma ligação evolutiva entres os vírus circulantes no estado do Mato do Grosso e na região Nordeste. Apesar dos mais de 14 mil casos de Chikungunya registrados durante o surto no Mato Grasso em 2018 (Ministério da Saúde do Brasil, 2023c), apenas 6 genomas virais estavam disponíveis para aquele estado. O sequenciamento de 24 genomas do CHIKV indicou que o surto de 2018 provavelmente originou-se a partir de um evento de dispersão do CHIKV circulante no Nordeste do Brasil e sugeriu que a região Centro-Oeste tem atuado como fonte de transmissão viral através das fronteiras, uma vez que as sequências do Mato Grosso eram proximamente

relacionadas às sequências coletadas no Paraguai também em 2018 (Gräf *et al.*, 2021).

Ambos os trabalhos evidenciaram a quantidade reduzida de dados genômicos disponíveis referentes a surtos localizados que apresentaram, no entanto, números elevados de casos de Chikungunya. Além disso, a escassa informação genética disponível possuía limitada representatividade tanto na escala espacial quanto temporal. Diante desse contexto, o estudo mais recente apresentado neste trabalho descreveu os resultados de um amplo projeto de vigilância realizado em 2021 e 2022, a fim de fornecer informações genômicas atualizadas para uma caracterização ampla, a nível nacional, da diversidade genética e da história da dispersão da linhagem ECSA do CHIKV no Brasil. Este trabalho colaborativo com os LACENs de 12 estados brasileiros permitiu a geração de 422 genomas do CHIKV, um número que contribuiu para um aumento expressivo do número de genomas disponíveis no banco de dados do GenBank, embora seja inferior a porcentagem sugerida de 0,5% para sequenciamento de casos confirmados em um surto a fim de aumentar a chance de detecção célere de linhagens com potencial epidêmico (Brito *et al.*, 2022).

Reconstruções filogenéticas utilizando os novos genomas coletados em quatro regiões geográficas do Brasil revelaram a emergência de dois subclados (clado I e clado II) da linhagem ECSA, com data de divergência estimada entre janeiro e fevereiro de 2018 e com origem estimada na região Sudeste. Diferenças identificadas no perfil de composição de sequências dos clados revelaram o clado I como sendo mais diverso, no entanto, tais diferenças podem mudar à medida que a linhagem ECSA se expande dentro do país. Além disso, essas diferenças de composição podem refletir as distintas redes de transmissão que provavelmente envolvem a intensa troca de vírus entre as regiões Sudeste e Nordeste. Análises filogeográficas realizadas neste trabalho revelaram que a região Nordeste continua atuando como fonte de transmissão do CHIKV para outros estados e sugeriram que o Brasil poderá contribuir para expansão da linhagem ECSA para outros países da América Latina através da dispersão internacional do vírus mediada pelo movimento de pessoas nas fronteiras.

A expansão internacional da linhagem ECSA se torna ainda mais preocupante quando se combina fatores como a presença de vetor competente, condições climáticas e sociais favoráveis em grandes centros urbanos e a emergência de mutações que criam condições que podem modificar a paisagem adaptativa do CHIKV (Bartholomeeusen *et al.*, 2023; Chen *et al.*, 2021; Costa-da-Silva *et al.*, 2017; Freitas

et al., 2021; Tsetsarkin *et al.*, 2014). A comparação entre as taxas de substituições sinônimas e não-sinônimas do gene do envelope viral indicou que os clados I e II estiveram sujeitos a seleção positiva. Embora mutações sinônimas e não-sinônimas tenham sido identificadas exclusivamente para cada clado, não foi identificado nas sequências brasileiras a presença da substituição E1-A226V responsável por aumentar a adaptação do vírus ao mosquito *A. albopictus* (Tsetsarkin and Weaver, 2011; Tsetsarkin *et al.*, 2007). De qualquer forma, o monitoramento e análise de substituições sinônimas são relevantes visto que tais alterações também tem o potencial de promover mudanças adaptativas que podem alterar a estrutura do RNA viral e resultar em expressão diferencial e adaptação ao hospedeiro (Chen *et al.*, 2013; Faure *et al.*, 2017; Morley *et al.*, 2018).

A análise das sequências dos clados I e II também revelou uma alta frequência de substituições do tipo transições que podem ser consequência da ação de repostas do sistema imune do hospedeiro mediadas por desaminases do tipo AID/APOBEC e ADAR (Soares-Schanoski *et al.*, 2019). A ação dessas desaminases através da introdução de mutações pode contribuir significativamente para diversificação genética e a adaptação evolutiva do CHIKV, assim impulsionando a contínua expansão da linhagem ECSA em um contexto favorável como o dos centros urbanos brasileiros (Piontkivska *et al.*, 2021; Simmonds and Ansari, 2021).

6 CONCLUSÃO

Neste trabalho, o uso da tecnologia portátil de sequenciamento por nanoporos permitiu gerar no total 454 genomas do CHIKV partir de amostras clínicas coletadas em 13 estados brasileiros nos anos de 2021 e 2022. Dessa forma foi possível contribuir para o aumento expressivo (aumento de 151,33%) do número total de sequências da linhagem ECSA circulante no Brasil depositadas no banco de dados do GenBank.

Análises filogenéticas dos novos genomas revelaram a reintrodução da linhagem ECSA do CHIKV no estado do Rio Grande do Norte, onde este vírus causou um surto em 2019. Essas análises também indicaram que o CHIKV se dispersou da região Nordeste para a região Centro-Oeste, onde foi registrado um surto de mais de 12 mil casos de febre Chikungunya apenas no estado do Mato Grosso em 2018. A proximidade filogenética das sequências do Mato Grosso com as sequências do Paraguai coletadas em 2018 e a análise filogeográfica da localização provável do ancestral comum mais recente dessas sequências sugeriram que a região Centro-Oeste pode estar atuando como fonte de dispersão do CHIKV para o Paraguai. Os dados genômicos do vírus circulante na epidemia de 2021-2022 possibilitaram a reconstrução da história evolutiva atualizada da linhagem ECSA no Brasil e indicaram que a região Nordeste continua atuando como principal fonte de dispersão do CHIKV no Brasil, alimentando uma frequente rede de intercâmbio de vírus entre as regiões Nordeste e Sudeste. Portanto, o estabelecimento da tecnologia de sequenciamento na vigilância de vírus epidêmicos tem contribuído para o monitorando oportuno e a caracterização extensiva da expansão de linhagem ECSA no Brasil.

Este trabalho também descreveu a emergência de dois subclados da linhagem ECSA do CHIKV circulante no Brasil recentemente. Embora os dois clados possuam uma data de emergência estimada para o mesmo período (janeiro-fevereiro de 2018), eles apresentaram diferenças no perfil mutacional e de composição de sequências. O clado I apresentou-se mais diverso com sequências obtidas em amostras coletadas em 2021 e 2022 em 14 estados brasileiros cobrindo a região Nordeste, Sul, Sudeste e Centro-Oeste. O clado II, por sua vez, continha sequências majoritariamente da região Nordeste coletadas em 2022.

A análise comparativa das sequências revelou que o clado II apresentou o maior número de variantes de nucleotídeo único (n=27) das quais três representaram

substituições não sinônimas, enquanto o clado I continha 13 variantes de nucleotídeo único, entre elas 6 substituições não-sinônimas identificadas em duas proteínas não-estruturais e três estruturais. Além disso, duas mutações (E1-T98A e E2-V264A) foram identificadas entre as sequências dos clados I e II. Tais mutações já haviam sido associadas ao aumento da adaptação do CHIKV a mosquitos do gênero *Aedes* spp. Ademais, uma alta frequência de transições do tipo citosina para timina foram identificadas entre as novas sequências dos dois clados. Argumenta-se que essas transições são resultado da ação de desaminases do grupo APOBEC e ADAR1 que atuam na regulação de mecanismos antivirais do hospedeiro e que podem contribuir para diversificação genética e a adaptação evolutiva da população viral.

Neste trabalho também foi discutido sobre a dificuldade observada no diagnóstico clínico das doenças causadas por arbovírus devido à similaridade dos sintomas e à existência de reação cruzada nos testes sorológicos para vírus pertencentes ao mesmo grupo taxonômico. Logo, tem sido argumentado sobre a necessidade de desenvolvimento de testes de diagnóstico mais sensíveis, específicos, rápidos e que permitam a análise de múltiplas amostras simultaneamente. Nesse sentido, o presente trabalho também desenvolveu um protocolo de detecção multiplex de arbovírus baseado na tecnologia de sequenciamento por nanoporos que permite a rápida geração de uma grande quantidade de dados de sequências nucleotídicas virais.

O protocolo Ampli-Flack integra a reação em cadeia da polimerase com transcrição reversa (RT-PCR) e um conjunto de *primers* selecionados para a amplificação simultânea do material genético do CHIKV e diferentes vírus pertencentes ao gênero *Orthoflavivirus* em uma única reação. A integração do protocolo com a tecnologia de sequenciamento por nanoporos permite o sequenciamento simultâneo de até 96 amostras em uma única reação. A praticidade e portabilidade do MinION empregado no protocolo Ampli-Flack possibilitaram a recuperação de longas sequências (~1000 bases) em até 6h que podem ser usadas para a identificação acurada de diferentes linhagens virais em uma única reação.

A validação do protocolo Ampli-Flack com mostras clínicas positivas para DENV-1, DENV-2, CHIKV, YFV e ZIKV resultou na identificação correta das linhagens virais em 83% das amostras testadas. Além disso, as sequências nucleotídicas geradas a partir das amostras clínicas apresentaram sinal filogenético suficiente para a reconstrução acurada da filogenia para a genotipagem dos vírus mencionados.

Portanto os resultados apresentados indicam que o protocolo Ampli-Flack é capaz de recuperar sequências para a identificação acurada das linhagens de múltiplas espécies de arbovírus e a sua praticidade aliada a portabilidade do sequenciador MinION possibilita que esse protocolo seja utilizado em atividades de campo em locais com infraestrutura limitada.

REFERÊNCIAS BIBLIOGRÁFICAS

ACEVEDO, N. *et al.* Zika Virus, Chikungunya Virus, and Dengue Virus in Cerebrospinal Fluid from Adults with Neurological Manifestations, Guayaquil, Ecuador. **Frontiers in microbiology**, v. 8, p. 42, 24 Jan. 2017.

ADELINO, T. É. R. *et al.* Field and classroom initiatives for portable sequence-based monitoring of dengue virus in Brazil. **Nature Communications**, v. 12, n. 1, p. 2296, 16 Apr. 2021.

ANDRIOLI, D. C.; BUSATO, M. A.; LUTINSKI, J. A. Spatial and temporal distribution of dengue in Brazil, 1990 - 2017. **Plos One**, v. 15, n. 2, p. e0228346, 13 Feb. 2020.

ARTSOB, H.; LINDSAY, R.; DREBOT, M. Arboviruses. *In*: **International encyclopedia of public health**. [s.l.] Elsevier, 2017. p. 154–160.

AZEVEDO COSTA, E. *et al.* West Nile virus detection in horses in three Brazilian states. **BioRxiv**, 6 Jan. 2021.

BALLOUX, F. *et al.* From theory to practice: translating whole-genome sequencing (WGS) into the clinic. **Trends in Microbiology**, v. 26, n. 12, p. 1035–1048, Dec. 2018.

BARTHOLOMEEUSEN, K. *et al.* Chikungunya fever. **Nature reviews. Disease primers**, v. 9, n. 1, p. 17, 6 Apr. 2023.

BHARADWAJ, M. *et al.* Diagnosing point-of-care diagnostics for neglected tropical diseases. **PLoS Neglected Tropical Diseases**, v. 15, n. 6, p. e0009405, 17 Jun. 2021.

BLACK, A. *et al.* Ten recommendations for supporting open pathogen genomic analysis in public health. **Nature Medicine**, v. 26, n. 6, p. 832–841, 11 Jun. 2020.

BRASIL, P. *et al.* Zika virus infection in pregnant women in rio de janeiro. **The New England Journal of Medicine**, v. 375, n. 24, p. 2321–2334, 15 Dec. 2016.

BRITO, A. F. *et al.* Global disparities in SARS-CoV-2 genomic surveillance. **Nature Communications**, v. 13, n. 1, p. 7003, 16 Nov. 2022.

C. ELEGANS SEQUENCING CONSORTIUM. Genome sequence of the nematode C. elegans: a platform for investigating biology. **Science**, v. 282, n. 5396, p. 2012–2018, 11 Dec. 1998.

CANDIDO, D. S. *et al.* Evolution and epidemic spread of SARS-CoV-2 in Brazil. **Science**, v. 369, n. 6508, p. 1255–1260, 4 Sep. 2020.

CAUCHEMEZ, S. *et al.* Association between Zika virus and microcephaly in French Polynesia, 2013-15: a retrospective study. **The Lancet**, v. 387, n. 10033, p. 2125–2132, 21 May 2016.

CAUSEY, O. R.; MAROJA, O. M. Mayaro virus: a new human disease agent. III. Investigation of an epidemic of acute febrile illness on the river Guama in Pará, Brazil, and isolation of Mayaro virus as causative agent. **The American Journal of Tropical Medicine and Hygiene**, v. 6, n. 6, p. 1017–1023, Nov. 1957.

CHALKIAS, S. *et al.* A Bivalent Omicron-Containing Booster Vaccine against Covid-19. **The New England Journal of Medicine**, v. 387, n. 14, p. 1279–1291, 6 Oct. 2022.

CHARLYS DA COSTA, A. *et al.* Spread of chikungunya virus east/central/south african genotype in northeast brazil. **Emerging Infectious Diseases**, v. 23, n. 10, p. 1742–1744, Oct. 2017.

CHEN, R. *et al.* Chikungunya virus 3' untranslated region: adaptation to mosquitoes and a population bottleneck as major evolutionary forces. **PLoS Pathogens**, v. 9, n. 8, p. e1003591, 29 Aug. 2013.

___. ICTV virus taxonomy profile: togaviridae. **The Journal of General Virology**, v. 99, n. 6, p. 761–762, Jun. 2018.

___. Lineage Divergence and Vector-Specific Adaptation Have Driven Chikungunya Virus onto Multiple Adaptive Landscapes. **mBio**, v. 12, n. 6, p. e0273821, 21 Dec. 2021.

CLARO, I. M. *et al.* Shotgun metagenomic sequencing of the first case of monkeypox virus in Brazil, 2022. **Revista Do Instituto de Medicina Tropical de Sao Paulo**, v. 64, p. e48, 24 Jun. 2022.

COSTA-DA-SILVA, A. L. *et al.* First report of naturally infected Aedes aegypti with chikungunya virus genotype ECSA in the Americas. **PLoS Neglected Tropical Diseases**, v. 11, n. 6, p. e0005630, 14 Jun. 2017.

COSTA, É. A. *et al.* West nile virus in brazil. **Pathogens (Basel, Switzerland)**, v. 10, n. 7, 15 Jul. 2021.

CUNHA, M. D. P. *et al.* Outbreak of chikungunya virus in a vulnerable population of Sergipe, Brazil-A molecular and serological survey. **Journal of Clinical Virology**, v. 97, p. 44–49, Dec. 2017.

CUNHA, M. S. *et al.* Autochthonous transmission of east/central/south african genotype chikungunya virus, brazil. **Emerging Infectious Diseases**, v. 23, n. 10, p. 1737–1739, Oct. 2017.

DELISLE, E. *et al.* Chikungunya outbreak in Montpellier, France, September to October 2014. **Euro Surveillance**, v. 20, n. 17, 30 Apr. 2015.

DONALISIO, M. R.; FREITAS, A. R. R.; ZUBEN, A. Arboviruses emerging in Brazil: challenges for clinic and implications for public health. **Revista de Saúde Pública**, v. 51, n. n.30, p. 1–6, 10 Apr. 2017.

EDRIDGE, A. W. D.; HOEK, L. VAN DER. Emerging orthobunyaviruses associated with CNS disease. **PLoS Neglected Tropical Diseases**, v. 14, n. 10, p. e0008856, 28 Oct. 2020.

ELLIOTT, R. M. Orthobunyaviruses: recent genetic and structural insights. **Nature Reviews. Microbiology**, v. 12, n. 10, p. 673–685, Oct. 2014.

ESPOSITO, D. L. A.; FONSECA, B. A. L. DA. Will Mayaro virus be responsible for the next outbreak of an arthropod-borne virus in Brazil? **The Brazilian Journal of Infectious Diseases**, v. 21, n. 5, p. 540–544, 7 Jul. 2017.

EXCLER, J.-L. *et al.* Vaccine development for emerging infectious diseases. **Nature Medicine**, v. 27, n. 4, p. 591–600, 12 Apr. 2021.

FANTINATO, F. F. S. T. *et al.* Description of the first cases of Zika virus fever investigated in municipalities of the Brazilian Northeastern Region, 2015. **Epidemiologia e servicos de saude : revista do Sistema Unico de Saude do Brasil**, v. 25, n. 4, p. 683–690, 26 Sep. 2016.

FARIA, N. R.; AZEVEDO, R. DO S. DA S.; *et al.* Zika virus in the Americas: Early epidemiological and genetic findings. **Science**, v. 352, n. 6283, p. 345–349, 15 Apr. 2016.

FARIA, N. R.; SABINO, E. C.; *et al.* Mobile real-time surveillance of Zika virus in Brazil. **Genome Medicine**, v. 8, n. 1, p. 97, 29 Sep. 2016.

FARIA, N. R. *et al.* Establishment and cryptic transmission of Zika virus in Brazil and the Americas. **Nature**, v. 546, n. 7658, p. 406–410, 15 Jun. 2017.

FARIA, NUNO RODRIGUES *et al.* Genomic and epidemiological monitoring of yellow fever virus transmission potential. **BioRxiv**, 16 Apr. 2018.

FARIA, N R *et al.* Genomic and epidemiological monitoring of yellow fever virus transmission potential. **Science**, v. 361, n. 6405, p. 894–899, 31 Aug. 2018.

FAURE, G. *et al.* Adaptation of mRNA structure to control protein folding. **RNA Biology**, v. 14, n. 12, p. 1649–1654, 2 Dec. 2017.

FONSECA, L. M. D. S. *et al.* Oropouche virus detection in febrile patients' saliva and urine samples in salvador, bahia, brazil. **Japanese Journal of Infectious Diseases**, v. 73, n. 2, p. 164–165, 24 Mar. 2020.

FOOD AND DRUG ADMINISTRATION. **Moderna COVID-19 Vaccine (2023-2024 Formula) Authorized For Individuals 6 Months through 11 Years of Age**. Disponível em: <https://www.fda.gov/vaccines-blood-biologics/coronavirus-covid-19-cber-regulated-biologics/moderna-covid-19-vaccine>. Acesso em: 8 oct. 2023a.

___. **Pfizer-BioNTech COVID-19 Vaccine (2023-2024 Formula) Authorized For Individuals 6 Months through 11 Years of Age**. Disponível em: <https://www.fda.gov/vaccines-blood-biologics/coronavirus-covid-19-cber-regulated-biologics/pfizer-biontech-covid-19-vaccine>. Acesso em: 8 oct. 2023b.

FREITAS, L. P. *et al.* Spatio-temporal modelling of the first Chikungunya epidemic in an intra-urban setting: The role of socioeconomic status, environment and temperature. **PLoS Neglected Tropical Diseases**, v. 15, n. 6, p. e0009537, 18 Jun. 2021.

FRITSCH, H. *et al.* Retrospective Investigation in Horses with Encephalitis Reveals Unnoticed Circulation of West Nile Virus in Brazil. **Viruses**, v. 14, n. 7, 14 Jul. 2022.

GARDY, J. L.; LOMAN, N. J. Towards a genomics-informed, real-time, global pathogen surveillance system. **Nature Reviews. Genetics**, v. 19, n. 1, p. 9–20, Jan. 2018.

GIOVANETTI, M. *et al.* Yellow Fever Virus Reemergence and Spread in Southeast Brazil, 2016-2019. **Journal of Virology**, v. 94, n. 1, 12 Dec. 2019.

___. Genomic and epidemiological surveillance of zika virus in the amazon region. **Cell reports**, v. 30, n. 7, p. 2275- 2283.e7, 18 Feb. 2020.

___. A retrospective overview of zika virus evolution in the midwest of brazil. **Microbiology spectrum**, v. 10, n. 2, p. e0015522, 27 Apr. 2022.

___. Genomic epidemiology unveils the dynamics and spatial corridor behind the Yellow Fever virus outbreak in Southern Brazil. **Science Advances**, v. 9, n. 35, p. eadg9204, 1 Sep. 2023.

GIRE, S. K. *et al.* Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. **Science**, v. 345, n. 6202, p. 1369–1372, 12 Sep. 2014.

GISAID INITIATIVE. **GISAID Initiative**. Disponível em: <https://gisaid.org/>. Acesso em: 30 sep. 2023.

GOODWIN, S.; MCPHERSON, J. D.; MCCOMBIE, W. R. Coming of age: ten years of next-generation sequencing technologies. **Nature Reviews. Genetics**, v. 17, n. 6, p. 333–351, 17 May 2016.

GOULD, E. *et al.* Emerging arboviruses: Why today? **One health**, v. 4, p. 1–13, Dec. 2017.

GOULD, E. A. *et al.* Origins, evolution, and vector/host coadaptations within the Genus Flavivirus. *In*: Advances in virus research. [s.l.] Elsevier, 2003. v. 59p. 277–314.

GRÄF, T. *et al.* Epidemiologic history and genetic diversity origins of chikungunya and dengue viruses, paraguay. **Emerging Infectious Diseases**, v. 27, n. 5, p. 1393–1404, May 2021.

GRANDADAM, M. *et al.* Chikungunya virus, southeastern France. **Emerging Infectious Diseases**, v. 17, n. 5, p. 910–913, May 2011.

HILL, V. *et al.* Progress and challenges in virus genomic epidemiology. **Trends in Parasitology**, v. 37, n. 12, p. 1038–1049, Dec. 2021.

___. Toward a global virus genomic surveillance network. **Cell Host & Microbe**, v. 31, n. 6, p. 861–873, 14 Jun. 2023.

HOULDCROFT, C. J.; BEALE, M. A.; BREUER, J. Clinical and biological insights from viral genome sequencing. **Nature Reviews. Microbiology**, v. 15, n. 3, p. 183–192, Mar. 2017.

HUGHES, H. R. *et al.* ICTV virus taxonomy profile: peribunyaviridae. **The Journal of General Virology**, v. 101, n. 1, p. 1–2, Jan. 2020.

HUNSPERGER, E. Arboviruses. *In*: JORGENSEN, J. H. *et al.* (Eds.). . ***Manual of* Clinical Microbiology**. Washington, DC, USA: ASM Press, 2015. p. 1644–1659.

HU, T. *et al.* Next-generation sequencing technologies: An overview. **Human Immunology**, v. 82, n. 11, p. 801–811, Nov. 2021.

IMRIE, A. *et al.* Antibody to dengue 1 detected more than 60 years after infection. **Viral Immunology**, v. 20, n. 4, p. 672–675, Dec. 2007.

INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM *et al.* Initial sequencing and analysis of the human genome. **Nature**, v. 409, n. 6822, p. 860–921, 15 Feb. 2001.

JESUS, J. G. DE *et al.* Importation and early local transmission of COVID-19 in Brazil, 2020. **Revista Do Instituto de Medicina Tropical de Sao Paulo**, v. 62, p. e30, 11 May 2020.

JONES, R. *et al.* Arbovirus vectors of epidemiological concern in the Americas: A scoping review of entomological studies on Zika, dengue and chikungunya virus vectors. **Plos One**, v. 15, n. 2, p. e0220753, 6 Feb. 2020.

KERKHOF, K. *et al.* Reliable serological diagnostic tests for arboviruses: feasible or utopia? **Trends in Microbiology**, v. 28, n. 4, p. 276–292, Apr. 2020.

LANCIOTTI, R. S.; ROEHRIG, J. T. Arboviruses. *In*: DETRICK, B.; SCHMITZ, J. L.; HAMILTON, R. G. (Eds.). . **Manual of molecular and clinical laboratory immunology**. Washington, DC, USA: ASM Press, 2016. p. 648–657.

LEITE, J. A. *et al.* Implementation of a COVID-19 Genomic Surveillance Regional Network for Latin America and Caribbean region. **Plos One**, v. 17, n. 3, p. e0252526, 3 Mar. 2022.

LESSA-AQUINO, C. *et al.* Detection of East/Central/South African genotype Chikungunya virus during an outbreak in a southeastern state of Brazil. **Epidemiology and Infection**, v. 146, n. 16, p. 2056–2058, Dec. 2018.

LEWANDOWSKI, K. *et al.* Metagenomic Nanopore Sequencing of Influenza Virus Direct from Clinical Respiratory Samples. **Journal of Clinical Microbiology**, v. 58, n. 1, 23 Dec. 2019.

LWANDE, O. W. *et al.* Globe-trotting *Aedes aegypt*i and *Aedes albopictus*: risk factors for arbovirus pandemics. **Vector Borne and Zoonotic Diseases**, v. 20, n. 2, p. 71–81, Feb. 2020.

MAGALHAES, T. *et al.* The endless challenges of arboviral diseases in brazil. **Tropical medicine and infectious disease**, v. 5, n. 2, 9 May 2020.

MARDIS, E. R. DNA sequencing technologies: 2006-2016. **Nature Protocols**, v. 12, n. 2, p. 213–218, Feb. 2017.

MARTINS, L. C. *et al.* First isolation of West Nile virus in Brazil. **Memorias do Instituto Oswaldo Cruz**, v. 114, p. e180332, 17 Jan. 2019.

MARX, V. Method of the year: long-read sequencing. **Nature Methods**, v. 20, n. 1, p. 6–11, Jan. 2023.

MARZIALI, A.; AKESON, M. New DNA sequencing methods. **Annual Review of Biomedical Engineering**, v. 3, p. 195–223, 2001.

MAYER, S. V.; TESH, R. B.; VASILAKIS, N. The emergence of arthropod-borne viral diseases: A global prospective on dengue, chikungunya and zika fevers. **Acta Tropica**, v. 166, p. 155–163, Feb. 2017.

MEHTA, R. *et al.* The spectrum of neurological disease associated with Zika and chikungunya viruses in adults in Rio de Janeiro, Brazil: A case series. **PLoS Neglected Tropical Diseases**, v. 12, n. 2, p. e0006212, 12 Feb. 2018.

MELANDRI, V. *et al.* Serological detection of West Nile virus in horses and chicken from Pantanal, Brazil. **Memorias do Instituto Oswaldo Cruz**, v. 107, n. 8, p. 1073–1075, Dec. 2012.

MENEZES, D. *et al.* SARS-CoV-2 Genomic Surveillance in Brazil: A Systematic Review with Scientometric Analysis. **Viruses**, v. 14, n. 12, 5 Dec. 2022.

METSKY, H. C. *et al.* Zika virus evolution and spread in the Americas. **Nature**, v. 546, n. 7658, p. 411–415, 15 Jun. 2017.

MIAO, Q. *et al.* Microbiological Diagnostic Performance of Metagenomic Next-generation Sequencing When Applied to Clinical Practice. **Clinical Infectious Diseases**, v. 67, n. suppl_2, p. S231–S240, 13 Nov. 2018.

MINISTÉRIO DA SAÚDE DO BRASIL. **Monitoramento de casos de febre do Mayaro e febre do Oropouche até a Semana Epidemiológica 35, 2019: Experiência do Instituto Evandro Chagas (IEC/SVS/MS)**: Boletim Epidemiológico. Brasilia: Secretaria de Vigilância em Saúde, Oct. 2019.

___. **Situação epidemiológica da febre amarela no monitoramento 2019/2020**. Brasilia: Secretaria de Vigilância em Saúde, 16 Jan. 2020. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/edicoes/2020/>. Acesso em: 27 sep. 2023.

___. **Situação epidemiológica da síndrome congênita associada à infecção pelo vírus Zika: Brasil, 2015 a 2022**. Brasilia: Secretaria de Vigilância em Saúde e Ambiente, 3 Apr. 2023a. Disponível em: <http://plataforma.saude.gov.br/anomalias-congenitas/>. Acesso em: 27 sep. 2023.

___. **Boletim Epidemiológico Especial: Doença pelo Coronavírus – Covid-19**. Brasilia: Secretaria de Vigilância em Saúde e Ambiente, 29 Jun. 2023b. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/covid-19/2023/boletim-epidemiologico-no-151-boletim-coe-coronavirus/view>. Acesso em: 6 oct. 2023.

___. **TabNet Win32 3.2: Febre de Chikungunya - Notificações registradas no Sistema de Informação de Agravos de Notificação - Mato Grosso**. Disponível em: <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sinannet/cnv/chikunmt.def>. Acesso em: 9 oct. 2023c.

MORATORIO, G.; VIGNUZZI, M. Monitoring and redirecting virus evolution. **PLoS Pathogens**, v. 14, n. 6, p. e1006979, 7 Jun. 2018.

MORLEY, V. J. *et al.* Chikungunya virus evolution following a large 3'UTR deletion results in host-specific molecular changes in protein-coding regions. **Virus evolution**, v. 4, n. 1, p. vey012, Jan. 2018.

MUIK, A. *et al.* Exposure to BA.4/5 S protein drives neutralization of Omicron BA.1, BA.2, BA.2.12.1, and BA.4/5 in vaccine-experienced humans and mice. **Science Immunology**, v. 7, n. 78, p. eade9888, 23 Dec. 2022.

NASCIMENTO, V. A. DO *et al.* Oropouche virus detection in saliva and urine. **Memorias do Instituto Oswaldo Cruz**, v. 115, p. e190338, 27 Feb. 2020.

NAVECA, F. G. *et al.* Chikungunya virus outbreak in the Amazon region: replacement of the Asian genotype by an ECSA lineage. **BioRxiv**, 10 Dec. 2018.

___. Genomic, epidemiological and digital surveillance of Chikungunya virus in the Brazilian Amazon. **PLoS Neglected Tropical Diseases**, v. 13, n. 3, p. e0007065, 7 Mar. 2019.

NETO, J. P. N. *et al.* First isolation and genome sequence analysis of west nile virus in mosquitoes in brazil. **Tropical medicine and infectious disease**, v. 8, n. 4, 20 Apr. 2023.

NOGRADY, B. How cancer genomics is transforming diagnosis and treatment. **Nature**, v. 579, n. 7800, p. S10–S11, Mar. 2020.

NOGUEIRA, R. M. *et al.* Dengue epidemic in the stage of Rio de Janeiro, Brazil, 1990-1: co-circulation of dengue 1 and dengue 2 serotypes. **Epidemiology and Infection**, v. 111, n. 1, p. 163–170, Aug. 1993.

NUNES, M. R. T. *et al.* Emergence and potential for spread of Chikungunya virus in Brazil. **BMC Medicine**, v. 13, p. 102, 30 Apr. 2015.

OHST, C. *et al.* Reliable serological testing for the diagnosis of emerging infectious diseases. **Advances in Experimental Medicine and Biology**, v. 1062, p. 19–43, 2018.

OSANAI ET AL., C. H. **Surto de dengue em Boa Vista, Roraima (nota prévia)**. [s.l.] Revista do Instituto de Medicina Tropical de São Paulo, Feb. 1983. Disponível em: <https://patua.iec.gov.br/handle/iec/2804>. Acesso em: 3 oct. 2023.

PAN AMERICAN HEALTH ORGANIZATION. **Epidemiological Update: Yellow Fever (28 December 2021)**. Washington, D.C: PAHO, 28 Dec. 2021. Disponível em: <https://iris.paho.org/handle/10665.2/55653>. Acesso em: 27 sep. 2023.

___. **PAHO/WHO Data - National Dengue fever cases**. Disponível em: <https://www3.paho.org/data/index.php/en/mnu-topics/indicadores-dengue-en/dengue-nacional-en/252-dengue-pais-ano-en.html>. Acesso em: 26 sep. 2023a.

___. **PAHO/WHO Data - Zika - Weekly report**. Disponível em: <https://www3.paho.org/data/index.php/en/mnu-topics/zika-weekly-en/>. Acesso em: 27 sep. 2023b.

___. **Cases of chikungunya virus disease**. Disponível em: <https://www3.paho.org/data/index.php/en/mnu-topics/chikv-en/550-chikv-weekly-en.html>. Acesso em: 21 feb. 2023c.

___. **Nota Informativa: Doença por Febre do Nilo Ocidental Brasil - 20 de setembro de 2023 - OPAS/OMS**. Disponível em: <https://www.paho.org/pt/documentos/nota-informativa-doenca-por-febre-do-nilo-ocidental-brasil-20-setembro-2023>. Acesso em: 28 sep. 2023d.

___. **Annual Arbovirus Bulletin 2022**. [s.l.] Regional Office for the Americas of the World Health Organization, 26 Sep. 2023e. Disponível em:

<https://www3.paho.org/data/index.php/en/mnu-topics/indicadores-dengue-en/annual-arbovirus-bulletin-2022.html>. Acesso em: 27 sep. 2023.

PAUVOLID-CORRÊA, A. *et al.* Neutralising antibodies for West Nile virus in horses from Brazilian Pantanal. **Memorias do Instituto Oswaldo Cruz**, v. 106, n. 4, p. 467–474, Jun. 2011.

PAYNE, A. *et al.* BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. **Bioinformatics**, v. 35, n. 13, p. 2193–2198, 1 Jul. 2019.

PÉRISSÉ, A. R. S. *et al.* Zika, dengue and chikungunya population prevalence in Rio de Janeiro city, Brazil, and the importance of seroprevalence studies to estimate the real number of infected individuals. **Plos One**, v. 15, n. 12, p. e0243239, 17 Dec. 2020.

PIANTADOSI, A.; KANJILAL, S. Diagnostic approach for arboviral infections in the united states. **Journal of Clinical Microbiology**, v. 58, n. 12, 18 Nov. 2020.

PINHEIRO ET AL., F. F. Epidemia de vírus Oropouche em Belém - Nota prévia (Publicado originalmente em 1962). *In*: **Memórias do Instituto Evandro Chagas**. Belém: Instituto Evandro Chagas, 2002. v. 7.

PINHEIRO, F. P. *et al.* An outbreak of Mayaro virus disease in Belterra, Brazil. I. Clinical and virological findings. **The American Journal of Tropical Medicine and Hygiene**, v. 30, n. 3, p. 674–681, May 1981.

PIONTKIVSKA, H. *et al.* ADAR Editing in Viruses: An Evolutionary Force to Reckon with. **Genome Biology and Evolution**, v. 13, n. 11, 5 Nov. 2021.

POLLETT, S. *et al.* Genomic Epidemiology as a Public Health Tool to Combat Mosquito-Borne Virus Outbreaks. **The Journal of Infectious Diseases**, v. 221, n. Suppl 3, p. S308–S318, 28 Mar. 2020.

POSTLER, T. S. *et al.* Renaming of the genus Flavivirus to Orthoflavivirus and extension of binomial species names within the family Flaviviridae. **Archives of Virology**, v. 168, n. 9, p. 224, 10 Aug. 2023.

POWERS, A. M. *et al.* Re-emergence of Chikungunya and O'nyong-nyong viruses: evidence for distinct geographical lineages and distant evolutionary relationships. **The Journal of General Virology**, v. 81, n. Pt 2, p. 471–479, Feb. 2000.

___. Evolutionary relationships and systematics of the alphaviruses. **Journal of Virology**, v. 75, n. 21, p. 10118–10131, Nov. 2001.

POWERS, A. M. Genomic evolution and phenotypic distinctions of Chikungunya viruses causing the Indian Ocean outbreak. **Experimental Biology and Medicine**, v. 236, n. 8, p. 909–914, Aug. 2011.

PYBUS, O. G.; TATEM, A. J.; LEMEY, P. Virus evolution and transmission in an ever more connected world. **Proceedings. Biological Sciences / the Royal Society**, v. 282, n. 1821, p. 20142878, 22 Dec. 2015.

QUICK, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. **Nature**, v. 530, n. 7589, p. 228–232, 11 Feb. 2016.

RICO-HESSE ET AL.,. **ICTV Virus Taxonomy Profile: Family: Flaviviridae**. Disponível em: <https://ictv.global/report/chapter/flaviviridae/flaviviridae>. Acesso em: 24 sep. 2023.

RODRIGUES, N. C. P. *et al.* Temporal and Spatial Evolution of Dengue Incidence in Brazil, 2001-2012. **Plos One**, v. 11, n. 11, p. e0165945, 10 Nov. 2016.

SAHLIN, K.; MEDVEDEV, P. Error correction enables use of Oxford Nanopore technology for reference-free transcriptome analysis. **Nature Communications**, v. 12, n. 1, p. 2, 4 Jan. 2021.

SANGER, F.; NICKLEN, S.; COULSON, A. R. DNA sequencing with chain-terminating inhibitors. **Proceedings of the National Academy of Sciences of the United States of America**, v. 74, n. 12, p. 5463–5467, Dec. 1977.

SCHATZMAYR, H. G.; NOGUEIRA, R. M.; TRAVASSOS DA ROSA, A. P. An outbreak of dengue virus at Rio de Janeiro--1986. **Memorias do Instituto Oswaldo Cruz**, v. 81, n. 2, p. 245–246, 1986.

SCHUFFENECKER, I. *et al.* Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. **PLoS Medicine**, v. 3, n. 7, p. e263, Jul. 2006.

SEREIKA, M. *et al.* Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. **Nature Methods**, v. 19, n. 7, p. 823–826, 4 Jul. 2022.

SHENDURE, J. *et al.* DNA sequencing at 40: past, present and future. **Nature**, v. 550, n. 7676, p. 345–353, 19 Oct. 2017.

SILVA, J. R. *et al.* Serologic survey of West Nile virus in horses from Central-West, Northeast and Southeast Brazil. **Memórias do Instituto Oswaldo Cruz**, v. 108, n. 7, p. 921–923, Nov. 2013.

SILVA, N. I. O. *et al.* Recent sylvatic yellow fever virus transmission in Brazil: the news from an old disease. **Virology Journal**, v. 17, n. 1, p. 9, 23 Jan. 2020.

SIMMONDS, P.; ANSARI, M. A. Extensive C->U transition biases in the genomes of a wide range of mammalian RNA viruses; potential associations with transcriptional mutations, damage- or host-mediated editing of viral RNA. **PLoS Pathogens**, v. 17, n. 6, p. e1009596, 1 Jun. 2021.

SOARES-SCHANOSKI, A. *et al.* Systems analysis of subjects acutely infected with the Chikungunya virus. **PLoS Pathogens**, v. 15, n. 6, p. e1007880, 18 Jun. 2019.

SOUZA, T. M. L. *et al.* Emergence of the East-Central-South-African genotype of Chikungunya virus in Brazil and the city of Rio de Janeiro may have occurred years before surveillance detection. **Scientific Reports**, v. 9, n. 1, p. 2760, 26 Feb. 2019.

STOCKDALE, J. E.; LIU, P.; COLIJN, C. The potential of genomics for infectious disease forecasting. **Nature Microbiology**, v. 7, n. 11, p. 1736–1743, Nov. 2022.

TANABE, E. L. DE L. *et al.* Report of East-Central South African Chikungunya virus genotype during the 2016 outbreak in the Alagoas State, Brazil. **Revista Do Instituto de Medicina Tropical de Sao Paulo**, v. 60, p. e19, 23 Apr. 2018.

TANG, K. F.; OOI, E. E. Diagnosis of dengue: an update. **Expert Review of Anti-Infective Therapy**, v. 10, n. 8, p. 895–907, Aug. 2012.

TSETSARKIN, K. A.; WEAVER, S. C. Sequential adaptive mutations enhance efficient vector switching by chikungunya virus and its epidemic emergence. **PLoS Pathogens**, v. 7, n. 12, p. e1002412, 8 Dec. 2011.

TSETSARKIN, K. A. *et al.* A single mutation in chikungunya virus affects vector specificity and epidemic potential. **PLoS Pathogens**, v. 3, n. 12, p. e201, Dec. 2007.

___. Multi-peaked adaptive landscape for chikungunya virus evolution predicts continued fitness optimization in Aedes albopictus mosquitoes. **Nature Communications**, v. 5, p. 4084, 16 Jun. 2014.

VASCONCELOS, H. B. *et al.* Molecular epidemiology of Oropouche virus, Brazil. **Emerging Infectious Diseases**, v. 17, n. 5, p. 800–806, May 2011.

VIANA, R. *et al.* Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. **Nature**, v. 603, n. 7902, p. 679–686, Mar. 2022.

VIEIRA, M. A. C. S. *et al.* West nile virus encephalitis: the first human case recorded in brazil. **The American Journal of Tropical Medicine and Hygiene**, v. 93, n. 2, p. 377–379, Aug. 2015.

VIEIRA, M. A. DA C. E. S. *et al.* Potential role of dengue virus, chikungunya virus and Zika virus in neurological diseases. **Memorias do Instituto Oswaldo Cruz**, v. 113, n. 11, p. e170538, 29 Oct. 2018.

WANG, YUNHAO *et al.* Nanopore sequencing technology, bioinformatics and applications. **Nature Biotechnology**, v. 39, n. 11, p. 1348–1365, 8 Nov. 2021.

WAN, Y. K. *et al.* Beyond sequencing: machine learning algorithms extract biology hidden in Nanopore signal data. **Trends in Genetics**, v. 38, n. 3, p. 246–257, Mar. 2022.

WEAVER, S. C. *et al.* Zika, Chikungunya, and Other Emerging Vector-Borne Viral Diseases. **Annual Review of Medicine**, v. 69, p. 395–408, 29 Jan. 2018.

WENGER, A. M. *et al.* Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. **Nature Biotechnology**, v. 37, n. 10, p. 1155–1162, Oct. 2019.

WILDER-SMITH, A. *et al.* Epidemic arboviral diseases: priorities for research and public health. **The Lancet Infectious Diseases**, v. 17, n. 3, p. e101–e106, Mar. 2017.

WILSON, C. N.; MUSICHA, P.; BEALE, M. A. Genomic epidemiology on the move. **Nature Reviews. Microbiology**, v. 21, n. 2, p. 69, Feb. 2023.

WINOKUR, P. *et al.* Bivalent Omicron BA.1-Adapted BNT162b2 Booster in Adults Older than 55 Years. **The New England Journal of Medicine**, v. 388, n. 3, p. 214–227, 19 Jan. 2023.

WORLD HEALTH ORGANIZATION. **Dengue and severe dengue**. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>. Acesso em: 7 oct. 2023a.

___. **COVID-19 Epidemiological Update - 29 September 2023**. [s.l.] World Health Organization, 29 Sep. 2023b. Disponível em: <https://www.who.int/publications/m/item/covid-19-epidemiological-update---29-september-2023>. Acesso em: 6 oct. 2023.

XAVIER, J. *et al.* Circulation of chikungunya virus East/Central/South African lineage in Rio de Janeiro, Brazil. **Plos One**, v. 14, n. 6, p. e0217871, 11 Jun. 2019.

YIN, Y.; BUTLER, C.; ZHANG, Q. Challenges in the application of NGS in the clinical laboratory. **Human Immunology**, v. 82, n. 11, p. 812–819, Nov. 2021.

YOUNG, P. R. *et al.* Arbovirus Infections. *In*: **Manson's tropical infectious diseases**. [s.l.] Elsevier, 2014. p. 129-161.e3.

YOUNG, P. R. Arboviruses: A family on the move. **Advances in Experimental Medicine and Biology**, v. 1062, p. 1–10, 2018.

ZANLUCA, C. *et al.* First report of autochthonous transmission of Zika virus in Brazil. **Memorias do Instituto Oswaldo Cruz**, v. 110, n. 4, p. 569–572, 9 Jun. 2015.

APÊNDICE A – CLASSIFICAÇÃO DAS SEQUÊNCIAS DE NUCLEOTÍDEOS DO ZIKA VIRUS

Capítulo de livro: Classification of Zika virus sequences with respect to their species and subspecies. Publicado no livro Zika Virus Impact, Diagnosis, Control, and Models: Xavier *et al*., 2021.

Resumo: O vírus Zika (*Orthoflavivirus zikaense*) é um vírus transmitido por artrópodes identificado pela primeira vez em Uganda, na África, que causou grandes epidemias no hemisfério ocidental. Relatos de infecção pelo vírus Zika foram observados em muitos países africanos. Há evidências da circulação do vírus Zika na região asiática desde 1954, embora a primeira epidemia naquela região tenha sido relatada apenas em 2007. Argumenta-se que a linhagem asiática divergiu da linhagem africana por volta de 1950 e depois se espalhou para o Sul da Ásia, Pacífico e Américas. Este trabalho de revisão discute como a similaridade entre sequências de nucleotídeos é utilizada como um critério relevante para a demarcação de espécies virais e para classificar as diferentes linhagens do vírus Zika.

Chapter 3

Classification of Zika virus sequences with respect to their species and subspecies

Joilson Xavier[a], Stephane Tosta[a], Talita Adelino[a,b], Vagner Fonseca[a,c], Marta Giovanetti[a,d], and Luiz Carlos Junior Alcantara[a,d]

[a]*Department of Genetics, Ecology and Evolution, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil,* [b]*Central Laboratory of Public Health of Minas Gerais, Ezequiel Dias Foundation, Belo Horizonte, MG, Brazil,* [c]*Kwazulu-Natal Research Innovation and Sequencing Platform (KRISP), Department of Virology, National Health Laboratory Service (NHLS), University of KwaZulu-Natal—Nelson R. Mandela School of Medicine, Berea, Durban, South Africa,* [d]*Oswaldo Cruz Institute (IOC), Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, RJ, Brazil*

Abbreviations

DNA	deoxyribonucleic acid
ICTV	International Committee on Taxonomy of Viruses
NS5	nonstructural protein 5
ORF	open reading frame
RNA	ribonucleic acid
ViPR	virus pathogen resource

Introduction

Zika virus (ZIKV) is an arthropod-borne virus belonging to the family *Flaviviridae*, genus *Flavivirus*. Its genome is a positive-sense, single-stranded, and nonsegmented RNA of approximately 11 kilobases (Kuno and Chang, 2007).

ZIKV was first identified in 1947 in a rhesus monkey from the Zika Forest in Uganda (Dick, Kitchen, & Haddow, 1952a). For many years the virus was responsible for causing sporadic infections in countries of Africa and Asia, such as Tanzania, Nigeria, and Malaysia (Marchette, Garcia, & Rudnick, 1969; Dick, Kitchen, & Haddow, 1952b; Macnamara, 1954).

Until now the largest epidemic caused by ZIKV was reported in 2015 in Brazil and, subsequently, in several other countries in South, North, and Central America (Pan American Health Organization/World Health Organization, 2016). Estimates indicate that more than 800,000 cases of ZIKV infection were reported in the Americas during that epidemic (Pan American Health Organization/World Health Organization, 2017). A considerable increase in the number of microcephaly cases was reported during the 2015 ZIKV epidemic in Brazil, suggesting for the first time a possible association between ZIKV infection and congenital abnormalities (Schuler-Faccini, 2016; Ministério Da Saúde, Brasil, 2015).

Transmission of ZIKV in urban settings occurs mainly via mosquitoes from the genus *Aedes* spp., also responsible for the transmission of other mosquito-borne viruses. Other nonvector modes of ZIKV transmission include congenital, perinatal, and sexual (Foy et al., 2011; Musso et al., 2015; Calvet et al., 2016; Hoen et al., 2018; Pomar et al., 2018). Possible transmission by blood transfusion has also been described (Musso et al., 2014; Barjas-Castro et al., 2016).

In humans the incubation period from a mosquito bite to symptom onset is ~3–12 days (Ioos et al., 2014). Infection is likely asymptomatic in ~80% of cases (Ioos et al., 2014). When symptoms occur, they are typically mild, self-limiting, and nonspecific; similarity to other arbovirus infections (e.g., dengue and chikungunya viruses) may confound the diagnosis. This superimposition of symptoms, combined with the possibility of cross-reactions with other arboviruses in serological tests, makes the ZIKV diagnosis a complicated task. Commonly reported symptoms include fever, headache, joint pain, muscle pain, conjunctivitis, and maculopapular eruption (Ioos et al., 2014; de Paula Freitas et al., 2016). Severe neurologic sequelae have also been described in adults, including meningitis, meningoencephalitis, and Guillain-Barré syndrome

Zika Virus Impact, Diagnosis, Control, and Models. https://doi.org/10.1016/B978-0-12-820267-8.00003-0

(WHO, 2016). A surge in Guillain-Barré syndrome cases has been observed in Brazil, Colombia, El Salvador, Suriname, Venezuela, and French Polynesia during outbreaks; however, ZIKV has been laboratory confirmed in only some of these cases (WHO, 2016).

Viruses are characterized in terms of sequence differentiation, and viral species are defined as a monophyletic group of viruses whose properties can be distinguished from those of other species by multiple criteria (Adams, Lefkowitz, King, & Carstens, 2013). According to the International Committee on Taxonomy of Viruses (ICTV), the nucleotide and deduced amino acid sequence relatedness is a relevant criterion for the demarcation of species. Other criteria such as antigenic and ecological characteristics, geographic, vector, host, and disease associations are also important to resolve the demarcation of genetically closely related viruses (Simmonds et al., 2017).

Zika virus sequences from African countries

ZIKV is currently taxonomically classified according to the ICTV Online (10th) Report (Simmonds et al., 2017), and the latest change into ZIKV species classification occurred in 2018, after the creation of the Riboviria realm, a taxon that includes RNA viruses (ICTV, 2019).

The first investigative studies on ZIKV epidemiology, mostly through serological surveys, found a wide distribution in African countries, including Gabon, Côte d'Ivoire, Central African Republic, Egypt, Kenya, Nigeria, Senegal, Sierra Leone, Somalia, and Tanzania (Jan, Languillat, Renaudet, & Robin, 1978; Akoua-Koffi et al., 2001; Saluzzo, Gonzalez, Herve, & Georges, 1981; Geser, Henderson, & Christensen, 1970; Fagbami, 1979; Robin & Mouchet, 1975).

The first report of ZIKV was published in 1952, when Dick, Kitchen, and Haddow reported two distinct isolation events in time and from different hosts at Zika Forest, Uganda, Africa (Dick et al., 1952a). Firstly, in April 1947, ZIKV was isolated from sentinel rhesus monkey placed in cages at the Zika Forest. The second isolation occurred 9 months later, in January 1948, from *Aedes africanus* mosquitoes at the same location (Dick et al., 1952a).

The ZIKV prototype (strain MR766 from Uganda) used for the NCBI Reference Sequence (NC_012532.1) was fully sequenced and genomically characterized by Kuno and Chang (2007). The sequence of this African strain comprises 10,794 nucleotides, and its ORF encodes 3419 amino acids. Later, Lanciotti et al. (2008) generated partial *NS5* sequences from three other African strains (strains 41,524, 41,525, and 41,662) isolated in 1984, in Senegal, from *Aedes* spp. mosquitoes.

Zika virus sequences from the Asian region

Following the first cases of ZIKV infection registered in the African continent, ZIKV transmission has been identified in other regions by virus isolation and/or serological studies. Pieces of evidence of ZIKV infection in humans in the Asian region have been registered since 1954 by serological studies performed in Malaysia, India, Philippines, Vietnam, Thailand, and Indonesia (Hammon et al., 1958; Olson et al., 1981; Pond, 1963; Smithburn, Kerr, & Gatne, 1954). However, a strain of the virus (P6-740) was isolated for the first time in Asia only in 1966 from infected mosquitoes *A. aegypti* collected in Malaysia, and the first human case from that region was confirmed in Central Java, Indonesia, in 1977 (Marchette et al., 1969; Olson et al., 1981).

Despite the existence of evidence of ZIKV circulation in Southeast Asia since 1954, an epidemic caused by this virus was reported for the first time in 2007 after an outbreak of a dengue-like illness in the Yap Island, Federated States of Micronesia (Duffy et al., 2009). Strains associated with this epidemic were sequenced, and phylogenetic analysis on the generated *NS5* gene fragment (strain ZIKV 2007 EC) showed that this strain is distantly related to the African sequences and reflected its geographical origin in Southeast Asia (Lanciotti et al., 2008). Moreover the authors of that study also pointed out to the tree topology displaying three subclades, where the first two subclades represent two lineages of African origin (referred to as the West and East African lineages, strains Senegal/1984 and MR766, respectively) and the third subclade represents the Asian lineage comprising the Yap strain (ZIKV 2007 EC) (Lanciotti et al., 2008).

Evolution of the Asian lineage

A different study from 2012, after performing phylogenetic analysis on sequences derived from the Yap Island and from a 2010 ZIKV case in Cambodia, corroborated Lanciotti et al.'s (2008) results by showing two geographically distinct lineages of ZIKV (Asian and African lineages), where the Asian lineage was represented by strains from the Yap epidemic and Cambodia, suggesting that the Asian strain was circulating in Southeast Asia until reaching and triggering the 2007 epidemic in Yap Island (Haddow et al., 2012). A 2014 study found similar results and also estimated the sequences of events that might have represented the emergence and international spread of the Asian lineage (Faye et al., 2014).

Faye et al. (2014) used ZIKV *NS5* and *envelope* genes generated from 43 strains to estimate the time of origin of ZIKV to early 20th century, suggesting the emergence of the virus in Uganda, Africa, around 1920, and from there the virus dispersed to Malaysia, probably around 1945, and then Micronesia (Faye et al., 2014). Another performed Bayesian analysis estimated an alternative earlier time of the most recent common ancestor of ZIKV to around 1837 (Gong et al., 2017). The time of emergence of the Asian lineage in Southeast Asia, in turn, was estimated to be around 1950, supporting the hypothesis of the African origin of ZIKV (Pettersson et al., 2018).

Spread of the Asian lineage to Pacific and Americas

Expansion of ZIKV to the South Pacific was determined after a ZIKV outbreak was reported in French Polynesia in 2013 (Cao-Lormeau et al., 2014). Phylogenetic analysis of partial ZIKV sequences from that outbreak showed that the French Polynesia sequences are closely related to Cambodia 2010 and Yap State 2007 strains (Cao-Lormeau et al., 2014). This expansion to South Pacific was also supported by phylogenetic analysis using the complete coding sequence of the ZIKV strain (H/PF/2013) from the outbreak in French Polynesia in 2013 (Baronti et al., 2014).

Following this spread to South Pacific, it has been argued that a strain from the ZIKV Asian lineage circulating in French Polynesia was likely introduced in the Americas around early 2014 (Faria et al., 2017; Metsky et al., 2017; Hu et al., 2019). Epidemiological and genomic data supported the hypothesis that this introduction might have occurred in the northeast region of Brazil, where the ZIKV caused a large outbreak in 2015 before spreading to other countries in the Americas (Faria et al., 2017; Metsky et al., 2017; Pettersson et al., 2018). The first ZIKV sequences from the 2015 outbreak in Brazil were partial *envelope* sequences generated from patients presenting dengue-like symptoms, and these sequences were closely related to the French Polynesia strain (H/PF/2013) (Zanluca et al., 2015; Campos et al., 2015).

A complete ZIKV genome sequence was obtained from fetal brain tissue of an associated microcephaly case originated in northeast Brazil, and this strain (Natal RGN) has been used as a provisional NCBI Reference Sequence (NC_035889.1) along with the MR 766 African strain (see Fig. 3) (Mlakar et al., 2016). In the following years, more ZIKV complete genome sequences were generated from the epidemic in the American continent (see Fig. 1 and Table 1), and two simultaneous studies from 2017 provided together 164 ZIKV genomes from Brazil and other countries from the Americas to understand the genomic epidemiology of ZIKV (Faria et al., 2017; Metsky et al., 2017). The results from these two studies

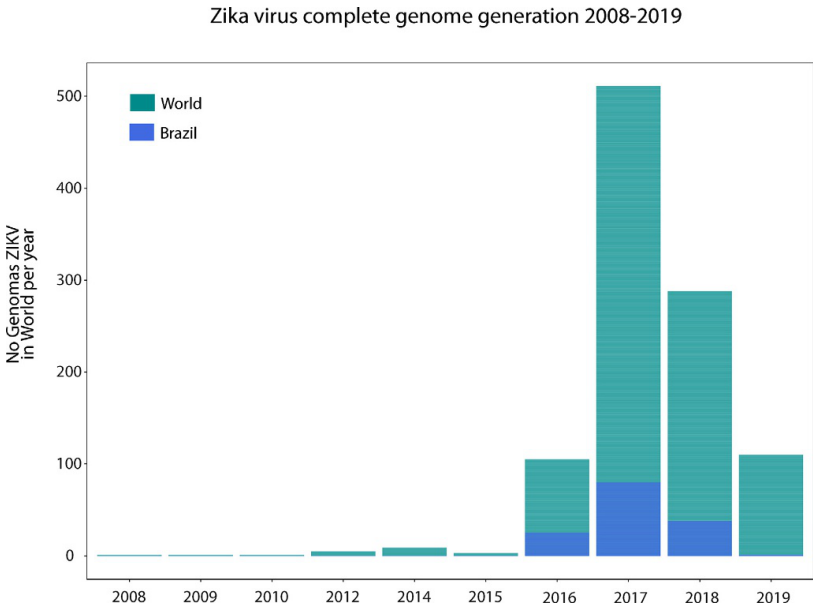


FIG. 1 Zika virus complete genome generation, 2008–19. A bar chart showing the number of Zika virus complete genomic sequences generated in each year. *Blue bars* represent the sequences from Brazilian isolates. The numbers were retrieved from the NCBI public database (www.ncbi.nlm.nih.gov).

TABLE 1 American ZIKV isolates sequenced between 2015 and 2017.

Accession number	Isolate	Lineage/ genotype	Country	Collection date	Host	Sample type	Sequence length (bp)
MH513599	BR/Sinop/H366/2015	Asian	Brazil	2015	Human	Plasma	10,658
KY785450	Zika virus/H. sapiens-wt/BRA/2016/ FC-6863-SER	Asian	Brazil	2016	Human	Serum	10,455
KU527068	Natal RGN	Asian	Brazil	2015	Human	Fetus' brain autopsy	10,808
KY631493	MEX_ENCB165	Asian	Mexico	2015	Human	Serum	10,683
MH063262	ZIKV/Homo_sapiens/Cuba/2017/ FL051Se	Asian	Cuba	2017	Human	Serum	10,619
MF783072	Zika virus/mosquito/Haiti/1855/2016	Asian	Haiti	2016	Mosquito	Mosquito	10,808
KX827268	ZIKV/Homo sapiens/USA/UT-1/2016	Asian	United States	2016	Human	Serum	10,787
KY785451	Zika virus/H. sapiens-wt/MTQ/2016/ FL-001-SAL	Asian	Martinique	2016	Human	Saliva	10,345
KY325467	Zika virus/H. sapiens-wt/USA/2016/ FLUR006	Asian	United States	2016	Human	Urine	10,742
MK028858	Zika virus/H. sapiens-tc/Panama/2015/ 259249	Asian	Panama	2015	Human	Serum	10,772
KY348640	SL1602	Asian	Suriname	2016	Human	Plasma	10,807
KX879604	ZIKV/EC/Esmeraldas/089/2016	Asian	Ecuador	2016	Human	Serum	10,810
KY348640	SL1602	Asian	Suriname	2016	Human	Plasma	10,807
KX893855	Zika virus/Homo sapiens/VEN/UF-2/ 2016	Asian	Venezuela	2016	Human	Serum	10,808
KU758877	17,271	Asian	French Guiana	2015	Human	Urine	10,364
MK049247	Aedes_aegypti/COL/FH04/2016	Asian	Colombia	2016	Mosquito	Mosquito	10,385
KY989971	FLA	Asian	Colombia	2015	Human	Serum	10,807
KY328289	HN16	Asian	Honduras	2016	Human	Serum	10,807
MH675628	CTS-223-16p	Asian	Puerto Rico	2016	Human	Blood	10,807
MK028858	Zika virus/H. sapiens-tc/Panama/2015/ 259249	Asian	Panama	2015	Human	Serum	10,772

A list of some ZIKV isolates from the Americas sequenced between the years 2015 and 2017. Sequence information were retrieved from the GenBank and Virus Pathogen Resource (ViPR) website.

are similar, and molecular clock analyses suggested that the virus emerged in the Americas, specifically in northeast Brazil, around early 2014, and from there the Asian strain disseminated internationally to other regions of the Americas, where thevirus might have circulated undetected for months before the first cases were locally confirmed (Faria et al., 2017; Metsky et al., 2017).

The proposition of three lineages

Since phylogenetic analyses have shown that the ZIKV sequences from the American epidemic clustered to form a single clade, these sequences have been classified as belonging to an American subclade derived from the Asian lineage circulating in Southeast Asia and the Pacific (see Fig. 2) (Faria et al., 2017; Gubler et al., 2017). The ZIKV phylogeny reconstructed by Gubler et al. (2017) showed a single clade comprising 93 ZIKV American strains from different countries from the Americas such as Brazil, Dominican Republic, Colombia, Ecuador, Guatemala, Haiti, Honduras, Mexico, Panama, Puerto Rico, Suriname, the United States, and Venezuela.

Besides the proposition of two major ZIKV lineages (African and Asian lineages), it has been argued that there might be a second African lineage sister to the previously defined African and Asian lineages (Gong et al., 2016; Shen et al., 2016). Using ZIKV *envelope* and *NS5* gene sequences available in GenBank, Shen et al. (2016) reconstructed the ZIKV phylogeny, whose tree topology revealed three ZIKV lineages (African lineage 1, African lineage 2, and Asian/American lineage) designated according to the geographical origin of the clustered strains. The African lineage 2 comprises strains from Senegal and Cote d'Ivoire, and it seems to have separated early from the other two lineages. Since there are no complete genomic sequences of the African lineage 2 and these three lineages appeared only in the *envelope* and *NS5* trees, it has been argued that sampling bias might explain why the African lineage 2 is absent in ZIKV phylogenies based on complete genomic sequences.

Regarding differences between ZIKV lineages, Smith et al. (2018) argued that mutations might influence phenotypic differences observed between ZIKV isolates from the Asian and African lineages, since this study showed approximately 75–100 amino acid residue variations between Asian and African isolates, while approximately 10–30 amino acid variations were observed between American and Asian isolates. A different study from 2018, by analyzing codon usage and nucleotide substitution rates in ZIKV genomes, identified that strains from the African lineage were more diverse than those from the Asian lineage and differential usage of codons was identified in genes from the Asian lineage. Moreover, that study also identified 131 nonsynonymous substitutions in their ZIKV dataset, and a higher selection pressure on the Asian lineage was also identified (Lin et al., 2018) (Figs. 3 and 4).

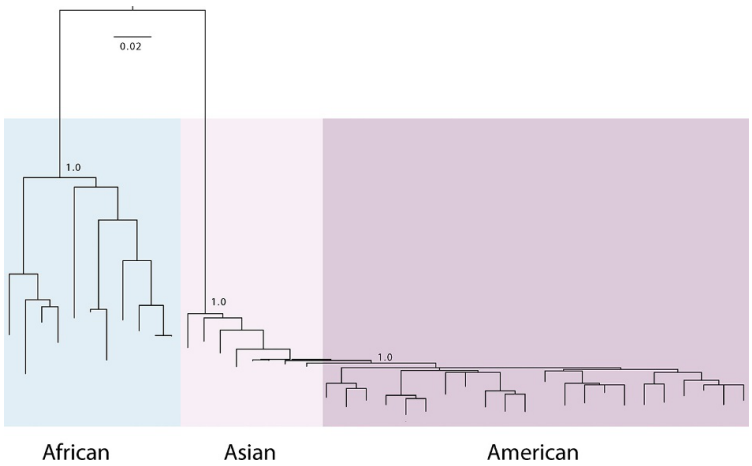


FIG. 2 Phylogenetic tree of Zika virus strains. A maximum likelihood tree reconstructed using complete genomic sequences of Zika virus strains from different countries. Closely related strains clustered to form three distinct groups (clades) that represent the proposed Zika virus lineages discussed in the main text (African, Asian, and American lineages). Sequences were retrieved from the GenBank public database (www.ncbi.nlm.nih.gov/genbank/). Phylogenetic analysis was performed using the IQ-TREE software.

FIG. 3 Alignment of two Zika virus sequences. Alignment of the NS1 gene of the Zika virus reference sequence (accession number: NC_012532) from NCBI database to another Zika virus sequence from Brazil set as provisional reference sequence from NCBI (accession number: NC_035889). This alignment shows a partial region of the NS1 gene, and each nucleotide is represented by a different color. Alignment was performed using MAFFT free alignment software and visualized using AliView.

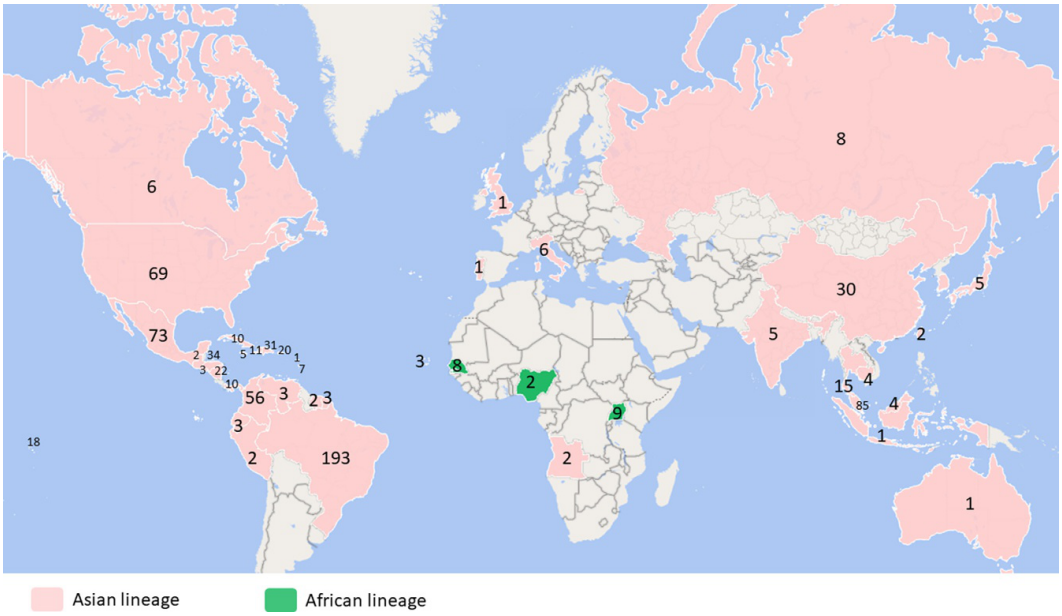


FIG. 4 Number of complete Zika virus sequences according to lineage by country. Geochart showing the number of Zika virus complete genomic sequences generated in each country (here, sequences >40% size of reference genome). Areas in *pink color* represent the sequences classified as Asian lineage, and the *green color* represents sequences classified as African lineage. The numbers were retrieved from the Virus Pathogen Resource (ViPR) public database (https://www.viprbrc.org/).

Conclusion

Recent epidemics caused by ZIKV have drawn attention from the international public due to its rapid spread, the high number of cases, and the neurological consequences associated with the infection. Such threat to public health posed by this virus influenced the World Health Organization's decision to include Zika in the list of priority diseases of the 2018 agenda for research and development, including surveillance and diagnostics. After the ZIKV epidemic in the Americas, an expressive increase in the number of complete and partial ZIKV genomic sequences generated from genomic epidemiology studies has been observed. The availability of this genomic information has allowed the characterization of the evolution and spread of viral lineages. Despite this increase in the amount of genomic data, little information is available on the genomic epidemiology of the virus in the African continent. Thus more genomic surveillance efforts in currently affected areas are required for a better understanding of viral evolution and dispersion dynamics that might help actions of prediction and control of future outbreaks.

Policy and procedures

Sequencing

To obtain genomic sequences from ZIKV, sequencing experiments are performed on clinical samples used as the source of ZIKV genetic material. Viral RNA can be extracted and purified from biological samples, such as saliva, serum, plasma,

and semen, using commercial RNA extraction kits. For the sequencing experiments, it has been commonly used protocols from DNA sequencer machines from Illumina or Oxford Nanopore. A detailed protocol for ZIKV sequencing was published in Quick et al. (2017)

Classification of sequences

Genomic sequences can be classified according to its evolutionary relationship with other isolates or specimens by using Bioinformatics software that calculate similarity among nucleotide sequences from a given dataset and perform serial groupings of those more closely related sequences.

Genetic sequence database

ZIKV sequences can be stored in public databases such as GenBank and Virus Pathogen Resource (ViPR). The GenBank is an example of an online public database that stores DNA sequences and related information. Researchers can store their genetic sequences in GenBank and get an accession number (sequence ID) unique for each sequence. This number allows a user to access information about sample collection and download a file containing the nucleotide sequence of interest.

Mini-dictionary of terms

Bayesian analysis. Here, it refers to a statistical method that uses probability to calculate the chance of a hypothesis.
Guillain-Barré syndrome. A disease in which the immune system attacks the nerves.
Maculopapular eruption. Skin redness with small, high, and soft lumps in the skin.
Meningoencephalitis. Inflammation of the brain and surrounding tissues, usually caused by infection.
Microcephaly. It is a medical condition in which there is a malformation of the brain of a newborn.
Molecular clock analyses. It is a method for estimating the evolutionary rate between species.
Monophyletic. A group that includes the common ancestor and all their descendants.
Phylogenetic analysis. A tool for estimating evolutionary relationships between genes, genomes, and species.
Phylogeny. Representation of the evolutionary relationship between groups of related species.
Subclades. A subgroup formed by evolutionarily related strains.

Key facts of Zika virus lineage

- The Zika virus reference strain MR766 was isolated from mosquitoes in Uganda in 1948, and its complete genomic sequence was published in 2007.
- A Zika virus strain from the Asian lineage was isolated for the first time in Asia in 1966.
- The first epidemic in Southeast Asia caused by an Asian lineage strain was reported in the Yap Island, Federated States of Micronesia, in 2007.
- Expansion of the Asian lineage to the South Pacific was identified when an epidemic was reported in French Polynesia in 2013.
- The Asian lineage reached the Americas and caused a large epidemic in 2015, and 164 complete genomic sequences were generated from that epidemic.
- Zika virus sequences from the epidemic in the Americas form a single distinct clade, named the American subclade.

Summary points

- Zika virus is an arthropod-borne virus first identified in Uganda, Africa, that has caused large epidemics in the western hemisphere.
- The use of nucleotide sequence relatedness is a relevant criterion for the demarcation of viral species.
- Reports of Zika virus infection have been observed throughout many African countries.
- There is evidence of the Zika virus circulation in the Asian region since 1954 although the first epidemic in that region was reported only in 2007.
- It is argued that the Asian lineage diverged from the African lineage around 1950 and then spread to South Asia, the Pacific, and the Americas.
- A 2016 study suggested that Zika virus sequences can be classified into three lineages.

References

Adams, M. J., Lefkowitz, E. J., King, A. M. Q., & Carstens, E. B. (2013). Recently agreed changes to the International Code of Virus Classification and Nomenclature. *Archives of Virology*, *158*(12), 2633–2639.

Akoua-Koffi, C., Diarrassouba, S., Benie, V. B., Ngbichi, J. M., Bozoua, T., Bosson, A., … Ehouman, A. (2001). Investigation surrounding a fatal case of yellow fever in Cote d'Ivoire in 1999. *Bulletin de la Societe de pathologie exotique (1990)*, *94*(3), 227–230.

Barjas-Castro, M. L., Angerami, R. N., Cunha, M. S., Suzuki, A., Nogueira, J. S., Rocco, I. M., … Stucchi, R. S. (2016). Probable transfusion-transmitted Zika virus in Brazil. *Transfusion*, *56*(7), 1684–1688.

Baronti, C., Piorkowski, G., Charrel, R. N., Boubis, L., Leparc-Goffart, I., & de Lamballerie, X. (2014). Complete coding sequence of Zika virus from a French polynesia outbreak in 2013. *Genome Announcements*, *2*(3). e00500-14.

Calvet, G., Aguiar, R. S., Melo, A. S., Sampaio, S. A., De Filippis, I., Fabri, A., … Tschoeke, D. A. (2016). Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: A case study. *The Lancet Infectious Diseases*, *16*(6), 653–660.

Campos, G. S., Bandeira, A. C., & Sardi, S. I. (2015). Zika virus outbreak, Bahia, Brazil. *Emerging Infectious Diseases*, *21*(10), 1885.

Cao-Lormeau, V. M., Roche, C., Teissier, A., Robin, E., Berry, A. L., Mallet, H. P., … Musso, D. (2014). Zika virus, French polynesia, South Pacific, 2013. *Emerging Infectious Diseases*, *20*(6), 1085.

de Paula Freitas, B., de Oliveira Dias, J. R., Prazeres, J., Sacramento, G. A., Ko, A. I., Maia, M., & Belfort, R. (2016). Ocular findings in infants with microcephaly associated with presumed Zika virus congenital infection in Salvador, Brazil. *JAMA Ophthalmology*, *134*(5), 529–535.

Dick, G. W. A., Kitchen, S. F., & Haddow, A. J. (1952a). Zika virus (I). Isolations and serological specificity. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *46*(5), 509–520.

Dick, G. W. A., Kitchen, S. F., & Haddow, A. J. (1952b). Zika virus (II). Pathogenicity and physical properties. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *46*(5), 521–534.

Duffy, M. R., Chen, T. H., Hancock, W. T., Powers, A. M., Kool, J. L., Lanciotti, R. S., … Guillaumot, L. (2009). Zika virus outbreak on Yap Island, Federated States of Micronesia. *New England Journal of Medicine*, *360*(24), 2536–2543.

Fagbami, A. H. (1979). Zika virus infections in Nigeria: Virological and seroepidemiological investigations in Oyo State. *Epidemiology and Infection*, *83*(2), 213–219.

Faria, N. R., Quick, J., Claro, I. M., Theze, J., de Jesus, J. G., Giovanetti, M., … Franco, L. C. (2017). Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature*, *546*(7658), 406.

Faye, O., Freire, C. C., Iamarino, A., Faye, O., de Oliveira, J. V. C., Diallo, M., & Zanotto, P. M. (2014). Molecular evolution of Zika virus during its emergence in the 20th century. *PLoS Neglected Tropical Diseases*, *8*(1), e2636.

Foy, B. D., Kobylinski, K. C., Foy, J. L. C., Blitvich, B. J., da Rosa, A. T., Haddow, A. D., … Tesh, R. B. (2011). Probable non-vector-borne transmission of Zika virus, Colorado, USA. *Emerging Infectious Diseases*, *17*(5), 880.

Geser, A., Henderson, B. E., & Christensen, S. (1970). A multipurpose serological survey in Kenya: 2. Results of arbovirus serological tests. *Bulletin of the World Health Organization*, *43*(4), 539.

Gong, Z., Gao, Y., & Han, G. Z. (2016). Zika virus: Two or three lineages? *Trends in Microbiology*, *24*(7), 521–522.

Gong, Z., Xu, X., & Han, G. Z. (2017). The diversification of Zika virus: Are there two distinct lineages? *Genome Biology and Evolution*, *9*(11), 2940–2945.

Gubler, D. J., Vasilakis, N., & Musso, D. (2017). History and emergence of Zika virus. *Journal of Infectious Diseases*, *216*(Suppl_10), S860–S867.

Haddow, A. D., Schuh, A. J., Yasuda, C. Y., Kasper, M. R., Heang, V., Huy, R., … Weaver, S. C. (2012). Genetic characterization of Zika virus strains: Geographic expansion of the Asian lineage. *PLoS Neglected Tropical Diseases*, *6*(2), e1477.

Hammon, W. M., Schrack, W. D., Jr., & Sather, G. E. (1958). Serological survey for arthropod-borne virus infections in the Philippines. *American Journal of Tropical Medicine and Hygiene*, *7*(3), 323–328.

Hoen, B., Schaub, B., Funk, A. L., Ardillon, V., Boullard, M., Cabié, A., … Douine, M. (2018). Pregnancy outcomes after ZIKV infection in French territories in the Americas. *New England Journal of Medicine*, *378*(11), 985–994.

Hu, T., Li, J., Carr, M. J., Duchêne, S., & Shi, W. (2019). The Asian lineage of Zika virus: Transmission and evolution in Asia and the Americas. *Virologica Sinica*, *34*(1), 1–8.

ICTV. (2019). *ICTV taxonomy history: Zika virus*. Available from https://talk.ictvonline.org/taxonomy/p/taxonomy-history?taxnode_id=201853123. (Accessed 13 December 2019).

Ioos, S., Mallet, H. P., Goffart, I. L., Gauthier, V., Cardoso, T., & Herida, M. (2014). Current Zika virus epidemiology and recent epidemics. *Medecine et maladies infectieuses*, *44*(7), 302–307.

Jan, C., Languillat, G., Renaudet, J., & Robin, Y. (1978). A serological survey of arboviruses in Gabon. *Bulletin de la Societe de pathologie exotique et de ses filiales*, *71*(2), 140–146.

Kuno, G., & Chang, G. J. (2007). Full-length sequencing and genomic characterization of Bagaza, Kedougou, and Zika viruses. *Archives of Virology*, *152*(4), 687–696.

Lanciotti, R. S., Kosoy, O. L., Laven, J. J., Velez, J. O., Lambert, A. J., Johnson, A. J., … Duffy, M. R. (2008). Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. *Emerging Infectious Diseases*, *14*(8), 1232.

Lin, D., Li, L., Xie, T., Yin, Q., Saksena, N., Wu, R., … Chen, X. (2018). Codon usage variation of Zika virus: The potential roles of NS2B and NS4A in its global pandemic. *Virus Research*, *247*, 71–83.

Macnamara, F. N. (1954). Zika virus: A report on three cases of human infection during an epidemic of jaundice in Nigeria. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *48*(2), 139–145.

Marchette, N. J., Garcia, R., & Rudnick, A. (1969). Isolation of Zika virus from Aedes aegypti mosquitoes in Malaysia. *American Journal of Tropical Medicine and Hygiene*, *18*(3), 411–415.

Metsky, H. C., Matranga, C. B., Wohl, S., Schaffner, S. F., Freije, C. A., Winnicki, S. M., … Lin, A. E. (2017). Zika virus evolution and spread in the Americas. *Nature*, *546*(7658), 411.

Ministério Da Saúde, Brasil. (2015). *Procedimentos preliminares a serem adotados para a vigilância dos casos de microcefalia no Brasil*. Nota Informativa n^0 01/2015. Available from https://portalarquivos2.saude.gov.br/images/pdf/2015/novembro/18/microcefalia-nota-informativa-17nov2015-c.pdf. (Accessed 31 October 2019).

Mlakar, J., Korva, M., Tul, N., Popović, M., Poljšak-Prijatelj, M., Mraz, J., … Vizjak, A. (2016). Zika virus associated with microcephaly. *New England Journal of Medicine*, *374*(10), 951–958.

Musso, D., Nhan, T., Robin, E., Roche, C., Bierlaire, D., Zisou, K., … Broult, J. (2014). Potential for Zika virus transmission through blood transfusion demonstrated during an outbreak in French Polynesia, November 2013 to February 2014. *Eurosurveillance*, *19*(14), 20761.

Musso, D., Roche, C., Robin, E., Nhan, T., Teissier, A., & Cao-Lormeau, V. M. (2015). Potential sexual transmission of Zika virus. *Emerging Infectious Diseases*, *21*(2), 359.

Pan American Health Organization/World Health Organization. (2016). *Timeline of the emergence of Zika virus in the Americas*. Washington, DC: PAHO/WHO. Available from https://www.paho.org/hq/index.php?option=com_content&view=article&id=11959:timeline-of-emergence-of-zika-virus-in-the-americas&Itemid=41711&lang=en. (Accessed 13 December 2019).

Pan American Health Organization/World Health Organization. (2017). *Zika suspected and confirmed cases reported by countries and territories in the Americas cumulative cases, 2015–2017*. Available from https://www.paho.org/hq/dmdocuments/2017/2017-ago-10-phe-ZIKV-casess.pdf. (Accessed 28 October 2019).

Pettersson, J. H. O., Bohlin, J., Dupont-Rouzeyrol, M., Brynildsrud, O. B., Alfsnes, K., Cao-Lormeau, V. M., … Musso, D. (2018). Re-visiting the evolution, dispersal and epidemiology of Zika virus in Asia. *Emerging Microbes and Infections*, *7*(1), 1–8.

Pomar, L., Vouga, M., Lambert, V., Pomar, C., Hcini, N., Jolivet, A., … Panchaud, A. (2018). Maternal-fetal transmission and adverse perinatal outcomes in pregnant women infected with Zika virus: Prospective cohort study in French Guiana. *BMJ*, *363*, k4431.

Pond, W. L. (1963). Arthropod-borne virus antibodies in sera from residents of South-East Asia. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *57*(5), 364–371.

Robin, Y., & Mouchet, J. (1975). Serological and entomological study on yellow fever in Sierra Leone. *Bulletin de la Societe de Pathologie Exotique et de ses Filiales*, *68*(3), 249–258.

Saluzzo, J. F., Gonzalez, J. P., Herve, J. P., & Georges, A. J. (1981). Serological survey for the prevalence of certain arboviruses in the human population of the south-east area of Central African Republic (author's transl). *Bulletin de la Societe de pathologie exotique et de ses filiales*, *74*(5), 490–499.

Schuler-Faccini, L. (2016). Possible association between Zika virus infection and microcephaly—Brazil, 2015. *MMWR Morbidity and Mortality Weekly Report*, *65*, 59–62.

Shen, S., Shi, J., Wang, J., Tang, S., Wang, H., Hu, Z., & Deng, F. (2016). Phylogenetic analysis revealed the central roles of two African countries in the evolution and worldwide spread of Zika virus. *Virologica Sinica*, *31*(2), 118–130.

Simmonds, P., Becher, P., Bukh, J., Gould, E. A., Meyers, G., Monath, T., … Stapleton, J. T. (2017). ICTV virus taxonomy profile: Flaviviridae. *The Journal of General Virology*, *98*(1), 2.

Smith, D. R., Sprague, T. R., Hollidge, B. S., Valdez, S. M., Padilla, S. L., Bellanca, S. A., … Haddow, A. D. (2018). African and Asian Zika virus isolates display phenotypic differences both in vitro and in vivo. *American Journal of Tropical Medicine and Hygiene*, *98*(2), 432–444.

World Health Organization (WHO). (2016). *Zika situation report: Neurological syndrome and congenital anomalies*. Available from http://www.who.int/emergencies/zika-virus/situation-report/5-february-2016/en/. (Accessed 13 December 2019).

Zanluca, C., Melo, V. C. A. D., Mosimann, A. L. P., Santos, G. I. V. D., Santos, C. N. D. D., & Luz, K. (2015). First report of autochthonous transmission of Zika virus in Brazil. *Memórias do Instituto Oswaldo Cruz*, *110*(4), 569–572.

Olson, J. G., et al. (1981). Zika virus, a cause of fever in Central Java, Indonesia. *Transactions of The Royal Society of Tropical Medicine and Hygiene*. https://doi.org/10.1016/0035-9203(81)90100-0.

Smithburn, K. C., Kerr, J. A., & Gatne, P. B. (1954). Neutralizing antibodies against certain viruses in the sera of residents of India. *The Journal of Immunology*, *72*(4), 248–257.

Quick, J., et al. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nature Protocols*. https://doi.org/10.1038/nprot.2017.066.

APÊNDICE B - VIGILÂNCIA GENÔMICA DO SARS-COV-2 EM MINAS GERAIS, BRASIL

Artigo: The ongoing COVID-19 epidemic in Minas Gerais, Brazil: insights from epidemiological data and SARS-CoV-2 whole genome sequencing. Publicado na revista Emerging Microbes & Infections: Xavier et al., 2020.

Resumo: Até 03 de Julho de 2020, mais de 10,7 milhões de casos da doença causada pelo vírus SARS-CoV-2, denominada COVID-19, foram relatados globalmente. Naquele período, mais de 61 mil casos e 4.205 mortes atribuídas à COVID-19 foram relatados no Brasil. O estado de Minas Gerais (MG), com a segunda maior população do país, apresentava naquele período um elevado risco de ser altamente afetado pela pandemia do COVID-19 devido ao seu grande tamanho populacional e fortes conexões com estados vizinhos como São Paulo e Rio de Janeiro. Até 28 de Abril, MG tinha registrado 71 mortes relacionadas à COVID-19 e a capital, Belo Horizonte, relatou 555 casos da doença. Neste trabalho foi empregado a técnica de sequenciamento por nanoporos, que permitiu a geração de 40 novas sequências do genoma completo do vírus SARS-CoV-2, que foram disponibilizadas em bancos de dados públicos, contribuindo com o aumento da quantidade de dados genômicos disponíveis sobre o vírus circulante no Brasil naquele período. As análises filogenéticas realizadas neste estudo forneceram informações sobre os múltiplos eventos de introdução do vírus no estado de MG. Este trabalho também mostrou que a epidemia de COVID-19, no final de abril de 2020 no estado de MG estava em expansão (R> 1) e altamente dispersa geograficamente, com muitos casos e mortes relatados fora da capital.

Emerging Microbes & Infections
2020, VOL. 9
https://doi.org/10.1080/22221751.2020.1803146

The ongoing COVID-19 epidemic in Minas Gerais, Brazil: insights from epidemiological data and SARS-CoV-2 whole genome sequencing

Joilson Xavier [a]*, Marta Giovanetti[a,b]*, Talita Adelino[a,c]*, Vagner Fonseca [a,d]*, Alana Vitor Barbosa da Costa[c], Adriana Aparecida Ribeiro[c], Katlin Nascimento Felicio[c], Clara Guerra Duarte[c], Marcos Vinicius Ferreira Silva[c], Álvaro Salgado[a], Mauricio Teixeira Lima[a], Ronaldo de Jesus[a], Allison Fabri[b], Cristiane Franco Soares Zoboli[c], Thales Gutemberg Souza Santos[c], Felipe Iani[a,c], Massimo Ciccozzi [k], Ana Maria Bispo de Filippis[b], Marilda Agudo Mendonça Teixeira de Siqueira[e], André Luiz de Abreu[f], Vasco de Azevedo[a], Dario Brock Ramalho[g], Carlos F. Campelo de Albuquerque[h], Tulio de Oliveira[d], Edward C. Holmes[i], José Lourenço[j]*, Luiz Carlos Junior Alcantara[a,b]* and Marluce Aparecida Assunção Oliveira[c]*

[a]Laboratório de Genética Celular e Molecular, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; [b]Laboratório de Flavivírus, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil; [c]Laboratório Central de Saúde Pública, Fundação Ezequiel Dias, Belo Horizonte, Brazil; [d]KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), College of Health Sciences, University of KwaZuluNatal, Durban 4001, South Africa; [e]Laboratório de Vírus Respiratórios e Sarampo, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil; [f]Coordenação Geral dos Laboratórios de Saúde Pública/Secretaria de Vigilância em Saúde, Ministério da Saúde, Brasília, Distrito Federal, Brazil; [g]Secretaria de Estado de Saúde de Minas Gerais, Belo Horizonte, Brazil; [h]Organização Pan-Americana da Saúde/Organização Mundial da Saúde, Brasília-DF, Brazil; [i]Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life and Environmental Sciences and School of Medical Sciences, University of Sydney, Sydney, NSW, Australia; [j]Department of Zoology, University of Oxford, Oxford OX1 3PS, UK; [k]University Campus Bio-Medico of Rome, Rome, Italy

ABSTRACT

The recent emergence of a coronavirus (SARS-CoV-2), first identified in the Chinese city of Wuhan in December 2019, has had major public health and economic consequences. Although 61,888 confirmed cases were reported in Brazil by 28 April 2020, little is known about the SARS-CoV-2 epidemic in this country. To better understand the recent epidemic in the second most populous state in southeast Brazil - Minas Gerais (MG) - we sequenced 40 complete SARS-CoV-2 genomes from MG cases and examined epidemiological data from three Brazilian states. Both the genome analyses and the geographical distribution of reported cases indicate for multiple independent introductions into MG. Epidemiological estimates of the reproductive number (R) using different data sources and theoretical assumptions suggest the potential for sustained virus transmission despite a reduction in R from the first reported case to the end of April 2020. The estimated date of SARS-CoV-2 introduction into Brazil was consistent with epidemiological data from the first case of a returned traveller from Lombardy, Italy. These findings highlight the nature of the COVID-19 epidemic in MG and reinforce the need for real-time and continued genomic surveillance strategies to better understand and prepare for the epidemic spread of emerging viral pathogens..

ARTICLE HISTORY Received 27 May 2020; Revised 6 July 2020; Accepted 26 July 2020

KEYWORDS SARS-CoV-2; genomic surveillance; Minas Gerais; southeast Brazil; pandemic; sequencing; genomic epidemiology

Introduction

The World Health Organization (WHO) office in China was informed about a cluster of new cases of pneumonia of unknown aetiology in the city of Wuhan (Hubei province), in late December 2019 [1]. Shortly afterwards, a new type of coronavirus, now termed SARS-CoV-2, was isolated and identified by Chinese authorities, with its genetic sequence shared with the international community on 10 January 2020 [2–5]. Phylogenetic analysis revealed that SARS-CoV-2 was similar to other (epidemic) betacoronaviruses, such as severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) [4,5], and was clearly related to coronaviruses isolated from bats and Malayan pangolins (*Manis javanica*) indicative of a zoonotic origin [2,5–7].

To date, more than 10.7 million cases of the disease caused by SARS-CoV-2, termed COVID-19, have been reported globally [8,9]. On 11 March 2020, the WHO declared a pandemic, prompting a dramatic increase in international concern and response [10]. On 26 February 2020, the first confirmed case of COVID-19 was reported in São Paulo (SP) state, Brazil [11]. Two months later (28 April 2020), 61,888 cases and 4,205 deaths attributed to COVID-19 had been reported in

Brazil [12]. Initial phylogenetic analysis using the first two SARS-CoV-2 complete genomes isolated in São Paulo from travellers returning from Italy revealed two independent introductions into the country relative to the data set available at that time [13].

The state of Minas Gerais (MG) is the second largest Brazilian state in terms of population size, estimated at approximately 21 million people [14]. Due to its large population size and strong connections to active neighbouring states such as São Paulo and Rio de Janeiro, the state of MG is likely to be highly affected by the COVID-19 pandemic.

Genetic analyses and surveillance allow the characterization of circulating viral lineages, the inference of introduction events and the reconstruction of transmission patterns [15]. Together with epidemiological data they constitute powerful tools to assist public health initiatives and preparedness. Herein, we present a summary of epidemiological data and the generation and analysis of 40 new SARS-CoV-2 genome sequences isolated from clinical samples of confirmed cases from MG. Our aim was to provide a preliminary epidemiological overview of SARS-CoV-2 circulation and introduction events into Minas Gerais.

Materials and methods

Ethics statement

Anonymised samples processed in this study were sent to the Central Public Health Laboratory/Octávio Magalhães Institute (IOM) of the Ezequiel Dias Foundation (FUNED), which belongs to public laboratories network from the *Brazilian Ministry of Health* (BMoH). They were previously obtained by the local health services for the routine diagnosis and epidemiological surveillance of SARS-CoV-2. The availability of these samples for research purposes during outbreaks of international concern is allowed under the terms of the 510/2016 Resolution of the National Ethical Committee for Research – Brazilian Ministry of Health (CONEP - Comissão Nacional de Ética em Pesquisa, Ministério da Saúde): this authorizes the use of clinical samples collected in the Brazilian Central Public Health Laboratories to accelerate knowledge building and contribute to surveillance and outbreak response. The study protocol was reviewed and approved by Research Ethics Committee of the Universidade Federal de Minas Gerais with approval No. 32912820.6.1001.5149.

Sample collection and RT-qPCR diagnosis

Samples used in this study were from residual anonymised clinical samples, with no or minimal risk to patients, provided for research and surveillance purposes as described above. *Swab* samples collected from COVID-19 suspected cases were collected throughout the state of MG and sent to IOM-FUNED facilities. At the IOM-FUNED facilities they were submitted to total RNA extraction with an automated protocol on the QIAsymphony platform using the DSP Virus/Pathogen Kit (Qiagen), following the manufacturer's recommendations. *The molecular diagnosis was performed on a 7500 Real-Time PCR System (Thermofisher Scientific), using a RT-qPCR singleplex kit for the SARS-CoV-2 envelope and RNA-dependent RNA polymerase genes developed by Bio-Manguinhos/Fiocruz (Rio de Janeiro, Brazil) and provided by the Brazilian Ministry of Health, following the manufacturer's recommendations*. We selected 48 samples with RT-qPCR positive results, collected up to 3 April 2020 from patients residing in different municipalities of the state of MG and presenting with symptoms such as fever, cough, headache, dyspnea, *sore throat and/or vomiting. Samples were selected based on cycle threshold (Ct) values* ≤ 32. *Associated epidemiological data, such as symptoms, travel history and municipality of residency, were collected from medical records accompanying the collected samples provided by IOM/FUNED*.

cDNA synthesis and sequencing multiplex PCR

For cDNA synthesis, the SuperScript IV Reverse Transcriptase kit (Invitrogen) was used following the manufacturer's instructions. The generated cDNA was subjected to multiplex PCR sequencing using Q5 High Fidelity Hot-Start DNA Polymerase (New England Biolabs) and a set of specific primers designed by the ARTIC Network (https://github.com/artic-network/artic-ncov2019/tree/master/primer_schemes/nCoV-2019/V1) for sequencing the complete SARS-CoV-2 genome [16]. PCR conditions have been previously reported in [16]. All experiments were performed in biosafety level-2 cabinet.

Whole genome sequencing

Amplified PCR products were purified using the 1x AMPure XP Beads (Beckman Coulter) following previously a published protocol [17]. Purified PCR products were quantified using the Qubit® dsDNA HS Assay Kits (Invitrogen), following the manufacturer's instructions. Of the 48 samples, only 40 contained sufficient DNA ($\geq 2ng/\mu L$) to proceed to library preparation. Sequencing libraries were prepared using the Oxford Nanopore Ligation Sequencing Kit (SQK-LSK109) following previously a published protocol [17]. Before pooling all samples, each sample was barcoded using the Native Barcoding Expansion kits (NBD104 and EXP-NBD114). After barcoding adaptor ligation, sequencing libraries were loaded on a flow cell (FLO-MIN106) for subsequent MinION sequencing,

programmed to run for six hours. Reads were *base-called* using Guppy and barcode demultiplexing was performed using qcat. Consensus sequences were generated by *de novo* assembling using Genome Detective and Coronavirus Typing Tool [18,19].

Phylogenetic analysis

Public SARS-CoV-2 complete genome sequences available up to 15 April 2020 were retrieved from the GISAID database (https://www.gisaid.org/). Sequences were aligned using MAFFT (FF-NS-2 algorithm) employing default parameters [20]. The alignment was manually curated to remove artefacts at the terminal regions and within the alignment using Aliview [21]. Phylogenetic analysis of these sequences was performed using the maximum likelihood method implemented in IQ-TREE (version 1.6.10), employing the best-fit model of nucleotide substitution according to the Bayesian Information Criterion (BIC) as indicated by the Model Finder application implemented in IQ-TREE [22]. The statistical robustness of individual nodes was determined using 1000 bootstrap replicates.

Lineage assessment was conducted using the Phylogenetic Assignment of Named Global Outbreak LINeages tool available at https://github.com/hCoV-2019/pangolin [23]. Four data sets of complete or near-complete SARS-CoV-2 genome sequences were generated. Data set 1 ($n = 3{,}102$) comprised the data reported in this study ($n = 40$) plus publicly available SARS-CoV-2 sequences ($n = 3{,}062$) from GISAID. Subsequently, to investigate the evolutionary behaviour of SARS-CoV-2 within the three different SARS-CoV-2 lineages (A, B and B.1), Bayesian molecular clock analysis was conducted on three smaller subsets of data for each of the three lineages identified in the ML phylogeny and containing isolates from MG (data set 2 *for* subset A $n = 100$; data set 3 *for* subset B $n = 84$; data set 4 *for* subset B.1 $n = 169$). ML trees from these three data subsets were inspected in TempEst v1.5.3 for the presence of a temporal (i.e. molecular clock) signal [24]. Linear regression of root-to-tip genetic distances against sampling dates indicated that the SARS-CoV-2 sequences evolve in a relatively-strong clock-like manner ($r = 0{\cdot}43$; $r = 0{\cdot}47$; $r = 0{\cdot}40$ from subset A; B and B.1, respectively) (Supplementary Figure S18). For more detailed Bayesian time-scaled phylogenetic analysis we employed the BEAST 1.10.4 program [25]. For this analysis we employed the strict molecular clock model, the HKY+Γ4 codon partitioned (CP)1 + 2,3 nucleotide substitution model and the exponential growth coalescent model [26]. We computed MCMC (Markov chain Monte Carlo) triplicate runs of 100 million states each, sampling every 10.000 steps for each data set. Convergence of MCMC chains was checked using Tracer v.1.7.1 [27]. Maximum clade credibility *trees* were summarized from the MCMC samples using TreeAnnotator after discarding 10% as burn-in.

Epidemiological data assembly

Data used in the epidemiological analysis were retrieved from https://github.com/wcota/covid19br [28].

Results

After the WHO declared the outbreak of SARS-CoV-2 a Public Health Emergency of International Concern (PHEIC) on 30 January 2020, the Brazilian government declared a Public Health Emergency of National Importance on 3 February 2020, enabling the introduction of measures to prevent and control virus spread [29]. Twenty-three days later, the first confirmed case in Brazil was reported in the city of São Paulo, related to a traveller returning from Lombardy, Italy (Figure 1) [11]. By the 28th April 2020, more than 61,888 COVID-19 cases were confirmed in Brazil, 1,578 of which were from MG (Figure 2A) [30]. Over this period, MG registered 71 COVID-19-related deaths and the capital city, Belo Horizonte, with an estimated population of 2.5 million people, reported 555 cases [30,31]. Figure 2A shows MG's epidemic (reported cases) curve compared to those of two other neighbouring states, São Paulo (SP) and Rio de Janeiro (RJ). Temporal trends reveal a growth in the number of cases in MG, although with SP accounting for the largest number of cases and deaths (24,041 cases; 2,049 deaths) by COVID-19 up to 28th April 2020.

Epidemiological analyses using cases and mortality time series data from three Brazilian states

Without access to the total number of tests in time and in each state, we obtained a crude estimate of the case fatality ratio (CFR) for MG, SP and RJ as the ratio between the number of reported deaths and cases [32]. Accordingly, CFR was found to increase with time in all states (Supplementary Figure S2), with means from date of first reported case up to the 28 April in each state as 2.67% for MG, 5.39% for RJ and 6.0% for SP. Although all such comparisons should be made with caution, the CFR in SP and RJ was consistently higher than reported in other localities; for example, 2.6% (95% CI 0·89-6·7) for the Diamond Princess cruise ship [33], and 3.67% (95% CI 3.56-3.80) and 1.2% (95% CI 0.3-2.7) and 1.4% (95% CI 0.9-2.1) for different Chinese regions [33–35].

We used the mortality time series (MTS) from MG, SP and RJ to project the (unobserved total) cumulative number of infections, making two main simplifying assumptions: first, that the infection fatality ratio (IFR) of SARS-CoV-2 would be similar in the Brazilian states to that reported elsewhere; and second, that the number of cumulative deaths in each state were well reported. We utilised

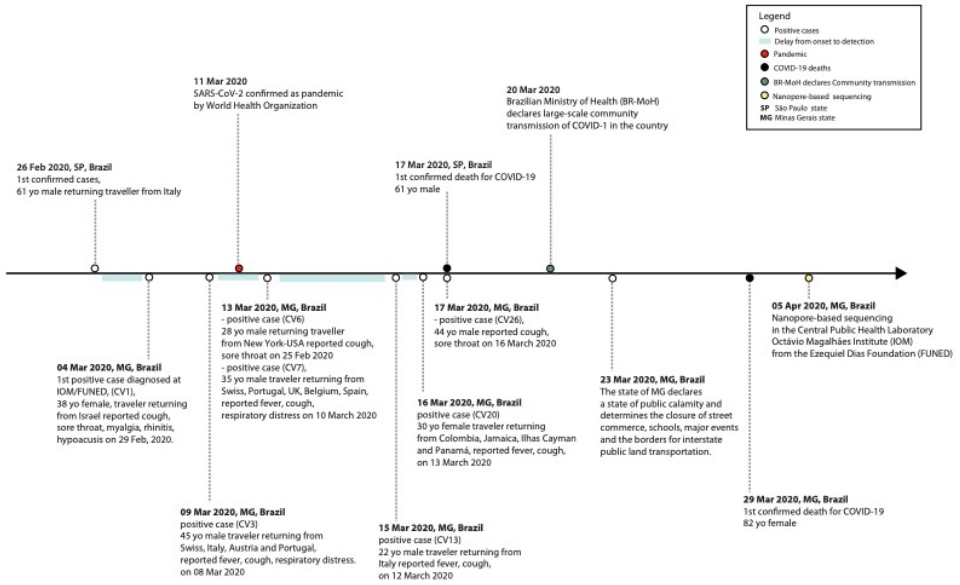


Figure 1. Timeline of key events following the first confirmed case of COVID-19 in Brazil. Events below the line occurred in Minas Gerais (MG) state, while national events are presented above the line. Codes in parentheses refer to the identification code (CV#) of the isolates from the cases described in this study.

the IFR estimated by Verity and colleagues (0.66%, CI 95% 0.39-1.33% [34]), for its general use in the modelling literature [36]. The cumulative number of infections in time was taken to be $I(t) = \frac{D(t)}{\frac{IFR}{100}}$, where $D(t)$ is the cumulative number of deaths. From $I(t)$ we further obtain the observation rate θ of reported cases from $\theta(t) = \frac{c(t)}{I(t)}$ where $c(t)$ is the number of reported cases in time. We found that the observation rate decreased in time for all states, a likely outcome of epidemic growth superseding tracing and testing efforts as the epidemic progressed (Supplementary Figure S1). By 28 April 2020, the last time point analysed, RJ and SP had similar observation rates at 7.6% and 7.74% (respectively), while in MG, where the epidemic started later, the observation rate was 15.3% (i.e. 1 reported case in 7 infections).

To compare transmission potential, we used reported cases (CTS) and mortality time series (MTS) from MG, SP and RJ states to estimate the (effective) reproduction number, R. For this, we performed maximum likelihood estimation of the (CTS and MTS) epidemic growth r using a phenomenological model, and two theoretical formulations on how r relates to R - one based on the SEIR epidemiological framework by Wallinga and colleagues [37], and another on the distribution of the serial interval [36] (see Supplementary Material for details). For all states R was found to decrease in time since the first reported case (Supplementary Figures S5 and S10). When considering the entire period from first reported case to the 28th of April, estimation methods estimated similar R values for each state (Supplementary Figures S6 and S11). For example, when using the CTS and serial interval formulation, R was 1.91 (CI 95% 1.2-3.1) for SP, 1.88 (CI 95% 1.27-2.8) for RJ and 1.82 (CI 95% 1.2-3.25) for MG.

Using geographic information from reported data in each state (Figure 2C), we found cases and deaths to be disproportionally reported in capital cities for the RJ and SP states but not for MG (Supplementary Figures S13-14). Typically, incidence (cases, deaths) are normalized per 100K individuals, taking into account the total population size of each state. Because of the very different spatial dispersion of cases and deaths in MG when compared to SP and RJ, we also calculated the effective population size; that is, the sum of the population sizes of all municipalities with reports. When using reported cases, we found that the effective population sizes were ~100%, ~100% and 64% of the total population sizes of RJ, SP and MG, respectively. When using reported deaths, the effective population sizes were ~95%, ~92%, and 35% of the total population sizes of RJ, SP and MG, respectively. Overall these numbers suggest that in MG cases and deaths have been reported only in a subset of the overall population, while in the other states SARS-CoV-2 appears widely dispersed. Incidence of reported cases per 100K using the effective population size was ~60 in SP, ~51 in RJ and ~7.85 in MG (Supplementary Figure

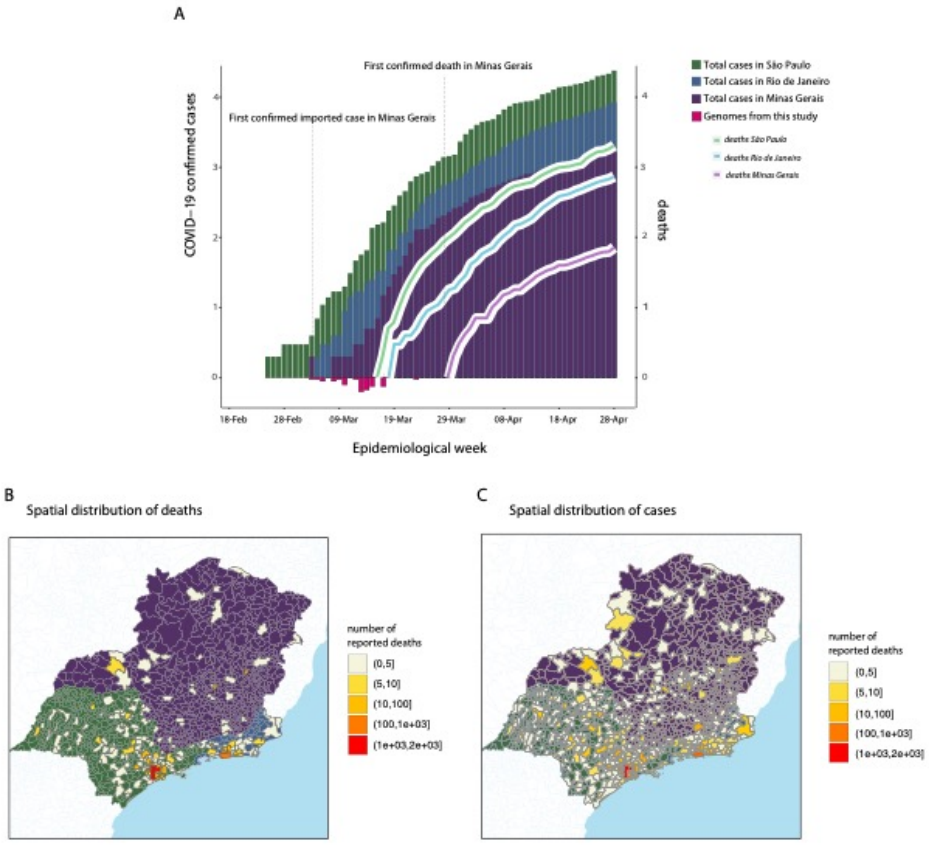


Figure 2. SARS-CoV-2 epidemic curve and spatial distribution of cases and deaths reported in the states of Minas Gerais (MG), São Paulo (SP) and Rio de Janeiro (RJ), Brazil. Panel A: Daily confirmed cases of COVID-19 in the state of MG. The X-axis represents the days from the first case in Brazil until 28 April 2020, while the Y-axis (left-hand side) represents the number of cases. The right-hand side of the Y-axis represents the number of deaths related to COVID-19. Y-axis numbers are represented as log10. Panels B and C: Maps with location (municipality) of (B) deaths and case (C) events, coloured by total number of reports. Different background colours highlight the boundaries of the three states: green for SP, purple for MG, blue for RJ.

S7), while incidence of deaths per 100K was ~5.56 in SP, ~4.69 in RJ and ~0.94 in MG (Supplementary Figure S12).

Phylogenetic analyses of 40 new SARS-CoV-2 sequences from Minas Gerais

In MG, samples from (clinically) suspected cases were screened at the Central Public Health Laboratory/Octávio Magalhães Institute (IOM) of the Ezequiel Dias Foundation (FUNED), which belongs to the public laboratories network of the Brazilian Ministry of Health (MoH). By 3rd April 2020, IOM/FUNED had performed 3,303 RT-qPCR tests for SARS-CoV-2 on swab samples from suspected cases. We used Nanopore sequencing to generate complete genomes from 40 COVID-19 patients in 15 different municipalities in MG (Table 1).

Of the 40 samples, 17 (42.5%) were from the state's capital (Belo Horizonte), while the other municipalities were represented by one or a maximum of three samples. These samples were from 17 females and 23 males, with collection dates ranging from 4 March 2020 to 26 March 2020 (Table 1). The median age of the patients was 35 years (ranging from 19-79 years old). The first sample that tested positive by RT-qPCR at IOM/FUNED was collected on 4th March 2020 (Table 1 and Figure 1). Selected samples had cycle threshold (Ct) values that ranged from 16.41 to 31.86 (median=22.945). We found no demographic variables (age, gender) to be statistically correlated with sample Ct (Supplementary Figure S17). The new sequences had a median genome coverage of 82.5% related to the reference genome NC_045512.3 (S1 Table). All sequences generated in this study have been submitted to the GISAID following WHO guidelines [38].

Table 1. Information on the 40 sequenced samples from Minas Gerais state.

Project-ID	Lab ID	Sample type	Ct value	Onset date	Collection date	Age	Sex	State	Municipality	Travel information
CV1	47/20	SWAB	20.54	29/02/20	04/03/20	38	F	MG	Ipatinga	Israel
CV2	115/20	SWAB	24.41	06/03/20	08/03/20	44	F	MG	Sete Lagoas	Portugal, Spain
CV3	135/20	SWAB	27.77	08/03/20	09/03/20	45	M	MG	Belo Horizonte	Italy, Switzerland, Austria, Portugal
CV4	242/20	SWAB	21.92	NA	09/03/20	65	M	MG	Juiz De Fora	USA
CV5	252/20	SWAB	29.93	12/03/20	12/03/20	32	M	MG	Belo Horizonte	..
CV6	298/20	SWAB	18.69	13/03/20	13/03/20	28	M	MG	Belo Horizonte	USA
CV7	352/20	SWAB	26.96	10/03/20	13/03/20	35	M	MG	Belo Horizonte	Switzerland, Portugal, England, Belgium, Spain
CV8	399/20	SWAB	22.61	06/03/20	13/03/20	34	M	MG	Belo Horizonte	..
CV9	428/20	SWAB	30.19	06/03/20	13/03/20	33	F	MG	Belo Horizonte	Sao Paulo (Brazil)
CV11	607/20	SWAB	27.92	10/03/20	16/03/20	40	F	MG	Mariana	Germany, Hungary, Czech Republic
CV12	615/20	SWAB	24.9	10/03/20	11/03/20	37	F	MG	Juiz De Fora	USA
CV13	660/20	SWAB	25.69	15/03/20	15/03/20	22	M	MG	Belo Horizonte	Italy
CV16	791/20	SWAB	20.64	15/03/20	16/03/20	52	M	MG	Belo Horizonte	..
CV17	809/20	SWAB	26.95	09/03/20	11/03/20	61	M	MG	Sete Lagoas	Portugal, Spain
CV18	833/20	SWAB	24.54	11/03/20	16/03/20	25	F	MG	Belo Horizonte	..
CV19	836/20	SWAB	22.04	11/03/20	16/03/20	22	M	MG	Belo Horizonte	Rio de Janeiro (Brazil)
CV20	838/20	SWAB	23.52	13/03/20	16/03/20	30	F	MG	Belo Horizonte	Colombia, Jamaica, Cayman Islands, Panama
CV21	842/20	SWAB	16.67	05/03/20	16/03/20	56	F	MG	Bom Despacho	-
CV22	895/20	SWAB	27.92	15/03/20	16/03/20	20	F	MG	Mariana	Germany
CV24	1028/20	SWAB	25.15	13/03/20	16/03/20	22	F	MG	Uberlândia	..
CV26	1078/20	SWAB	18.76	16/03/20	17/03/20	44	M	MG	Belo Horizonte	..
CV27	1166/20	SWAB	22.99	11/03/20	17/03/20	60	F	MG	Boa Esperança	..
CV28	1142/20	SWAB	22.21	13/03/20	17/03/20	46	F	MG	São João Del Rei	USA
CV31	1274/20	SWAB	17.38	16/03/20	17/03/20	35	M	MG	Betim	..
CV32	1290/20	SWAB	16.41	17/03/20	17/03/20	27	M	MG	Betim	..
CV33	1420/20	SWAB	18.79	14/03/20	17/03/20	35	M	MG	Sabara	..
CV34	1467/20	SWAB	22.31	16/03/20	18/03/20	48	F	MG	Belo Horizonte	..
CV35	1500/20	SWAB	24.06	07/03/20	18/03/20	75	M	MG	Poços De Caldas	Chile, Peru
CV36	1504/20	SWAB	24.07	18/03/20	18/03/20	50	F	MG	Muriae	..
CV40	1834/20	SWAB	23.97	18/03/20	19/03/20	29	M	MG	Belo Horizonte	..
CV41	1892/20	SWAB	22.84	16/03/20	18/03/20	20	F	MG	Serra Do Salitre	..
CV42	2119/20	SWAB	18.78	18/03/20	20/03/20	67	M	MG	São João Del Rei	..
CV43	2159/20	SWAB	24.81	14/03/20	17/03/20	19	F	MG	Patrocinio	..
CV44	2196/20	SWAB	23.47	17/03/20	18/03/20	19	F	MG	Patrocinio	..
CV45	2241/20	SWAB	22.85	14/03/20	20/03/20	58	M	MG	Muriae	Sao Paulo (Brazil)
CV46	2271/20	SWAB	22.9	19/03/20	20/03/20	35	M	MG	Belo Horizonte	..
CV47	2288/20	SWAB	22.4	17/03/20	19/03/20	35	M	MG	Belo Horizonte	..
CV48	2693/20	SWAB	22.43	19/03/20	20/03/20	74	M	MG	Varginha	..
CV49	2801/20	SWAB	20.95	16/03/20	22/03/20	30	M	MG	Belo Horizonte	..
CV50	5068/20	SWAB	31.86	20/03/20	26/03/20	44	M	MG	Mariana	..

Project-ID=sample identifier; Onset date= Symptoms onset date; F=Female; M=Male; MG=State of Minas Gerais; NA=Not Available.

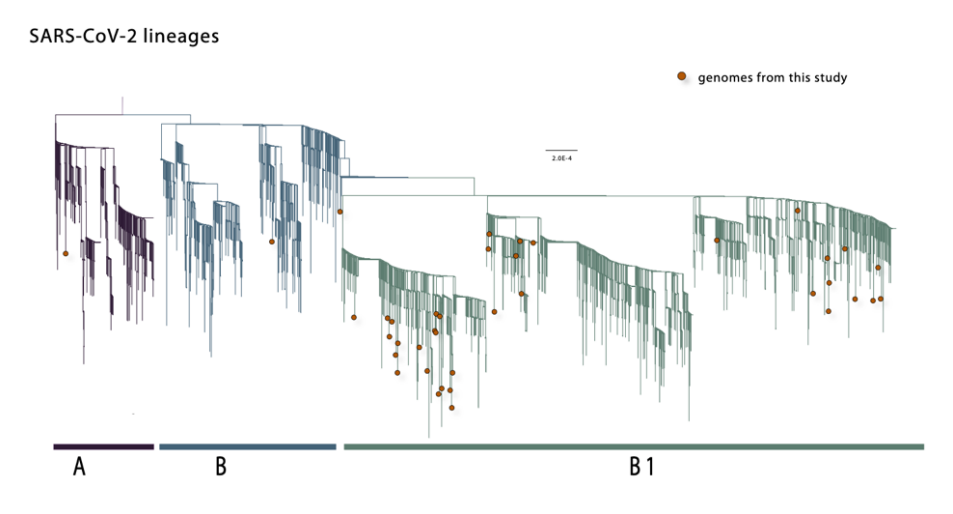


Figure 3. Phylogenetic analysis of the SARS-CoV-2 isolated in the state of MG, Brazil. Maximum likelihood phylogeny inferred using 40 genome sequences from SARS-CoV-2 generated in this study (marked with red circles) and 3062 sequences already deposited on the GISAID database. The tree is rooted between lineages A and B. Vertical branch lengths give the number of nucleotide substitutions per site.

Of the 17 (42.5%, n=40) sequenced cases with available travel history information, 14 (82.35%, n=17) reported international travel while three reported domestic travel (Table 1). Two among the later visited the city of São Paulo and one the city of Rio de Janeiro. Of the international travel-related cases, seven (50%) were linked to travel to European countries (Portugal, Spain, Italy, Switzerland, Austria, England, Belgium, Germany, Czech Republic, and Hungary), while six reported travel to countries in the Americas (USA, Colombia, Jamaica, Cayman Islands, Panama, Chile, and Peru). One case reported travel to Israel.

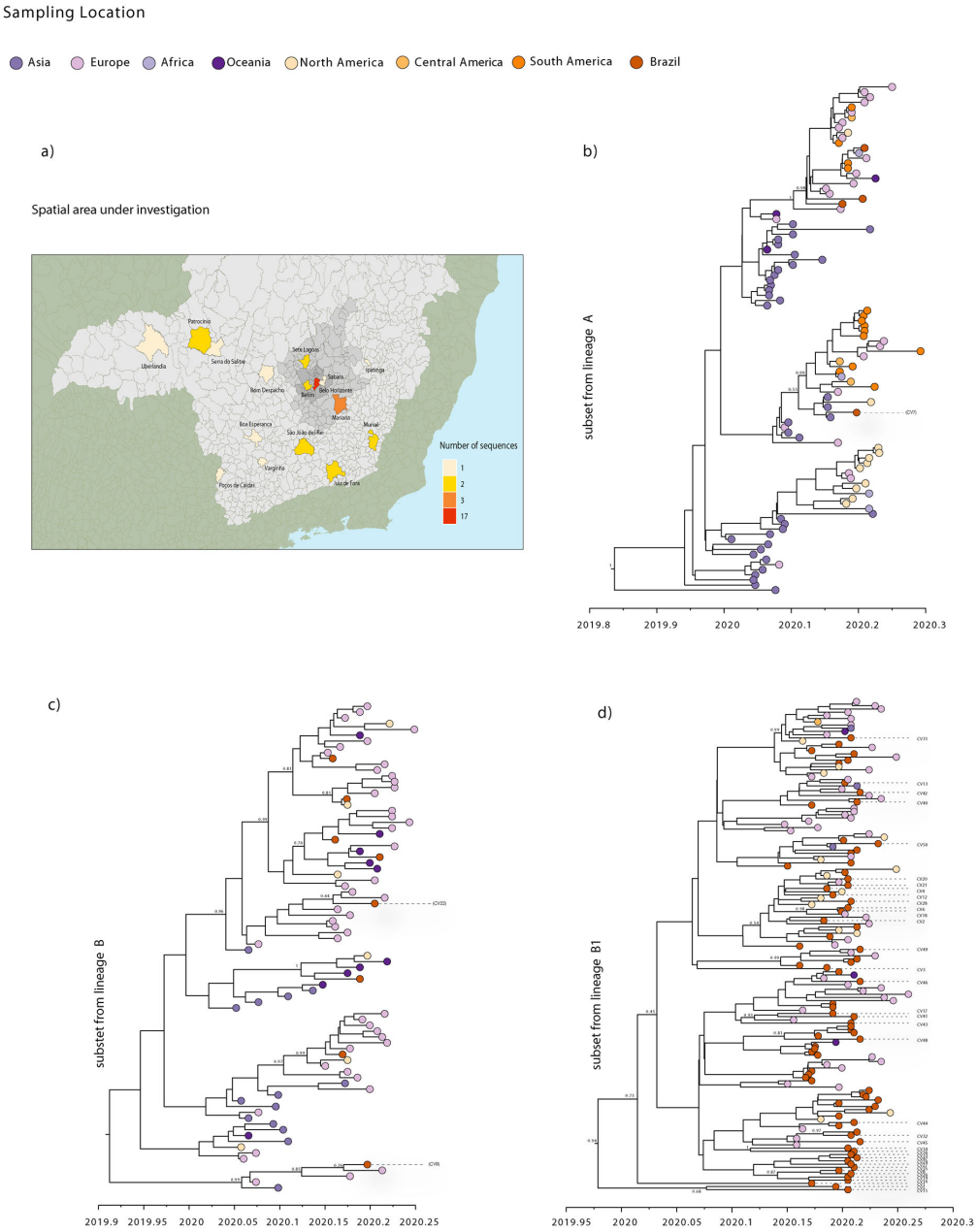


Figure 4. Bayesian analysis of the SARS-CoV-2 isolated in the state of MG, Brazil. (a) Map of the MG state showing the number of SARS-CoV-2 new sequences by patient´s municipality. (b) Molecular clock phylogeny of a subset of lineage A viruses, including one new sequence from MG. (c) Molecular clock phylogeny of a subset of lineage B viruses, including two new sequences from MG. (d) Molecular clock phylogeny of the subset from sublineage B.1, including 37 new sequences from MG. In all cases, the numbers along branches represent posterior probabilities while colours represent different sampling locations.

To explore the epidemiological history of the virus in MG, we performed a maximum likelihood (ML) phylogenetic analysis on the 40 new sequences combined with another 3,062 sequences deposited in GISAID up to 15 April 2020. We also made use of a publicly available tool for the automated lineage assignment within SARS-CoV-2 (https://github.com/hCoV-2019/pangolin) [23]. Our estimated phylogeny identified two major clades branching at the root of the tree (Figure 3). These two clades were named lineages A and B, following a recently proposed SARS-CoV-2 lineage nomenclature [39]: lineage A can be defined by the Wuhan/WH04/2020 sequence, while lineage B is represented by Wuhan-Hu-1. Lineage B can be further divided into a number of component sublineages, and our phylogenetic analysis revealed that the majority (n=37, 92.5%) of MG sequences belonged to lineage B.1 which contains sequences from a variety of countries including Australia, China, Canada, Malaysia, and the USA [40] (Figure 3, Table 1). Of these 37 B.1 sequences, 11 were isolates from cases that reported travel to European countries (CV2, CV3, CV11, CV13, CV17) or the Americas (CV4, CV6, CV12, CV20, CV28, CV35), in addition to the isolate CV1 from a traveller who returned from Israel. Aside from sublineage B.1, two sequences were assigned to lineage B in our ML phylogeny (isolates CV22 and CV36, with the former reporting travel to Germany), while one sequence could be assigned to lineage A (isolate CV7) who reported travel to European countries (full results from the Pangolin COVID-19 lineage assigner are presented in Supplementary Table S2).

To assess the evolution of these lineages in more detail and in time, we performed Bayesian time-measured phylogenetic analysis using a molecular clock model. We analysed three subsets of data (A, B, B.1) extracted from each lineage from the ML tree that included Brazilian sequences. Our maximum clade credibility (MCC) trees showed that most of MG's sequences were interspersed with viruses sampled from other countries (Figure 4b, c, d). This pattern, similar to that observed elsewhere [40–42], is also in accordance with our ML tree and with the epidemiological data, indicating that the appearance of these viruses were linked to travel exposure rather than community transmission, and reinforcing the idea that multiple independent introductions from overseas occurred in MG.

In contrast, some other SARS-CoV-2 sequences from MG grouped together, forming clusters that included sequences from Brazil and other countries: these are apparent in the data subset B.1 phylogeny (Figure 4d). However, these clusters have very low posterior probability support, likely due to the low genetic diversity of SARS-CoV-2 genomes [43–45]. Nonetheless, four clusters, each comprising only two MG sequences, showed posterior probabilities >80%. One of these clusters (Figure 4d), with a posterior probability of 100%, was formed by isolates CV34 and CV36 that arose from local contacts with a confirmed and a suspected COVID-19 case, respectively.

From the time-scaled phylogenies, we estimated the mean time of the most recent common ancestor (tMRCA) of the SARS-CoV-2 epidemic in Brazil to range from 27 January to 22 February 2020 (95% HPD interval; mean date of 10 February 2020). This is consistent with the start of reported cases in Brazil and with the epidemiological data from the first case confirmed in SP, regarding a traveller returning from Lombardy, Italy, on 21 February 2020 [11,13].

Discussion

We provide a preliminary characterization of the ongoing COVID-19 epidemic in MG, the second largest state in Brazil with respect to population size, where 1,578 cases were reported up to 28 April 2020. We report genomic data obtained by sequencing 40 SARS-CoV-2 infection cases confirmed by RT-qPCR in MG. We also collected epidemic data of case and death time series to estimate key epidemiological metrics such as the case fatality ratio, the observation rate and the effective reproduction number, R. Combined epidemiological and genomic data strongly suggest that the epidemic of SARS-CoV-2 infection in MG was associated with multiple independent importations from returned travellers.

Epidemiological analysis using case and death time series revealed that in MG, where case incidence was ~7.85, the distribution of cases and deaths was more spatially uniform (i.e. less focused on Belo Horizonte), while in SP and RJ states it was more centralized around capital cities, although with a wide variance. The estimated observation rate was 15.3% in MG, indicating that 1 case in 7 was reported in that state by 28 April 2020. Despite the different approaches used, the R values estimated for MG were similar to those estimated previously [46]. As all estimates of R were > 1, we conclude that virus transmission is not controlled and that the epidemic in MG will continue to grow unless stricter non-pharmaceutical interventions are implemented.

Our time-measured phylogeny revealed that MG's isolates are scattered throughout the tree, although some clusters suggestive of local transmission were also observed. This dispersed distribution has been observed in other studies such as from China [42] and New York [47], and is consistent with the self-declared travel history of patients to different countries in the Americas and Europe. From our phylogeny we inferred that multiple independent importations of SARS-CoV-2 infection from returned travellers had occurred in MG. This is noteworthy because it has

been argued that several introductions are necessary for an outbreak to establish in a new location [48]. The mean time of the most recent common ancestor of all Brazilian strains analysed in this study was estimated to be from 27 January to 22 February 2020, and range that includes the first COVID-19 confirmed case reported in 21 February 2020 [49]. These results illustrate that by combining genomic data and available epidemiological information we can generate more reliable phylogenetic inferences.

Despite the grouping of some MG sequences, because of the small sample size data which covers only about 30 days of the MG epidemic we cannot infer a close relationship between these sequences with certainty at this stage. In addition, the low genetic diversity of sequences available limits conclusions about SARS-CoV-2 directionality and spread based on genetic data alone. As noted elsewhere [43], such phylogenetic results should be approached with caution and considered as hypothesis-generating with respect to the transmission events of SARS-CoV-2 in a local setting.

In conclusion, at the end of April 2020, the COVID-19 epidemic in the state of MG was expanding (R>1) and highly geographically dispersed, with many cases and deaths reported away from the capital city. Genomic data and other epidemiological information from travel-related cases allowed us to identify several independent introductions in MG, helping to explain the geographical patchiness of reported cases and deaths. These initial insights based on the restricted data reveal that transmission is likely to continue in the near future and suggest there is considerable room to improve reporting. Increasing COVID-19 testing and SARS-CoV-2 genomic sequencing will undoubtedly help to understand how the virus is spreading in Brazil and thus greatly assist epidemic control.

Acknowledgments

We thank all the authors, originating and submitting laboratories that have kindly deposited and shared genome data on GISAID EpiCoV database, on which this research is based. An acknowledgment table can be found in Supplementary Table S3. We thank all personnel from Health Surveillance System from the state of MG that assisted with epidemiological data collection. We are also grateful for the support provided by the personnel from the Central Public Health Laboratory/Octávio Magalhães Institute (IOM) of the Ezequiel Dias Foundation (FUNED).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Pan American World Health Organization (VPGDI-003-FIO-19-2-2-30). M.G. is supported by Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ). V.F. and T.d.O. are supported by the South African Medical Research Council (MRC-RFA-UFSP-01-2013/UKZN HIVEPI) and the NIH H3AbioNet network, which is an initiative of the Human Health and Heredity in Africa Consortium (H3Africa). E.C.H. is supported by an Australian Research Council Australian Laureate Fellowship (FL170100022). J.L. is supported by a lectureship from the Department of Zoology, University of Oxford. J.X. and F.I. are supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. F.I. is also supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). A.S. has a scholarship from ZIKA - Announcement MCTIC/FNDCT-CNPq/MEC-CAPES/MS-Decit /No. 14/2016 - Prevention and Fight against Zika Virus. M.T.L. is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ). V.A. has a grant from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)(#88887.506611/202-00). The funders had no role in study design, data collection and analysis, writing and/or decision to publish the manuscript.

Data availability statement

SARS-CoV-2 genome sequences generated in this study have been deposited in the GISAID platform (https://www.gisaid.org/), accession numbers IDs EPI_ISL_429664 to EPI_ISL_429703.

ORCID

Joilson Xavier http://orcid.org/0000-0003-3916-2018
Vagner Fonseca http://orcid.org/0000-0001-5521-6448
Massimo Ciccozzi http://orcid.org/0000-0003-3866-9239

References

[1] World Health Organization (WHO). Novel coronavirus (2019-nCoV) Situation report - 1, 21 January 2020 [Internet]. New York: WHO Bull; 2020; Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4.
[2] Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet; 2020;395:565–574.
[3] Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579:265–269.
[4] Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020;382:727–733.
[5] Zhou P, Lou YX, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579:270–273.
[6] Zhang T, Wu Q, Zhang Z. Pangolin homology associated with 2019-nCoV. bioRxiv. 2020;1:1–23.
[7] Andersen KG, Rambaut A, Lipkin WI, et al. The proximal origin of SARS-CoV-2. Nat Med. 2020;26:450–452.
[8] World Health Organization (WHO). Coronavirus disease (COVID-19) Situation Report–165, 3 July 2020 [Internet]. 2020. Available from: https://www.who.int/

docs/default-source/coronaviruse/situation-reports/20200703-covid-19-sitrep-165.pdf?sfvrsn=b27a772e_2.
[9] Gorbalenya AE, Baker SC, Baric RS, et al. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol [Internet]. 2020;5:536–544. doi:10.1038/s41564-020-0695-z.
[10] World Health Organization (WHO). Coronavirus disease 2019 (COVID-19). Situation Report – 51 11 March 2020 [Internet]. 2020. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10.
[11] Ministério da Saúde. Boletim Epidemiológico Especial-14 / SE 18-26 de abril de 2020 COE-COVID19 [Internet]. 2020. Available from: https://portalarquivos.saude.gov.br/images/pdf/2020/April/27/2020-04-27-18-05h-BEE14-Boletim-do-COE.pdf.
[12] World Health Organization (WHO). Coronavirus disease (COVID-19) Situation Report – 103, 02 May 2020 [Internet]. 2020. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200502-covid-19-sitrep-103.pdf?sfvrsn=d95e76d8_4.
[13] Jaqueline Goes de J, Sacchi C, Claro I, et al. First cases of coronavirus disease (COVID-19) in Brazil, South America (2 genomes, 3rd March 2020) [Internet]. Virol. genome reports. 2020. p. 1. Available from: http://virological.org/t/first-cases-of-coronavirus-disease-covid-19-in-brazil-south-america-2-genomes-3rd-march-2020/409.
[14] Instituto Brasileiro de Geografia e Estatística - IBGE. Cidades e Estados: Minas Gerais [Internet]. 2020. Available from: https://www.ibge.gov.br/cidades-e-estados/mg.html.
[15] Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. Nat Rev Genet. 2018;19:9–20.
[16] Quick J. nCoV-2019 sequencing protocol. protocols.io; 2020.
[17] Quick J, Grubaugh ND, Pullan ST, et al. Multiplex PCR method for MinION and Illumina sequencing of zika and other virus genomes directly from clinical samples. Nat Protoc. 2017;12:1261–1276.
[18] Cleemput S, Dumon W, Fonseca V, et al. Genome detective coronavirus typing tool for rapid identification and characterization of novel coronavirus genomes. Bioinformatics. 2020;36:3552–3555.
[19] Vilsker M, Moosa Y, Nooij S, et al. Genome detective: an automated system for virus identification from high-throughput sequencing data. Bioinformatics. 2019;35:871–873.
[20] Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019;20:1160–1166.
[21] Laarson A. Aliview: a fast and lightweight alignment viewer and editor for large data sets. Bioinformatics. 2014;30:3276–3278.
[22] Nguyen LT, Schmidt HA, Von Haeseler A, et al. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–274.
[23] O'Toole Á, McCrone J. Phylogenetic Assignment of Named Global Outbreak LINeages [Internet]. 2020. Available from: https://github.com/hCoV-2019/pangolin.
[24] Rambaut A, Lam TT, Max Carvalho L, et al. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016;2:1–7.
[25] Suchard MA, Lemey P, Baele G, et al. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 2018;4:1–16.
[26] Griffiths RC, Tavaré S. Sampling theory for neutral alleles in a varying environment. Philos Trans R Soc Lond B Biol Sci. 1994;344:403–410.
[27] Rambaut A, Drummond AJ, Xie D, et al. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol. 2018;67:901–904
[28] Costa W. Confirmed cases and deaths of COVID-19 in Brazil, at municipal (city) level [Internet]. 2020. Available from: https://github.com/wcota/covid19br.
[29] Ministério da Saúde. Secretaria-Geral da Presidência da República - Imprensa Nacional: Portaria No 188, De 3 De Fevereiro De 2020 [Internet]. Brasília: Diário Of. da União; Available from: http://www.in.gov.br/en/web/dou/-/portaria-n-188-de-3-de-fevereiro-de-2020-241408388.
[30] Secretaria de Estado de Saúde de Minas Gerais. Boletim Epidemiológico COVID-19: Doença causada pelo coronavírus –19 28 de abril de 2020 [Internet]. 2020. Available from: https://saude.mg.gov.br/images/noticias_e_eventos/000_2020/mar_abr_maio/28.04.2020_Boletim_epidemiologico_COVID-19_MG.pdf.
[31] Instituto Brasileiro de Geografia e Estatística - IBGE. Cidades e Estados - Belo Horizonte [Internet]. 2020. Available from: https://www.ibge.gov.br/cidades-e-estados/mg/belo-horizonte.html.
[32] Spychalski P, Błażyńska-Spychalska A, Kobiela J. Estimating case fatality rates of COVID-19. Lancet Infect Dis. 2020;20:774–775.
[33] Russell TW, Hellewell J, Jarvis CI, et al. Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. Eurosurveillance. 2020;25:1–5.
[34] Verity R, Okell LC, Dorigatti I, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infect Dis. 2020;20:669–677.
[35] Wu JT, Leung K, Bushman M, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nat Med. 2020;26:506–510.
[36] Flaxman S, Mishra S, Gandy A, et al. Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. Imp Coll London. 2020.
[37] Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc R Soc B Biol Sci. 2007;274:599–604.
[38] Moorthy V, Henao M, Preziosi M, et al. Data sharing during the novel coronavirus public health emergency of international concern. Bull World Health Organ [Internet]. New York: 2020: 4–6. Available from: https://www.who.int/bulletin/online_first/20-251561.pdf.
[39] Rambaut A, Holmes EC, Hill V, et al. A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. bioRxiv. 2020.
[40] Nextstrain. Genomic epidemiology of novel coronavirus - Global subsampling [Internet]. 2020. Available from: https://nextstrain.org/ncov/global?branchLabel=clade&c=clade_membership&r=location&s=Canada/BC_64686/2020.
[41] Munnink BBO, Nieuwenhuijse DF, Stein M, et al. Rapid SARS-CoV-2 whole genome sequencing for

informed public health decision making in the Netherlands. bioRxiv. 2020.
[42] Lu J, du Plessis L, Liu Z, et al. Genomic Epidemiology of SARS-CoV-2 in Guangdong Province, China. Cell [Internet]. 2020 [cited 2020 May 22]; Available from: https://linkinghub.elsevier.com/retrieve/pii/S0092867420304864.
[43] Mavian C, Marini S, Prosperi M, et al. A snapshot of SARS-CoV-2 genome availability up to April 2020 and its implications (Preprint). JMIR Public Heal Surveill [Internet]. 2020 [cited 2020 May 22]; Available from: http://preprints.jmir.org/preprint/19170/accepted.
[44] Mavian C, Marini S, Manes C, et al. Regaining perspective on SARS-CoV-2 molecular tracing and its implications. medRxiv. 2020.
[45] Rambaut A. Phylodynamic Analysis | 176 genomes | 6 Mar 2020 [Internet]. virological.org. 2020. Available from: http://virological.org/t/phylodynamic-analysis-176-genomes-6-mar-2020/356.
[46] Mellan TA, Hoeltgebaum HH, Mishra S, et al. Report 21: Estimating COVID-19 cases and reproduction number in Brazil. medRxiv. 2020: 20096701.
[47] Gonzalez-Reiche AS, Hernandez MM, Sullivan M, et al. Introductions and early spread of SARS-CoV-2 in the New York city area. medRxiv. 2020.
[48] Kucharski AJ, Russell TW, Diamond C, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis. 2020;20:553–558.
[49] Jesus Jd, Sacchi C, Candido DdS, et al. Importation and early local transmission of COVID-19 in Brazil, 2020. Rev Inst Med Trop Sao Paulo. 2020;62:e30.

ANEXOS

Lista de artigos completos publicados em periódicos durante o Doutorado.

1. Adelino, Talita Émile Ribeiro; Giovanetti, Marta; Fonseca, Vagner; **Xavier, Joilson**; De Abreu, Álvaro Salgado; Do Nascimento, Valdinete Alves; Demarchi, Luiz Henrique Ferraz; Oliveira, Marluce Aparecida Assunção; Da Silva, Vinícius Lemes; De Mello, Arabela Leal E. Silva; Cunha, Gabriel Muricy; Santos, Roselene Hans; De Oliveira, Elaine Cristina; Júnior, Jorge Antônio Chamon; De Melo Iani, Felipe Campos; De Filippis, Ana Maria Bispo; De Abreu, André Luiz; De Jesus, Ronaldo; De Albuquerque, Carlos Frederico Campelo; Rico, Jairo Mendez; Do Carmo Said, Rodrigo Fabiano; Silva, Joscélio Aguiar; De Moura, Noely Fabiana Oliveira; Leite, Priscila; Frutuoso, Lívia Carla Vinhal; *et al.* Field and classroom initiatives for portable sequence-based monitoring of dengue virus in Brazil. Nature Communications, v.12, p.2296 - , 2021.

2. Giovanetti, Marta; Alcantara, Luiz Carlos Junior; Dorea, Alfredo Souza; Ferreira, Qesya Rodrigues; Marques, Willian De Almeida; Junior Franca De Barros, Jose; Adelino, Talita Emile Ribeiro; Tosta, Stephane; Fritsch, Hegger; Iani, Felipe Campos De Melo; Mares-Guia, Maria Angélica; Salgado, Alvaro; Fonseca, V. S.; **Xavier, J**.; Lopes, Elisson Nogueira; Soares, Gilson Carlos; Castro Amarante, Maria Fernanda De; Azevedo, Vasco; Kruger, Alícia; Correa Matta, Gustavo; Paineiras-Domingos, Laisa Liane; Colonnello, Claudia; Bispo De Filippis, Ana Maria; Montesano, Carla; Colizzi, Vittorio; Barreto, Fernanda Khouri. Promoting Responsible Research and Innovation (RRI) During Brazilian Activities of Genomic and Epidemiological Surveillance of Arboviruses. Frontiers In Public Health, v.9, p.1 - , 2021.

3. Slavov, Svetoslav Nanev; Giovanetti, Marta; Dos Santos Bezerra, Rafael; Fonseca, V. S.; Santos, Elaine Vieira; Rodrigues, Evandra Strazza; Adelino, Talita; **Xavier, J**.; Borges, Josiane Serrano; Evaristo, Mariane; De Carvalho Pereira, Glauco; Yamamoto, Aparecida Yulie; Clé, Diego Villa; Calado, Rodrigo Tocantins; Covas, Dimas Tadeu; Alcantara, Luiz Carlos Junior; Kashima, Simone. Molecular surveillance

of the on-going SARS-COV-2 epidemic in Ribeirao Preto City, Brazil. Infection Genetics And Evolution, v.93, p.104976 - , 2021.

4. Costa, Érica Azevedo; Giovanetti, Marta; Silva Catenacci, Lilian; Fonseca, V. S.; Aburjaile, Flávia Figueira; Chalhoub, Flávia L. L.; **Xavier, J.**; Campos De Melo Iani, Felipe; Da Cunha E Silva Vieira, Marcelo Adriano; Freitas Henriques, Danielle; Medeiros, Daniele Barbosa De Almeida; Guedes, Maria Isabel Maldonado Coelho; Senra Álvares Da Silva Santos, Beatriz; Gonçalves Silva, Aila Solimar; De Pino Albuquerque Maranhão, Renata; Da Costa Faria, Nieli Rodrigues; Farinelli De Siqueira, Renata; De Oliveira, Tulio; Ribeiro Leite Jardim Cavalcante, Karina; Oliveira De Moura, Noely Fabiana; Pecego Martins Romano, Alessandro; Campelo De Albuquerque, Carlos F.; Soares Feitosa, Lauro César; Martins Bayeux, José Joffre; Bertoni Cavalcanti Teixeira, Raffaella; Lisboa Lobato, Osmaikon; Da Costa Silva, Silvokleio; Bispo De Filippis, Ana Maria; Venâncio Da Cunha, Rivaldo; Lourenço, José; Alcantara, Luiz Carlos Junior. West Nile Virus in Brazil. Pathogens, v.10, p.896 - , 2021.

5. Giovanetti, Marta; Pereira, Luiz Augusto; Adelino, Talita Émile Ribeiro; Fonseca, Vagner; **Xavier, Joilson**; De Araújo Fabri, Allison; Slavov, Svetoslav Nanev; Da Silva Lemos, Poliana; De Almeida Marques, William; Kashima, Simone; Lourenço, José; De Oliveira, Tulio; Campelo De Albuquerque, Carlos Frederico; Freitas, Carla; Peterka, Cassio Roberto Leonel; Da Cunha, Rivaldo Venancio; Mendonça, Ana Flávia; Lemes Da Silva, Vinícius; Alcantara, Luiz Carlos Junior. A Retrospective Overview of Zika Virus Evolution in the Midwest of Brazil. Microbiology Spectrum, v.x, p.1 - , 2022.

6. Giovanetti, Marta; Pereira, Luiz Augusto; Santiago, Gilberto A.; Fonseca, V. S.; Mendoza, Maria Paquita García; De Oliveira, Carla; De Moraes, Laise; **Xavier, J**.; Tosta, Stephane; Fristch, Hegger; De Castro Barbosa, Emerson; Rodrigues, Evandra Strazza; Figueroa-Romero, Dana; Padilla-Rojas, Carlos; Cáceres-Rey, Omar; Mendonça, Ana Flávia; De Bruycker Nogueira, Fernanda; Venancio Da Cunha, Rivaldo; De Filippis, Ana Maria Bispo; Freitas, Carla; Peterka, Cassio Roberto Leonel; De Albuquerque, Carlos Frederico Campelo; Franco, Leticia; Méndez Rico, Jairo Andrés; Muñoz-Jordán, Jorge L.; Lemes Da Silva, Vinícius; Alcantara, Luiz Carlos Junior. Emergence of Dengue Virus Serotype 2 Cosmopolitan Genotype, Brazil. Emerging Infectious Diseases (Online), v.28, p.1725 - 1727, 2022.

7. Giovanetti, Marta; Slavov, Svetoslav Nanev; Fonseca, Vagner; Wilkinson, Eduan; Tegally, Houriiyah; Patané, José Salvatore Leister; Viala, Vincent Louis; San, Emmanuel James; Rodrigues, Evandra Strazza; Santos, Elaine Vieira; Aburjaile, Flavia; **Xavier, J**.; Fritsch, Hegger; Adelino, T. E. R.; Pereira, Felicidade; Leal, Arabela; Iani, Felipe Campos De Melo; De Carvalho Pereira, Glauco; Vazquez, Cynthia; Sanabria, Gladys Mercedes Estigarribia; Oliveira, Elaine Cristina De; Demarchi, Luiz; Croda, Julio; Dos Santos Bezerra, Rafael; Paola Oliveira De Lima, Loyze; *et al*. Genomic epidemiology of the SARS-CoV-2 epidemic in Brazil. Nature Microbiology, v.1, p.1 - 14, 2022.

8. Giovanetti, Marta; Fonseca, Vagner; Wilkinson, Eduan; Tegally, Houriiyah; San, Emmanuel James; Althaus, Christian L; **Xavier, Joilson**; Nanev Slavov, Svetoslav; Viala, Vincent Louis; Ranieri Jerônimo Lima, Alex; Ribeiro, Gabriela; Souza-Neto, Jayme A; Fukumasu, Heidge; Lehmann Coutinho, Luiz; Venancio Da Cunha, Rivaldo; Freitas, Carla; Campelo De A E Melo, Carlos F; Navegantes De Araújo, Wildo; Do Carmo Said, Rodrigo Fabiano; Almiron, Maria; De Oliveira, Tulio; Coccuzzo Sampaio, Sandra; Elias, Maria Carolina; Covas, Dimas Tadeu; Holmes, Edward C; Lourenço, José; Kashima, Simone; De Alcantara, Luiz Carlos Junior. Replacement of the Gamma by the Delta variant in Brazil: Impact of lineage displacement on the ongoing pandemic. Virus Evolution, v.8, p.1 - 1, 2022.

9. Fritsch, Hegger; Giovanetti, Marta; **Xavier, J**.; Adelino, Talita Emile Ribeiro; Fonseca, V. S.; De Jesus, Jaqueline Góes; De Jesus, Ronaldo; Freitas, Carla; Peterka, Cassio Roberto Leonel; Campelo De Albuquerque, Carlos Frederico; Bispo De Filippis, Ana Maria; Da Cunha, Rivaldo Venâncio; Silva, Erniria Carvalhais; Alcantara, Luiz Carlos Junior; Iani, Felipe Campos De Melo. Retrospective Genomic Surveillance of Chikungunya Transmission in Minas Gerais State, Southeast Brazil. Microbiology Spectrum, v.1, p.1 - , 2022.

10. Fritsch, Hegger; Pereira, Felicidade Mota; Costa, Erica Azevedo; Fonseca, V. S.; Tosta, Stephane; **Xavier, J**.; Levy, Flavia; Oliveira, Carla De; Menezes, Gabriela; Lima, Jaqueline; Santos, Lenisa; Silva, Luciana; Nardy, Vanessa; Astete, Marcela Kelly Gómez; Santos, Beatriz Senra Álvares Da Silva; Aguiar, Nágila Rocha; Guedes, Maria

Isabel Maldonado Coelho; Faria, Guilherme Canhestro De; Furtini, Ronaldo; Drumond, Safira Rachel Milanez; Cunha, Gabriel Muricy; Souza, Marcia São Pedro Leal; Jesus, Ronaldo De; Guimarães, Sara A. Franco; Nuno, Italo Coelho; Santana, Ian Carlos Brito De; Sá, José Eduardo Ungar De; Santos, George Roma; Silva, Willadesmon Santos; Guedes, Thiago Ferreira; Araújo, Emerson Luiz Lima; Said, Rodrigo Fabiano Do Carmo; Albuquerque, Carlos Frederico Campelo De; Peterka, Cassio Roberto Leonel; Romano, Alessandro Pecego Martins; Cunha, Rivaldo Venâncio Da; Filippis, Ana Maria Bispo De; Leal E Silva De Mello, Arabela; Giovanetti, Marta; Alcantara, Luiz Carlos Junior. Retrospective Investigation in Horses with Encephalitis Reveals Unnoticed Circulation of West Nile Virus in Brazil. Viruses, v.14, p.1540 - , 2022.

11. Castilho De Arruda, Larissa Domingues; Giovanetti, Marta; Fonseca, V. S.; Zardin, Marina Castilhos Souza Umaki; Lichs, Gislene Garcia De Castro; Asato, Silvia; Esposito, Ana Olivia Pascoto; Tokeshi Müller, Miriam**; Xavier, J**.; Fritsch, Hegger; Lima, Mauricio; De Oliveira, Carla; Santos, Elaine Vieira; Maziero, Livia De Mello Almeida; Frias, Danila Fernanda Rodrigues; Ahad Das Neves, Danielle; Ferreira Da Silva, Liliane; Rodrigues Barretos, Ellen Caroline; Tsuha Oshiro, Paulo Eduardo; Modafari Goday, Bianca; Lemos Dos Santos, Jéssica Klener; Kashima, Simone; Albuquerque, Carlos F. C. De; Said, Rodrigo Fabiano Do Carmo; Rosewell, Alexander; Demarchi, Luiz Henrique Ferraz; Croda, Julio; Alcantara, Luiz Carlos Junior; Cavalheiro Maymone Gonçalves, Crhistinne. Dengue Fever Surveillance in Mato Grosso do Sul: Insights from Genomic Analysis and Implications for Public Health Strategies. Viruses, v.15, p.1790 -, 2023.

12. Giovanetti, Marta; Pinotti, Francesco; Zanluca, Camila; Fonseca, Vagner; Nakase, Taishi; Koishi, Andrea C.; Tscha, Marcel; Soares, Guilherme; Dorl, Gisiane Gruber; Marques, Antônio Ernesto M.L; Sousa, Renato; Adelino, T. E. R.; **Xavier, J**.; De Oliveira, Carla; Patroca, Sandro; Guimaraes, Natalia Rocha; Fritsch, Hegger; Mares-Guia, Maria Angélica; Levy, Flavia; Passos, Pedro Henrique; Da Silva, Vinicius Leme; Pereira, Luiz Augusto; Mendonça, Ana Flávia; De Macêdo, Isabel Luana; Ribeiro De Sousa, Davi Emanuel; *et al*. Genomic epidemiology unveils the dynamics and spatial corridor behind the Yellow Fever virus outbreak in Southern Brazil. Science Advances, v.9, p.eadg9204, 2023.

13. Fritsch, Hegger; Moreno, Keldenn; Lima, Italo Andrade Barbosa; Santos, Cleiton Silva; Costa, Bernardo Grattival Gouvea; De Almeida, Breno Lima; Dos Santos, Ronald Alves; Francisco, Marcos Vinicius Lima De Oliveira; Sampaio, Maria Paula Souza; De Lima, Maricelia Maia; Pereira, Felicidade Mota; Fonseca, Vagner; Tosta, Stephane; **Xavier, J**.; De Oliveira, Carla; Adelino, T. E. R.; De Mello, Arabela Leal E Silva; Gräf, Tiago; Alcantara, Luiz Carlos Junior; Giovanetti, Marta; De Siqueira, Isadora Cristina. Phylogenetic Reconstructions Reveal the Circulation of a Novel Dengue Virus-1V Clade and the Persistence of a Dengue Virus-2 III Genotype in Northeast Brazil. Viruses, v.15, p.1073 -, 2023.

14. Giovanetti, Marta; Vazquez, Cynthia; Lima, Mauricio; Castro, Emerson; Rojas, Analia; De La Fuente, A G; Aquino, Carolina; Cantero, Cesar ; Fleitas, Fatima ; Torales, Juan; Barrios, Julio; Ortega, Mj; Gamarra, Maria Liz; Villalba, Shirley; Alfonzo, T; **Xavier, J.**; Adelino, T. E. R.; Fritsch, Hegger; Iani, F. C. M.; Pereira, G. C.; De Oliveira, Carla; Schuab, G.; Rodrigues, Evandra S.; Haddad, Simone Kashima; Leite, Juliana; Gresh, Lionel; Franco, Leticia; Tegally, Houriiyah; Van Voorhis, Wesley C.; Lessels, R.; De Filippis, Ana Maria Bispo; Oieda, A.; Sequera, Guillermo; Montoya, Romeo; Holmes, Edward C.; Oliveira, Tulio De; Rico, Jairo Mendez; Lourenco, J.; Fonseca, Vagner; Alcantara, L. C. J. Rapid Epidemic Expansion of Chikungunya Virus East/Central/South African Lineage, Paraguay. Emerging Infectious Diseases (Online), v.29, p.eid2909.230523 - , 2023.

15. Vazquez, Cynthia; Alcantara, Luiz Carlos Junior; Fonseca, Vagner; Lima, Mauricio; **Xavier, J**.; Adelino, T. E. R.; Fritsch, Hegger; Castro, Emerson; De Oliveira, Carla; Schuab, Gabriel; Lima, Alex Ranieri Jerônimo; Villalba, Shirley; Gomez De La Fuente, Andrea; Rojas, Analia; Cantero, Cesar; Fleitas, Fatima; Aquino, Carolina; Ojeda, Andrea; Sequera, Guillermo; Torales, Juan; Barrios, Julio; Elias, Maria Carolina; Iani, Felipe C. M.; Ortega, Maria Jose; Gamarra, Maria Liz; Montoya, Romeo; Rodrigues, Evandra Strazza; Kashima, Simone; Sampaio, Sandra Coccuzzo; Coluchi, Norma; Leite, Juliana; Gresh, Lionel; Franco, Leticia; Lourenço, José; Rico, Jairo Mendez; Bispo De Filippis, Ana Maria; Giovanetti, Marta. Retrospective Spatio-Temporal Dynamics of Dengue Virus 1, 2 and 4 in Paraguay. Viruses, v.15, p.1275 - , 2023.

16. Stoffella-Dutra, Ana Gabriella; De Campos, Bruna Hermine; Bastos E Silva, Pedro Henrique; Dias, Karolina Lopes; Da Silva Domingos, Iago José; Hemetrio, Nadja Simbera; **Xavier, J**.; Iani, Felipe; Fonseca, V. S.; Giovanetti, Marta; De Oliveira, Leonardo Camilo; Teixeira, Mauro Martins; Lobato, Zelia Ines Portela; Ferreira, Helena Lage; Arns, Clarice Weis; Durigon, Edison; Drumond, Betânia Paiva; Alcantara, Luiz Carlos Junior; De Carvalho, Marcelo Pires Nogueira; De Souza Trindade, Giliane. SARS-CoV-2 Spillback to Wild Coatis in Sylvatic-Urban Hotspot, Brazil. Emerging Infectious Diseases, v.29, p.664 - 667, 2023.

17. De Mello Almeida Maziero, Lívia De Mello Almeida; Giovanetti, Marta; Fonseca, V. S.; Zardin, Marina Castilhos Souza Umaki; De Castro Lichs, Gislene Garcia De Castro; De Rezende Romera, Grazielli Rocha De Rezende; Tsuha, Daniel Henrique; Frias, Danila Fernanda Rodrigues; Escandolhero, Valdir Castanho; Demarchi, Luiz Henrique; Domingues Castilho, Larissa Domingues; Barbosa, Karine Ferreira; Tebet, Danielle Galindo Martins; **Xavier, J**.; Fritsch, Hegger; Lima, Mauricio; Oliveira, Carla De; Santos, Elaine Vieira; Kashima, Simone; Said, Rodrigo Fabiano Do Carmo; Rosewell, Alexander; Croda, Julio; Alcantara, Luiz Carlos Junior; Gonçalves, Crhistinne Cavalheiro Maymon. Unveiling the Impact of the Omicron Variant: Insights from Genomic Surveillance in Mato Grosso do Sul, Midwest Brazil. Viruses, v.15, p.1604 -, 2023.