UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE CIÊNCIAS BIOMÉDICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM IMUNOLOGIA E PARASITOLOGIA APLICADAS

“Análise da variabilidade genômica de SARS-CoV-2 circulantes no estado de Minas Gerais”

GIULIA MAGALHÃES FERREIRA

Setembro, 2020

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE CIÊNCIAS BIOMÉDICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM IMUNOLOGIA E PARASITOLOGIA APLICADAS

“Análise da variabilidade genômica de SARS-CoV-2 circulantes no estado de Minas Gerais”

GIULIA MAGALHÃES FERREIRA

Dissertação apresentada ao colegiado do Programa de Pós-Graduação em Imunologia e Parasitologia Aplicadas como parte de obtenção do título de Mestre.

Orientadora: Ana Carolina Gomes Jardim

UBERLÂNDIA

Setembro – 2020

Ficha Catalográfica Online do Sistema de Bibliotecas
da UFU com dados informados pelo(a) próprio(a)
autor(a).

F383 Ferreira, Giulia Magalhães, 1995-
2020 Análise da variabilidade genômica de SARS-CoV-2 de variantes circulantes no estado de Minas Gerais [recurso eletrônico] : SARS-CoV-2 GENOME SEQUECING ANALYSIS IN
MINAS GERAIS, BRAZIL / Giulia Magalhães Ferreira. - 2020.

Orientadora: Ana Carolina Gomes Jardim.
Dissertação (Mestrado) - Universidade Federal de
Uberlândia, Pós-graduação em Imunologia e Parasitologia Aplicadas.
Modo de acesso: Internet.
Disponível em: http://doi.org/10.14393/ufu.di.2020.661 Inclui bibliografia.

1. Imunologia. I. Jardim, Ana Carolina

Bibliotecários responsáveis pela estrutura de acordo com o

AACR2: Gizele Cristine Nunes do Couto - CRB6/2091

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Coordenação do Programa de Pós-Graduação em Imunologia e Parasitologia Aplicada
Av. Amazonas, s/n, Bloco 4C, Sala 4C218 - Bairro Umuarama, Uberlândia-MG, CEP 38400-902
Telefone: (34) 3225-8672 - www.imunoparasito.ufu.br - coipa@ufu.br

ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Imunologia e Parasitologia Aplicadas				
Defesa de:	Dissertação de Mestrado número 266 do PPIPA				
Data:	um de outubro de dois mil e vinte	Hora de início:	14h30	Hora de encerramento:	17h43
Matrícula do Discente:	11822IPA002				
Nome do Discente:	Giulia Magalhães Ferreira				
Título do Trabalho:	Análise da variabilidade genômica de *SARS-CoV-2* de variantes circulantes no estado de Minas Gerais				
Área de concentração:	Imunologia e Parasitologia Aplicadas				
Linha de pesquisa:	Biologia das interações entre patógenos e seus hospedeiros				
Projeto de Pesquisa de vinculação:	Rede de pesquisa em doenças infecciosas humanas e animais no estado de Minas Gerais				

Reuniu-se, por vídeo conferência web, a Banca Examinadora designada pelo Colegiado do Programa de Pós-graduação em Imunologia e Parasitologia Aplicadas, assim composta pelos Titulares: Profa. Dra. Sabrina Royer– ICBIM/UFU; Prof. Dr. Alex Martins Machado– UFMS/MT e Profª Drª Ana Carolina Gomes Jardim – ICBIM/UFU (Presidente), orientadora da candidata.

Iniciando os trabalhos a presidente da mesa, Profª Drª Ana Carolina Gomes Jardim, apresentou a Comissão Examinadora e a candidata, agradeceu a presença do público, e concedeu a discente a palavra para a exposição do seu trabalho. A duração da apresentação da discente, o tempo de arguição e de resposta foram conforme as normas do programa.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovada.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.

Documento assinado eletronicamente por **Ana Carolina Gomes Jardim**, **Professor(a) do Magistério Superior**, em 01/10/2020, às 19:24, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Sabrina Royer**, **Professor(a) do Magistério Superior**, em 02/10/2020, às 13:59, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Alex Martins Machado**, **Usuário Externo**, em 02/10/2020, às 14:53, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2290709** e o código CRC **71E048B1**.

AGRADECIMENTOS

Agradeço a Deus, por estar sempre comigo, me guiado, e cuidado da minha vida, sem Ele não teria chegado aqui.

Agradeço aos meus pais Mário e Geisa e minha irmã Giovana pelo amor incondicional e pelo carinho mesmo em momentos que não mereci. Agradeço a eles por todo o apoio e paciência que tiveram comigo ao longo desses anos. Sem o apoio deles conquistar esse título não seria possível.

Agradeço à minha orientadora Carol por todos os ensinamentos, por ter investido no meu crescimento pessoal e profissional, por ter confiado em mim, e por ter me incentivado e buscado meios para que eu trabalhasse em uma linha de pesquisa que fosse do meu agrado. Muito obrigada por todos os conselhos de orientadora e as vezes de mãe\amiga.

Agradeço à toda equipe do CADDE, por terem me recebido tão bem em São Paulo, possibilitando que eu me sentisse em casa. Sou grata a todos os ensinamentos e momentos de descontração vividos no laboratório. Em especial, agradeço a Jaqueline Goes pela disposição em me ensinar e ajudar, por toda a calma, conversas e amizade.

Aos meus colegas do laboratório de virologia da UFU, agradeço pela amizade, pelas brincadeiras, pela ajuda e apoio que me deram, mesmo quando estive fisicamente distante. Essa equipe vale ouro. Com certeza minha experiência não seria a mesma sem vocês.

''Os que semeiam em lágrimas segarão com alegria. Aquele que leva a preciosa semente, andando e chorando, voltará, sem dúvida, com alegria, trazendo consigo os seus molhos.'' Salmos 126: 5,6

LISTA DE ABREVIATURAS

α-CoV	*alphacoronavirus*
β-CoV	*betacoronavirus*
d-CoV	*deltacoronavirus*
γ-CoV	*gammacoronavirus*
ACE-2	Enzima conversora de angiotensina
CoVs	Coronavírus
COVID-19	Doença do coronavírus
CTD	Domínio C-terminal
E	Proteína de Envelope
ELISA	Ensaio de imunoabsorção enzimática
HCoV-229E	Coronavirus humano 229E
HCoV-NL63	Coronavirus humano NL63
HCoV-OC43	Coronavirus humano OC43
HCoV-HK1U	Coronavirus humano HKU1
IgM	Imunoglobulina M
IgG	Imunoglobulina G
INP	Intervenção não farmacêutica
IL-2	Interleucina 2
IL-7	Interleucina 7
kDa	Kilodalton
M	Proteína de Membrana
MERS-CoV	Coronavírus da sindrome respiratória aguda do Oriente Médio
MG	Minas Gerais
N	Nucleocapsídeo
NTD	Domínio N terminal
nsp	non-structural proteins (Proteínas não estruturais)
OMS	Organização Mundial da Saúde
ORFs	Open Reading Frame (Regiões de leitura aberta)
PCR	Reação em cadeia da polimerase
RNA	Ribonucleicacid (Ácido ribonucleico
RBD	Domínio de ligação do receptor
S	Proteína spike

SUMÁRIO

RESUMO

O coronavírus da síndrome respiratória aguda grave 2 (SARS-CoV-2) foi identificado pela primeira vez em Wuhan-China, como o agente causador da doença do coronavírus 2019 (COVID-19). Desde o surgimento do primeiro caso em São Paulo, em 26 de fevereiro de 2020, mais de 3.000.000 de casos e 106.000 mortes foram notificados no Brasil. Na fase inicial da epidemia, o SARS-CoV-2 se espalhou localmente, no entanto, com o tempo, esse vírus se disseminou para outras regiões do país. Neste estudo, o sequenciamento do genoma do SARS-CoV-2 e análises filogenéticas foram realizados a partir de sequencias obtidas do processamento de 20 amostras clínicas com diagnóstico para COVID-19, provenientes de 9 cidades de MG, a fim de avaliar a caracterização genética das variantes virais circulantes no estado, de março a maio de 2020. As análises demonstraram a circulação da linhagem B.1 nos locais investigados e substituições nucleotídicas foram observadas nas regiões genômicas relacionadas a importantes estruturas virais. Além disso, as sequências geradas neste estudo agrupam filogeneticamente com isolados do estado de São Paulo, sugerindo uma rota de transmissão entre esses 2 estados. Alternativamente, grupos monofiléticos de sequências de MG e de outros estados ou país foram observados, indicando eventos de introdução independente do vírus. Esses resultados reforçam a necessidade de vigilância genômica para o entendimento da atual propagação de patógenos virais emergentes.

Palavras-chave: SARS-CoV-2, COVID-19, vigilância genômica, Minas Gerais, sequenciamento genômico , linhagem B1

ABSTRACT

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first identified in Wuhan-China, as the causative agent of the coronavirus disease 2019 (COVID-19). Since notified in São Paulo on 26th February 2020, more than 3,000,000 cases and 106,000 deaths were reported in Brazil. In the early epidemic phase, SARS-CoV-2 spread locally, however, over time, this virus was disseminated to other regions of the country. Here we performed genome sequencing and phylogenetic analysis of SARS-CoV-2 of 20 clinical samples of COVID-19 confirmed cases from 9 cities of MG, in order to evaluate the genetic characterization of circulating viral strains in the state from March to May 2020. Our analyses demonstrated the circulation of B.1 lineage isolates in the investigated locations and nucleotide substitutions were observed into the genomic regions related to important viral structures. Additionally, sequences generated in this study clustered with isolates from Sao Paulo state, suggesting a dissemination route between these 2 states. Alternatively, monophyletic groups of sequences from MG and other states or country were observed, indicating independent events of virus introduction. These results reinforce the need of genomic surveillance for understand the ongoing spread of the emerging viral pathogens.

Keyword: SARS-CoV-2, COVID-19, genomic surveillance, Minas Gerais, genome sequencing, B1 lineage

CAPÍTULO I

Fundamentação Teórica

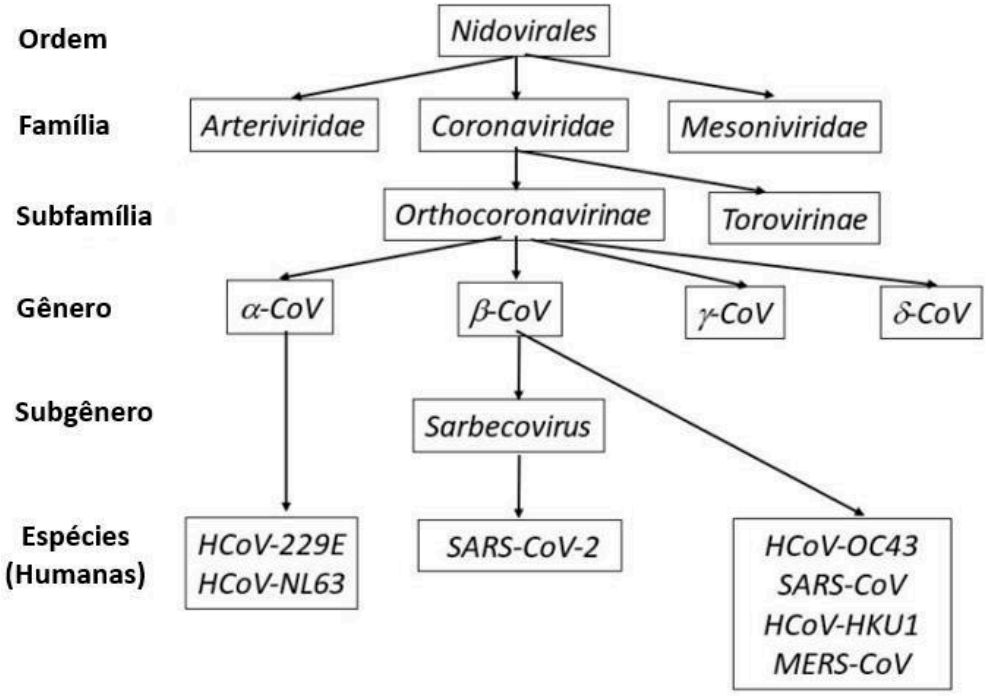
Coronavírus da síndrome respiratória aguda grave 2

INTRODUÇÃO

História e classificação dos Coronavírus.

Em 1965, Tyrell e Bynoe identificaram um microorganismo desconhecido em uma amostra de cultura de tecido traqueal de um paciente que apresentava um resfriado comum (TYRRELL; BYNOE, 1966). Após análises, esse novo grupo de vírus foi denominado como Coronavírus (CoVs), devido a aparência de coroa da sua superfície (TYRRELL et al., 1975). Os CoVs pertencem a ordem *Nidovirales*, família *Coronaviridae* e subfamília *Orthocoronavirinae*, e são classificados em quatro gêneros: alphacoronavirus (α-CoV), betacoronavirus (β-CoV), gammacoronavirus, (γ-CoV) e deltacoronavirus (d-CoV) (WOO et al., 2012) **(Figura 1)**.

Figura 1. Esquema de classificação dos Coronavírus humanos



Adaptado de (MALIK,2020)

Os CoVs são os agentes causadores de uma variedade de doenças letais em pássaros e mamíferos, fator que gera impacto na indústria agrícola (LEE, 2015). Alguns foram originalmente classificados como enzoóticos, limitando a infecção a animais, porém, progrediram para zoonoses, e atualmente também causam infecções em humanos

(MALIK, 2020). Até o momento, sete espécies conhecidas são zoonóticas, dentre elas, os CoVs humanos 229E (HCoV-229E) e NL63 (HCoV-NL63), pertencentes ao gênero alphacoronavirus, e os CoVs humano OC43 (HCoV-OC43), HKU1 (HCoV-HKU1), da síndrome respiratória aguda grave (SARS-CoV), da síndrome respiratória aguda grave 2 (SARS-CoV-2) e da síndrome respiratória do Oriente Médio (MERS-CoV), que pertencem aos betacoronavirus (VAN BOHEEMEN et al., 2012). Quatro dessas espécies (HCoV-229E, HCoVNL63, HCoV-OC43 e HCoV-HKU1) causam infecções em humanos, com sintomas similares a de um resfriado comum (Payne, 2017). No entanto, SARS-CoV, MERS-CoV e SARS-CoV-2 são capazes infectar o trato respiratório e desencadear uma série de sintomas como febre, tosses, pneumonia e dispneia, sendo que essas manifestações podem se agravar e levar o paciente ao óbito (LAI et al., 2020). Devido a possibilidade de apresentar manifestações mais severas e causar epidemias, essas últimas espécies tem despertado maior atenção da sociedade.

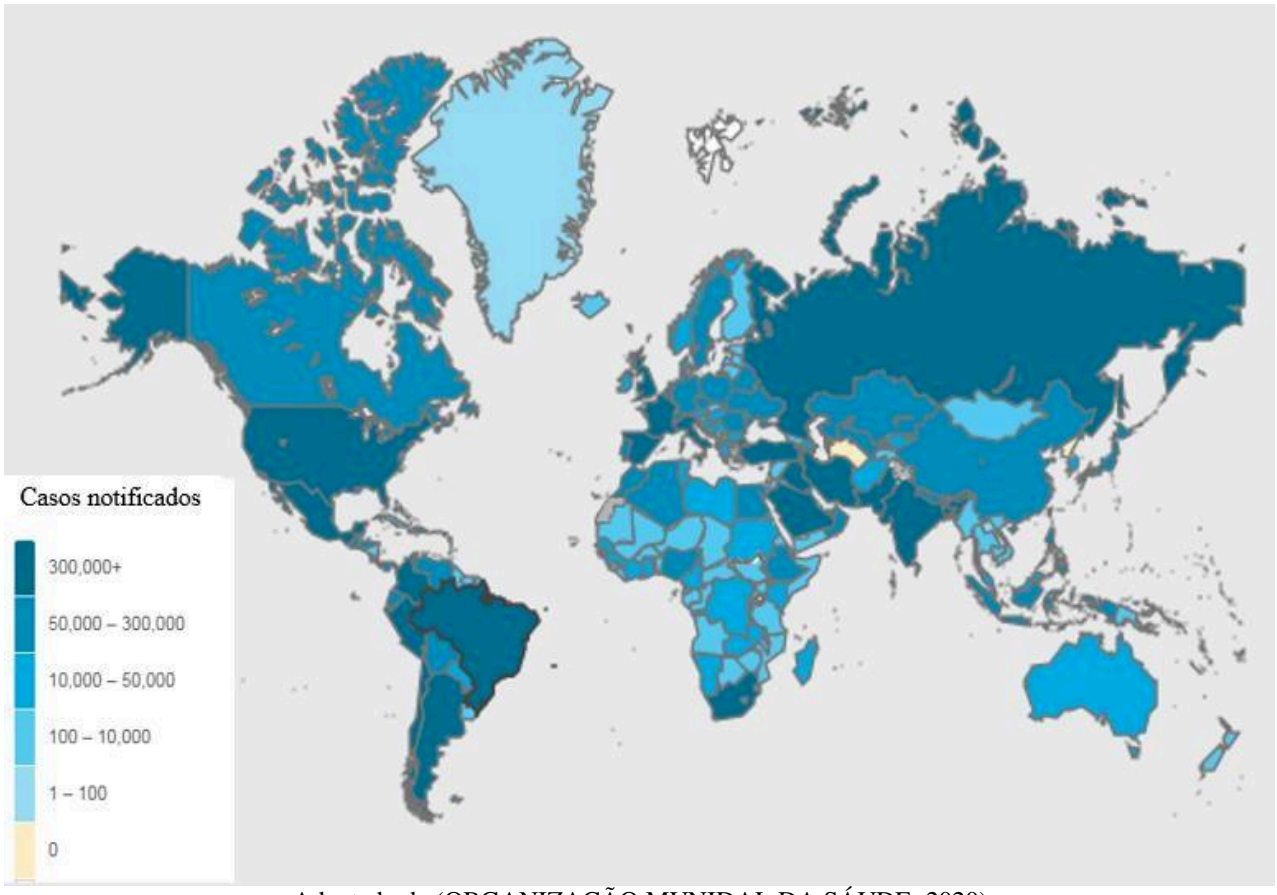
Entre os anos de 2002 e 2003, os primeiros casos da Síndrome Respiratória Aguda Grave (SARS, *Severe Acute Respiratory Syndrome*) ocorreram na China, causada pelo SARS-CoV (KSIAZEK et al., 2003). O vírus se espalhou rapidamente para países da América do Norte, América do Sul, Europa e Ásia. Entre 1 de novembro de 2002 e 8 de julho de 2003, mais de 8 mil casos e 812 mortes foram notificadas, entretanto, essa epidemia foi controlada (KSIAZEK et al., 2003; WOO et al., 2009). Ainda não se sabe ao certo como esse vírus passou a circular em humanos, já que os Civetas das palmeiras asiáticas eram os reservatórios naturais de SARS-CoV (GUAN et al., 2003).

Em 2012 foi isolado no Oriente Médio outro CoV com caráter epidêmico, distinto do SARS-CoV, o MERS-CoV (ZAKI et al., 2012). Este vírus foi o agente causador de infecções do trato respiratório com taxa de mortalidade entre 20 e 40% nas fases iniciais do surto, com preocupações de que esse CoV se disseminasse globalmente (AZHAR et al., 2019). Porém, embora casos esporádicos ocorreram por um período, o surto não se intensificou (MACKAY; ARDEN, 2015). Algum tempo após o primeiro relato, pesquisadores identificaram a presença de MERS-CoV em aproximadamente 90% dos Camelos da Arábia Saudita, sugerindo que esse animal possa ter sido o principal reservatório desse vírus (ALAGAILI et al., 2014; MACKAY; ARDEN, 2015).

Em 31 de Dezembro de 2019, a Organização Mundial da Saúde (OMS) foi reportada sobre a ocorrência de casos de pneumonia de etiologia desconhecida, que acometia a cidade de Wuhan, na província de Hubei, China (Organização Mundial da

Saúde, 2020). Em 11 de Fevereiro de 2020, a OMS nomeou essa nova pneumonia como Doença do Coronavírus 2019 (COVID-19, *Coronavirus Disease 2019*). Apesar de não ter sido originado a partir do SARS-CoV, o novo CoV foi nomeado pela Comissão Internacional de Classificação Viral como SARS-CoV-2 (CORONAVIRIDAE STUDY GROUP OF THE INTERNATIONAL COMMITTEE ON TAXONOMY OF, 2020), visto que o sequenciamento do genoma desse novo patógeno revelou maior similaridade SARS-CoV do que com o MERS-CoV (MALIK, 2020). Até 25 de setembro de 2020 foram relatados pela OMS mais de 30 milhões de casos e 970 mil mortes por COVID-19 no mundo (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2020) **(Figura 2)**. Devido a sua disseminação global e ao grande número de casos, em 11 de março de 2020 a OMS decretou situação de pandemia para o SARS-CoV-2. Devido à inexistência de antivirais licenciados contra o SARS-CoV-2 (SANTOS et al., 2020) vários países têm usado intervenções não farmacêuticas (INP) na tentativa de controlar a pandemia (CANDIDO et al., 2020) para combater o COVID-19. No entanto, apesar dessas medidas apresentarem efeitos positivos, estudos têm demonstrado que a transmissão do SARS-CoV-2 de indivíduos assintomáticos ou pré-sintomáticos dificulta os esforços da saúde pública para combater (ARONS et al., 2020; MIZUMOTO et al., 2020).

Figura 2. Número de casos notificados de SARS-CoV-2 no mundo



Adaptado de (ORGANIZAÇÃO MUNIDAL DA SÁUDE, 2020)

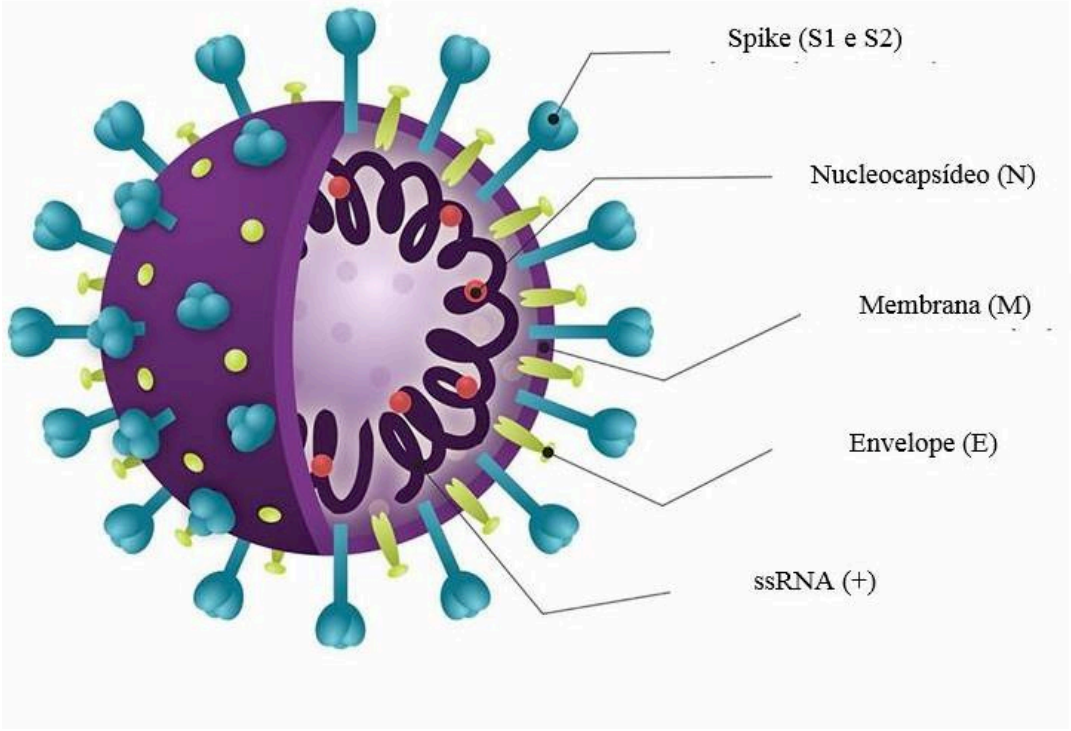
Origem do SARS-CoV-2

Desde a descoberta do SARS-CoV-2, a origem desse novo CoV vem sendo investigada (ANDERSEN et al., 2020; CIOTTI et al., 2020). A hipótese de que o SARS-CoV-2 seria resultado de procedimentos laboratoriais não foi sustentada, visto que os dados de similaridade genética entre o SARS-CoV-2 e outros Covs demostraram que o novo CoV não deriva de um vírus previamente conhecido (ALMAZAN et al., 2014). A análise do genoma viral demonstrou que o SARS-CoV-2 apresenta características genéticas especificas, como um sítio polibásico nas subunidades da proteína *spike* (S1 e S2,) permitindo que ocorra a clivagem por furinas e outras proteases. Essas proteases concedem uma maior afinidade ao receptor da enzima conversora de angiotensina 2 (ACE2), presente em algumas células do hospedeiro. Essa otimização da afinidade de ligação ao receptor da célula do hospedeiro distingue o SARS-CoV-2 dos outros coronavírus existentes (ANDERSEN et al., 2020; NAO et al., 2017). Foi observado que o SARS-CoV-2 é altamente semelhante aos CoVs de morcegos, podendo esse animal ser reservatório desse vírus (WU et al., 2020). Adicionalmente, outros CoVs de caráter epidêmico como o MERS-CoV (ALAGAILI et al., 2014) e o SARS-CoV (GUAN et al., 2003) apresentam animais como reservatório, fato que reforça a hipótese da origem desse novo vírus.

Estrutura da partícula viral do SARS-CoV-2

Assim com os membros da família *Coronaviridae*, o SARS-CoV-2 é um vírus envelopado, com genoma de RNA de fita simples com cerca de 30.000 pares de bases (MALIK, 2020). As proteínas *spike* (S), de membrana (M), de envelope (E) e do nucleocapsídeo (N) são os principais constituintes da estrutura do vírus (FEHR; PERLMAN, 2015; MALIK, 2020) **(Figura 3)**. O envelope viral é constituído por uma bicamada lipídica, derivada das membranas da célula hospedeira, onde estão inseridas as proteínas S, M e E. O nucleocapsídeo é formado por unidades da proteína N, que envolvem o RNA viral (FEHR; PERLMAN, 2015; MALIK, 2020) **(Figura 3)**.

Figura 3. Estrutura da partícula viral do SARS-CoV-2



(SANTOS et al., 2020)

Além de desempenharem um papel fundamental na estrutura da partícula do vírus, algumas proteínas estruturais também estão envolvidas em outros processos relacionados ao ciclo de replicação viral (MALIK, 2020). A proteína S (~150 kDa) auxilia a adsorção do vírus aos receptores de superfície da célula hospedeira, resultando na fusão da membrana celular ao envelope viral, com consequente entrada do vírus nas células (GLOWACKA et al., 2011; MALIK, 2020). Para evadir a ação dos anticorpos neutralizantes, alguns CoVs utilizam, no momento da infecção, da estratégia de fusão da célula infectada com as células adjacentes, formando células multinucleadas, estratégia mediada pela proteína S (GLOWACKA et al., 2011; QIAN; DOMINGUEZ; HOLMES, 2013). Uma protease do hospedeiro, semelhante a furina, cliva a proteína S em dois polipeptídeos, S1 e S2 (ABRAHAM et al., 1990; MALIK, 2020).

A proteína M (~25-30 kDa) define a forma do envelope viral, sendo a mais abundante dentre as proteínas citadas (ARMSTRONG et al., 1984; MALIK, 2020). Esta proteína se apresenta em duas conformações diferentes, permitindo a curvatura da membrana e, ao mesmo tempo, a ligação com o nucleocapsídeo, deixando este mais estável (NEUMAN et al., 2011). A associação entre as proteínas M e E compõem o

envelope viral, sendo essa interação suficiente para a produção e liberação de partículas semelhantes a vírus (VLPs, *Virus-Like Particles*) (MALIK, 2020).

Dentre as principais proteínas estruturais, a proteína E é a menor delas, apresentando entre 8 a12 kDa (NIETO-TORRES et al., 2011; WANG, H. et al., 2020). É uma proteína transmembranica que possui atividade de canal iônico (NIETO-TORRES et al., 2011; WANG, H. et al., 2020). Durante a replicação, a proteína E é abundantemente expressa dentro da célula infectada, sendo parte incorporada ao envelope viral, contribuindo para a montagem e brotamento do vírus (NIETO-TORRES et al., 2011). Estudos realizados com CoVs que não continham a proteína E demostraram que os títulos virais obtidos foram reduzidos, sugerindo a produção de progênie defectiva, e portanto, demonstrando a importância da proteína E na maturação e produção do vírus (ORTEGO et al., 2002)

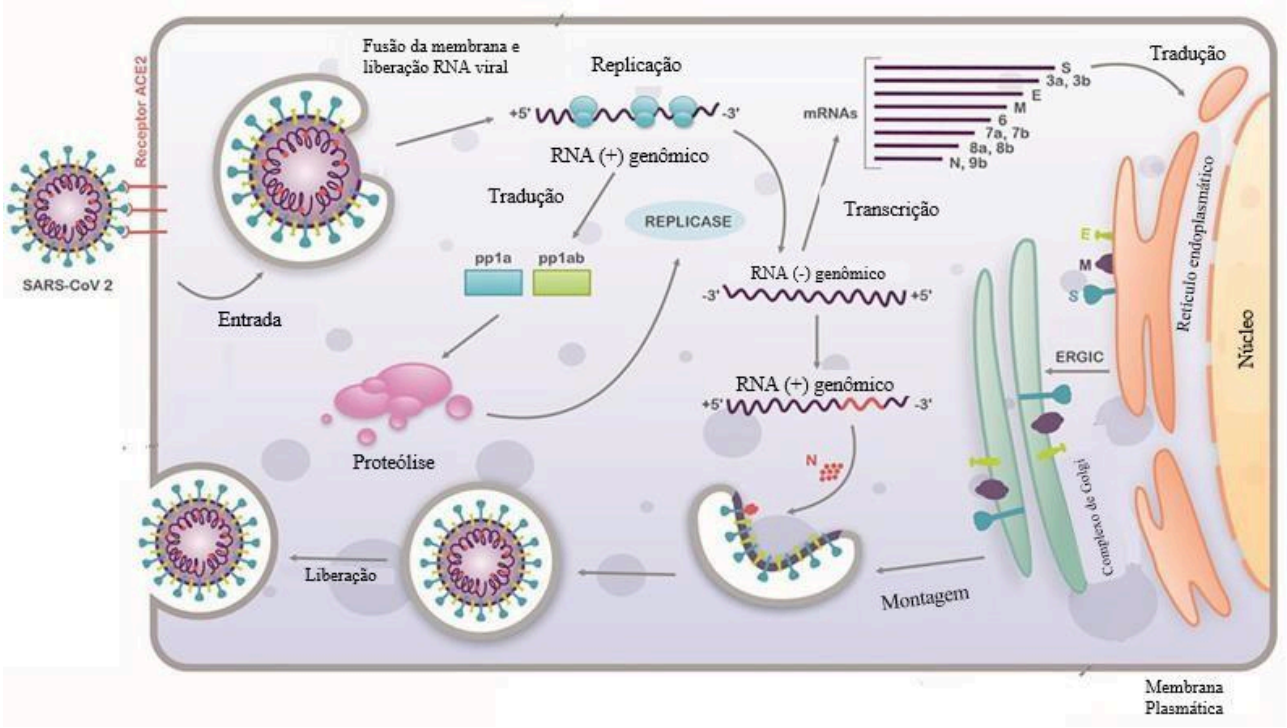
A única proteína que se liga ao genoma do SARS-CoV-2 é a proteína N (DE HAAN; ROTTIER, 2005), sendo composta por dois domínios separados, um domínio N-terminal (NTD) e um domínio C-terminal (CTD). Sugere-se que ambos contribuam para a ligação ideal ao RNA (CHANG et al., 2006). Essa proteína também está envolvida na montagem viral, resultando na formação do vírion (MALIK, 2020; TOOZE; TOOZE; WARREN, 1984).

Replicação do SARS-CoV-2

Devido à identificação recente do SARS-CoV-2, o processo de replicação no hospedeiro ainda não foi completamente elucidado. Os mecanismos gerais de entrada na célula hospedeira, replicação e liberação seguem características que foram descritos para outros CoVs e que foram parcialmente confirmados para SARS-CoV-2. A ligação do SARS-CoV-2 a célula hospedeira é iniciada pela interação entre a proteína viral S e o receptor celular ACE-2, processo relacionado ao tropismo do vírus pelos tecidos do hospedeiro que expressam receptor, como o pulmonar, gastrointestinal, hepático e o renal (CHU et al., 2020; HOFFMANN et al., 2020). A interação entre o ACE2 e o domínio de ligação ao receptor (RBD, *Receptor Binding Domain*) da proteína S desencadeia a endocitose da partícula viral, que é realizada através da formação de endossomos (RABI et al., 2020). A proteína S é dividida em duas subunidades, S1 e S2 (MALIK, 2020), e durante a endocitose a subunidade S2 é exposta, permitindo a fusão do envelope viral

com a membrana do endossomo e a consequente liberação do capsídeo no citoplasma celular (BELOUZARD; CHU; WHITTAKER, 2009) **(Figura 4)**.

Figura 4. Esquema representativo do ciclo replicativo do SARS-CoV-2



Adaptado de (SANTOS et al., 2020)

O ciclo de replicação segue para a próxima etapa, a de liberação do RNA viral. O RNA do SARS-CoV-2, que possui sentido positivo, é traduzido a partir de duas regiões de leitura aberta (ORFs, *Open reading frame*), ORF1a e ORF1b. A poliproteína pp1a é codificada pela ORF1a e posteriormente clivada em 11 proteínas não-estruturais (nsps), enquanto o ORF1b codifica a poliproteína pp1ab, que é clivada em 15 nsps (KIM et al., 2020). O complexo replicase-transcriptase é responsável pela replicação e transcrição de nove RNAs subgenômicos (sgRNAs), que atuam como mRNAs de genes acessórios e estruturais (FEHR; PERLMAN, 2015). As proteínas estruturais S, E e M são direcionadas para um compartimento intermediário entre o reticulo endoplasmático e o complexo de Golgi, onde os genomas virais são encapsulados pelas proteínas N e posteriormente associados a outras proteínas estruturais, formando os vírions (FEHR; PERLMAN, 2015; LI, X. et al., 2020). Após as proteínas M, E e N se ligarem e as proteínas S serem incorporadas, os vírions são transportados para a superfície da célula e liberados via exocitose (KIM et al., 2020) **(Figura 4)**.

Variabilidade genômica do SARS-CoV-2

Os vírus de RNA utilizamdurante o processo replicativo RNA polimerases virais que não apresentam atividade corretiva, tornando frequente a ocorrência de mutações e/ou eventos de recombinação (DUFFY, 2018; UDDIN et al., 2020). Este fato pode desempenhar um papel importante na evolução do SARS-CoV-2. Em relação à variabilidade do genoma viral, análises filogenéticas demonstraram que as variantes de SARS-CoV-2 se agruparam em dois subtipos principais, designados como L e S, que são diferenciados por polimorfismo de nucleotídeo, sendo que o tipo L foi encontrado em aproximadamente 70% dos casos e foi observado como sendo a forma mais agressiva e contagiosa em comparação ao subtipo S (TANG et al., 2020; UDDIN et al., 2020). Análises filogenéticas adicionais demonstraram que o vírus parece estar se distinguindo em 3 grupos, sendo o A e C encontrado principalmente nas Américas e na Europa e o B na Ásia Oriental (FORSTER et al., 2020).

Com o intuito de padronizar a nomenclatura das linhagens de SARS-CoV-2, mais de 23 mil sequencias do genoma viral foram avaliados e resultaram na identificação de 3 grupos principais: A, B e B.1 (RAMBAUT et al., 2020). Dentre esses grupos, seis linhagens derivam da linhagem A (A.1 a A.6) e duas de A.1 (A.1.1 e A.3) (RAMBAUT et al., 2020). Também foram descritas 16 linhagens derivadas diretamente do grupo B (B.1 a B.16) (RAMBAUT et al., 2020). Até o momento, B.1 é a linhagem predominante globalmente e foi subdividida em aproximadamente 70 sublinhagens. A linhagem B.2 apresenta seis sublineagens descendentes (RAMBAUT et al., 2020). Essa variabilidade genética viral ocorre provavelmente devido as grandes introduções de casos importados, seguidos por aumento exponencial na transmissão local (RAMBAUT et al., 2020).

Estudos recentes desenvolvidos no Brasil mostraram que as sequências de SARS-CoV-2 se agruparam em 3 clados, que se caracterizam por grupos de sequencias com diferentes trocas nucleotídicas e regiões de maior circulação. Como exemplos: i) sequencias agrupadas no clado 1 apresentam uma substituição de guanina por timina na região codificante da proteína S (G25088T) e são mais encontradas na região de São Paulo; ii) sequências agrupadas no clado 2 apresentam duas substituições nucleotídicas (T27299C e T28148C), referentes a ORF6 e a nucleoproteína, respectivamente, e estão distribuídas ao longo do país; iii) sequências agrupadas no clado 3 são mais encontradas no Ceará (CANDIDO et al., 2020). Como ancestral comum mais recente, os clados apresentam sequências datadas a partir de 28 de fevereiro (clado1), 22 de fevereiro

(clado2) e 11 de março (clado3). Esses dados sugerem que a transmissão entre regiões já havia sido estabelecida no início de março (CANDIDO et al., 2020).

Epidemiologia SARS-CoV-2 no Brasil

No Brasil, o primeiro caso confirmado de COVID-19 foi notificado no estado de São Paulo (SP) em 26 de fevereiro de 2020 (JESUS et al., 2020). Análise das duas primeiras sequências do genoma completo do SARS-CoV-2, isoladas de pacientes brasileiros que haviam retornado recentemente da Itália, demonstraram duas introduções independentes do vírus no país (JESUS et al., 2020). Durante a fase inicial da epidemia, o SARS-CoV-2 foi disseminado localmente, no entanto, apesar das intervenções para prevenir a dispersão do vírus ao longo do tempo, grandes centros urbanos se tornaram responsáveis pela disseminação do vírus para outras localidades. De acordo com estimativas recentes, mais de 100 introduções internacionais de vírus foram observadas no Brasil (CANDIDO et al., 2020). Atualmente, o SARS-CoV-2 é responsável por causar 4.529.201 de casos de COVID-19 e mais de 106.000 mortes no Brasil (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2020).

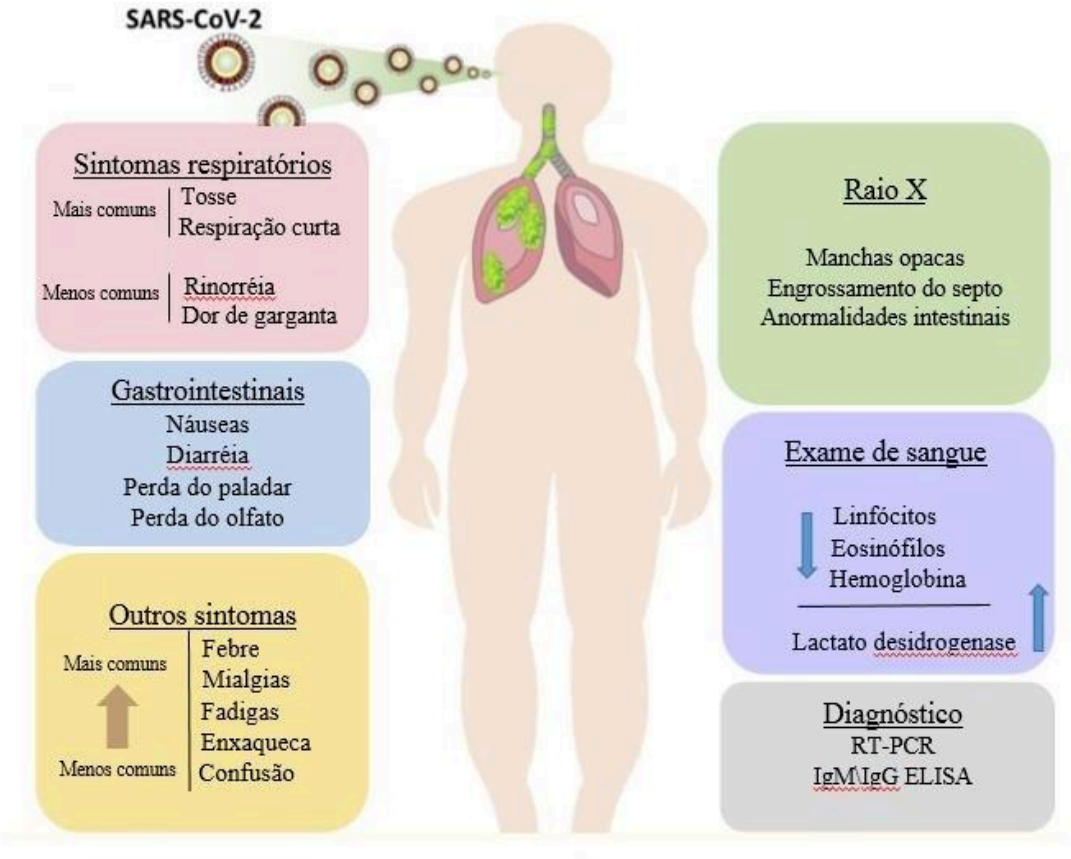
Minas Gerais (MG) é o segundo estado mais populoso e o quarto em relação a área do Brasil, além de ser o terceiro estado que mais contribui para o crescimento do Produto Interno Bruto (PIB) do país (INSTITUTO BRASILEIRO DE GEOGRAFIA E ECONOMIA, 2020). MG faz fronteira com SP, Bahia, Rio de Janeiro, Goiás e Mato Grosso do Sul (INSTITUTO BRASILEIRO DE GEOGRAFIA E ECONOMIA, 2020). Belo Horizonte, a capital de MG, é o maior centro urbano e financeiro da América Latina (INSTITUTO BRASILEIRO DE GEOGRAFIA E ECONOMIA, 2020). Devido ao seu grande tamanho populacional, situação econômica e acesso facilitado a outros estados importantes para a economia, MG representa uma região de disseminação da SARS-CoV-2, com potencial para agravar a situação (XAVIER et al., 2020). Até o momento, 278.701 casos e 6.983 mortes de COVID-19 foram relatados em MG (MINISTÉRIO DA SAÚDE. BOLETIM EPIDEMIOLÓGICO DE MINAS GERAIS).

A doença do Coronavírus 2019 e a transmissão do SARS-CoV-2 entre humanos

Pacientes com COVID-19 podem apresentar diferentes condições clinicas, que podem variar de assintomáticas a graves (CIOTTI et al., 2020). Os sintomas incluem

febre, tosse, falta de ar, corrimento nasal (rinorréia) (WANG et al., 2020), vômito, diarreia e dor abdominal, além da perda do paladar e olfato (CHEN, N. et al., 2020). Por causa da baixa oxigenação no sangue, alguns pacientes podem apresentar confusão mental, irritabilidade, agitação, tonturas e desmaios, além de pneumonia acompanhada de uma inflamação exacerbada do tecido, podendo levar a morte (CIOTTI et al., 2020; WANG et al., 2020)**.** Pacientes infectados com SARS-CoV-2 geralmente apresentam na radiografia de tórax múltiplas manchas opacas e aumento do septo (WANG et al., 2020; ZHU et al., 2020). Apesar de o principal alvo desse CoV ser o pulmão, a ampla distribuição de receptores ACE2 em outros órgãos pode levar ao desenvolvimento de doenças cardiovasculares, gastrointestinais, renais, hepáticas e danos oculares (CIOTTI et al., 2020; HAMMING et al., 2004).A avaliação dos exames de sangue desses pacientes demonstra uma queda na contagem de linfócitos e eosinófilos, valores de hemoglobina mais baixos, e o aumento da contagem de neutrófilos e lactato desidrogenase (LIPPI; PLEBANI, 2020) **(Figura 5)**.

Figura 5. Esquema dos sintomas e alterações causadas pela COVID-19



Adaptado de (TU et al., 2020)

Apesar do órgão alvo principal desse coronavírus ser o pulmão, a ampla distribuição de receptores ACE2 em outros órgãos pode levar ao desenvolvimento de doenças cardiovasculares, gastrointestinais, renais, hepáticas e danos oculares (CIOTTI et al., 2020; HAMMING et al., 2004).Alguns pacientes apresentam uma repentina piora do quadro clínico, muitas vezes evoluindo a óbito, devido a falência dos órgãos (CIOTTI et al., 2020). Geralmente esses pacientes apresentam a "tempestade de citoquinas" (*cytokine storm*) (WANG et al., 2020). Esse processo se dá pelo aumento de vários componentes do sistema imune, como as interleucinas IL-2 e IL-7, os granulócitos e fatores de necrose tumoral e estimulador de colônia (HUANG et al., 2020). Além disso, níveis elevados de ferritina e IL-6 levam a um processo de hiperinflamação induzida pelo vírus (RUAN et al., 2020).

A infecção pode ser caracterizada como sintomática, pré-sintomática ou assintomática (REF). A infecção sintomática é definida pela presença de sintomas, sendo o principal modo de transmissão entre os indivíduos infectados e não infectados (MALIK, 2020). Infecção pré-sintomática ocorre quando o indivíduo ainda não apresentou sintomas, porém, a disseminação viral acontece. Casos documentados de transmissão em pacientes pre-sintomaticos sugerem que a carga viral é suficiente para permitir a transmissão mesmo antes do aparecimento dos sintomas (MALIK, 2020; WEI et al., 2020). Por fim, a infecção assintomática ocorre durante o período de incubação viral, que pode durar até 14 dias (MALIK, 2020).

Assim como acontece com outros vírus respiratórios, a transmissão do SARS-CoV-2 ocorre principalmente pela via respiratória (HAN et al., 2020; LEUNG et al., 2020), que apresenta alta eficácia e infectividade (LEUNG et al., 2020). A infecção se dá quando há contato com gotículas respiratórias expelidas por indivíduos infectados ou por contato com objetos e superfícies contaminadas (LI, Q. et al., 2020; LIU et al., 2020). A via fecal-oral pode ser uma possível rota de transmissão, já que há relatos da presença de RNA de SARS-CoV-2 em amostras de fezes de pacientes com COVID-19 (HOLSHUE et al., 2020).

Diagnóstico da COVID-19

Para detecção do SARS-CoV-2, as primeiras abordagens escolhidas são as que envolvem a detecção molecular do vírus, como o teste da reação em cadeia da polimerase

(PCR), que é considerada a técnica de padrão ouro no diagnóstico desse patógeno (AHN et al., 2020; CORMAN et al., 2020). A OMS recomenda que as amostras de pacientes com suspeita de SARS-CoV-2 utilizadas para o teste molecular sejam de isolados de espécimes do trato respiratório (incluindo nasal e esfregaços faríngeos, escarro ou lavado broncoalveolar) (AHN et al., 2020). A sequência isolada do primeiro caso SARS-CoV-2 foi utilizada para produção de *primers* e sondas específicos para detectar o SARSCoV-2, os quais usam como alvos as regiões virais das ORFs 1a e 1b, E e N (AHN et al., 2020; CORMAN et al., 2020).

Testes sorológicos que detectam anticorpos anti- SARS-CoV-2 também são métodos diagnósticos importantes com um curto tempo de resposta para o diagnóstico da COVID-19 (FEHR; PERLMAN, 2015). Diversos testes sorológicos vem sendo desenvolvidos e comercializados desde o início da pandemia, como o ensaio de imunoabsorção enzimática (ELISA) e ensaios de quimiluminescência, dentre eles, o teste de ELISA é o mais utilizado (CHEN, Y. et al., 2015; CIOTTI et al., 2020). Os testes de ELISA podem detectar anticorpos da fase ativa da infecção (IgM) e\ou da fase pós infecção (IgG) (AHN et al., 2020; CHEN, Y. et al., 2015), geralmente utilizando antígenos de proteínas recombinantes das proteínas S e N (CIOTTI et al., 2020).

Estudos comparativos entre o teste de RT-PCR e o teste de ELISA demostraram que, quando utilizados em amostras coletadas na primeira semana de sintomas, o PCR se mostrou 66.7% mais sensível do que o teste de ELISA (ZHAO et al., 2020). Entretanto, a sensibilidade dos testes imunoenzimáticos aumentaram quando aplicados nas amostras coletadas no período de 8-12 dias (ZHAO et al., 2020).

Importância da vigilância genômica

A avaliação dos genomas virais tem potencial para fornecer informações valiosas sobre a evolução contínua e epidemiológica do vírus durante pandemias e epidemias e desempenham um papel importante na vigilância e no controle dessas doenças. Caso haja uma incapacidade de diagnosticar infecções empiricamente, devido à velocidade das epidemias ou a falta de insumos para realizar os testes, a epidemiologia genômica pode ser usada para estimar a taxa de replicação do vírus na população, bem como carga de infecção, permitindo aos profissionais de saúde alcançar decisões para desacelerar o contágio (UDDIN et al., 2020). Um exemplo disso é a sugestão que Candido e

colaboradores (2020) fizeram sobre o início da dispersão de SARS-CoV-2 no Brasil. Segundo dados obtidos nesse estudo, quando a medida que restringia viagens internacionais foi estabelecida, o vírus já havia se disseminado no país (CANDIDO et al., 2020).

O sequenciamento é o passo inicial para a continuidade das pesquisas que visam o tratamento ou a prevenção das infecções. As informações genéticas desses vírus podem auxiliar no desenvolvimento de fármacos e/ou vacinas. Além disso, é possível saber qual a rota de transmissão viral através do sequenciamento genômico e análises filogenéticas, epidemiológicas e de filogeografia. Sendo assim, comparando vírus de populações diferentes, se torna possível traçar sua rota de dispersão e buscar meios de evita-las. O monitoramento contínuo das eventuais alterações que esses vírus possam sofrer se faz necessário com o objetivo de compreender os rumos da epidemia e assim evitar os problemas causados por sua disseminação.

OBJETIVOS

O presente trabalho teve como objetivo avaliar a variabilidade genômica de SARS-CoV-2 de amostras coletadas em cidades localizadas ao sul do estado de Minas Gerais, a fim de investigar as variantes circulantes nessas regiões.

Objetivos específicos

- Sequenciar amostras de SARS-CoV-2 provenientes de regiões de Minas Gerais com poucas ou nenhuma sequência depositada no banco de dados, com o objetivo de avaliar quais linhagens de SARS-CoV-2 estão circulando nos municípios localizados ao sul do estado;
- Investigar a presença de possíveis substituições nucleotídicas nas variantes circulantes dessas regiões;
- Avaliar o padrão de agrupamento das sequências geradas em relação a sequências referentes a amostras coletadas no estado de MG, e de outras localidades do Brasil e do mundo
- Analisar o estabelecimento e propagação do SARS-CoV-2 em alguns municípios ao sul de Minas Gerais, para o melhor entendimento da circulação das variantes e linhagens do vírus.

REFERÊNCIAS BIBLIOGRAFICAS

ABRAHAM, S. et al. Deduced sequence of the bovine coronavirus spike protein and identification of the internal proteolytic cleavage site. Virology, v. 176, n. 1, p. 296-301, May 1990.https://doi.org/10.1016/0042-6822(90)90257-R

AHN, D. G. et al. Current Status of Epidemiology, Diagnosis, Therapeutics, and Vaccines for Novel Coronavirus Disease 2019 (COVID-19). J Microbiol Biotechnol, v. 30, n. 3, p. 313-324, Mar 28 2020.https://doi.org/10.4014/jmb.2003.03011

ALAGAILI, A. N. et al. Middle East respiratory syndrome coronavirus infection in dromedary camels in Saudi Arabia. mBio, v. 5, n. 2, p. e00884-14, Feb 25 2014. https://doi.org/10.1128/mBio.01002-14

ALMAZAN, F. et al. Coronavirus reverse genetic systems: infectious clones and replicons. Virus Res, v. 189, p. 262-70, Aug 30 2014. https://doi.org/10.1016/j.virusres.2014.05.026

ANDERSEN, K. G. et al. The proximal origin of SARS-CoV-2. Nat Med, v. 26, n. 4, p. 450-452, Apr 2020.https://doi.org/10.1038/s41591-020-0820-9

ARMSTRONG, J. et al. Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus. Nature, v. 308, n. 5961, p. 751-2, Apr 19-25 1984.https://doi.org/10.1038/308751a0

ARONS, M. M. et al. Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. N Engl J Med, v. 382, n. 22, p. 2081-2090, May 28 2020.

AZHAR, E. I. et al. The Middle East Respiratory Syndrome (MERS). Infect Dis Clin North Am, v. 33, n. 4, p. 891-905, Dec 2019.https://doi.org/10.1016/j.idc.2019.08.001

BELOUZARD, S.; CHU, V. C.; WHITTAKER, G. R. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. Proc Natl Acad Sci U S A, v. 106, n. 14, p. 5871-6, Apr 7 2009. https://doi.org/10.1073/pnas.0809524106

CANDIDO, D. S. et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. Science, v. 369, n. 6508, p. 1255-1260, Sep 4 2020. https://doi.org/10.1126/science.369.6508.1203-o

CHANG, C. K. et al. Modular organization of SARS coronavirus nucleocapsid protein. J Biomed Sci, v. 13, n. 1, p. 59-72, Jan 2006.https://doi.org/10.1007/s11373-005-9035-9

CHEN, N. et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet, v. 395, n. 10223, p. 507- 513, Feb 15 2020.https://doi.org/10.1016/S0140-6736(20)30211-7

CHEN, Y. et al. A sensitive and specific antigen detection assay for Middle East respiratory syndrome coronavirus. Emerg Microbes Infect, v. 4, n. 4, p. e26, Apr 2015.
https://doi.org/10.1038/emi.2015.26

CHU, H. et al. Comparative tropism, replication kinetics, and cell damage profiling of SARS- CoV-2 and SARS-CoV with implications for clinical manifestations, transmissibility, and laboratory studies of COVID-19: an observational study. Lancet Microbe, v. 1, n. 1, p. e14-e23, May 2020.https://doi.org/10.1016/S2666-5247(20)30004-5

CIOTTI, M. et al. The COVID-19 pandemic. Crit Rev Clin Lab Sci, v. 57, n. 6, p. 365-388, Sep 2020.https://doi.org/10.1080/10408363.2020.1783198

CORMAN, V. M. et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR.Euro Surveill, v. 25, n. 3, Jan 2020.https://doi.org/10.2807/1560-7917.ES.2020.25.21.2001035

CORONAVIRIDAE STUDY GROUP OF THE INTERNATIONAL COMMITTEE ON TAXONOMY OF, V. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol, v. 5, n. 4, p. 536-544, Apr 2020.https://doi.org/10.1038/s41564-020-0695-z

DE HAAN, C. A.; ROTTIER, P. J. Molecular interactions in the assembly of coronaviruses. Adv Virus Res, v. 64, p. 165-230, 2005.https://doi.org/10.1016/S0065-3527(05)64006-7

FEHR, A. R.; PERLMAN, S. Coronaviruses: an overview of their replication and pathogenesis. Methods Mol Biol, v. 1282, p. 1-23, 2015.https://doi.org/10.1007/978-1-4939-2438-7_1

GLOWACKA, I. et al. Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. J Virol, v. 85, n. 9, p. 4122-34, May 2011.
https://doi.org/10.1128/JVI.02232-10

GUAN, Y. et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science, v. 302, n. 5643, p. 276-8, Oct 10 2003.https://doi.org/10.1126/science.1087139

HAMMING, I. et al. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. J Pathol, v. 203, n. 2, p. 631-7, Jun 2004.https://doi.org/10.1002/path.1570

HAN, Q. et al. Uncertainties about the transmission routes of 2019 novel coronavirus. Influenza Other Respir Viruses, v. 14, n. 4, p. 470-471, Jul 2020. https://doi.org/10.1111/irv.12735

HOFFMANN, M. et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell, v. 181, n. 2, p. 271-280 e8, Apr 16 2020.https://doi.org/10.1016/j.cell.2020.02.052

HOLSHUE, M. L. et al. First Case of 2019 Novel Coronavirus in the United States. N Engl J Med,v. 382, n. 10, p. 929-936, Mar 5 2020.https://doi.org/10.1056/NEJMoa2001191

HUANG, C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet, v. 395, n. 10223, p. 497-506, Feb 15 2020. https://doi.org/10.1016/S0140-6736(20)30183-5

JESUS, J. G. et al. Importation and early local transmission of COVID-19 in Brazil, 2020. Rev Inst Med Trop Sao Paulo, v. 62, p. e30, 2020.

KIM, D. et al. The Architecture of SARS-CoV-2 Transcriptome. Cell, v. 181, n. 4, p. 914-921 e10, May 14 2020.https://doi.org/10.1016/j.cell.2020.04.011

KSIAZEK, T. G. et al. A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med, v. 348, n. 20, p. 1953-66, May 15 2003.

LEE, C. Porcine epidemic diarrhea virus: An emerging and re-emerging epizootic swine virus.Virol J, v. 12, p. 193, Dec 22 2015.https://doi.org/10.1186/s12985-015-0421-2

LEUNG, N. H. L. et al. Respiratory virus shedding in exhaled breath and efficacy of face masks.Nat Med, v. 26, n. 5, p. 676-680, May 2020. https://doi.org/10.1038/s41591-020-0843-2

LI, Q. et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med, v. 382, n. 13, p. 1199-1207, Mar 26 2020.

LI, X. et al. Molecular immune pathogenesis and diagnosis of COVID-19. J Pharm Anal, v. 10, n. 2, p. 102-108, Apr 2020.https://doi.org/10.1016/j.jpha.2020.03.001

LIPPI, G.; PLEBANI, M. The critical role of laboratory medicine during coronavirus disease 2019 (COVID-19) and other viral outbreaks. Clin Chem Lab Med, v. 58, n. 7, p. 1063-1069, Jun 25 2020.https://doi.org/10.1515/cclm-2020-0240

LIU, J. et al. Community Transmission of Severe Acute Respiratory Syndrome Coronavirus 2, Shenzhen, China, 2020. Emerg Infect Dis, v. 26, n. 6, p. 1320-1323, Jun 2020.https://doi.org/10.3201/eid2606.200239

MACKAY, I. M.; ARDEN, K. E. MERS coronavirus: diagnostics, epidemiology and transmission.Virol J, v. 12, p. 222, Dec 22 2015.https://doi.org/10.1186/s12985-015-0439-5

MALIK, Y. A. Properties of Coronavirus and SARS-CoV-2. Malays J Pathol, v. 42, n. 1, p. 3-11, Apr 2020.

MIZUMOTO, K. et al. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. Euro Surveill, v. 25, n. 10, Mar 2020.https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000180

NAO, N. et al. Genetic Predisposition To Acquire a Polybasic Cleavage Site for Highly Pathogenic Avian Influenza Virus Hemagglutinin. mBio, v. 8, n. 1, Feb 14 2017.https://doi.org/10.1128/mBio.02298-16

NEUMAN, B. W. et al. A structural analysis of M protein in coronavirus assembly and morphology. J Struct Biol, v. 174, n. 1, p. 11-22, Apr 2011.https://doi.org/10.1016/j.jsb.2010.11.021

NIETO-TORRES, J. L. et al. Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein. Virology, v. 415, n. 2, p. 69-82, Jul 5 2011.https://doi.org/10.1016/j.virol.2011.03.029

ORTEGO, J. et al. Generation of a replication-competent, propagation-deficient virus vector based on the transmissible gastroenteritis coronavirus genome. J Virol, v. 76, n. 22, p. 11518- 29, Nov 2002.https://doi.org/10.1128/JVI.76.22.11518-11529.2002

QIAN, Z.; DOMINGUEZ, S. R.; HOLMES, K. V. Role of the spike glycoprotein of human Middle East respiratory syndrome coronavirus (MERS-CoV) in virus entry and syncytia formation. PLoS One, v. 8, n. 10, p. e76469, 2013. https://doi.org/10.1371/journal.pone.0076469

RABI, F. A. et al. SARS-CoV-2 and Coronavirus Disease 2019: What We Know So Far. Pathogens, v. 9, n. 3, Mar 20 2020.https://doi.org/10.3390/pathogens9030231

RUAN, Q. et al. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. Intensive Care Med, v. 46, n. 5, p. 846-848, May 2020.https://doi.org/10.1007/s00134-020-05991-x

SANTOS, I. A. et al. Antivirals Against Coronaviruses: Candidate Drugs for SARS-CoV-2 Treatment? Front Microbiol, v. 11, p. 1818, 2020.https://doi.org/10.3389/fmicb.2020.01818

TOOZE, J.; TOOZE, S.; WARREN, G. Replication of coronavirus MHV-A59 in sac- cells: determination of the first site of budding of progeny virions. Eur J Cell Biol, v. 33, n. 2, p. 281- 93, Mar 1984.

TYRRELL, D. A. et al. Coronaviridae. Intervirology, v. 5, n. 1-2, p. 76-82, 1975. https://doi.org/10.1159/000149883

TYRRELL, D. A.; BYNOE, M. L. Cultivation of viruses from a high proportion of patients with colds. Lancet, v. 1, n. 7428, p. 76-7, Jan 8 1966. https://doi.org/10.1016/S0140-6736(66)92364-6

WANG, D. et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA, v. 323, n. 11, p. 1061-1069, Mar 17 2020.https://doi.org/10.1001/jama.2020.1585

WEI, W. E. et al. Presymptomatic Transmission of SARS-CoV-2 - Singapore, January 23-March 16, 2020. MMWR Morb Mortal Wkly Rep, v. 69, n. 14, p. 411-415, Apr 10 2020.https://doi.org/10.15585/mmwr.mm6914e1

WOO, P. C. et al. Coronavirus diversity, phylogeny and interspecies jumping. Exp Biol Med (Maywood), v. 234, n. 10, p. 1117-27, Oct 2009. https://doi.org/10.3181/0903-MR-94

WOO, P. C. et al. Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. J Virol, v. 86, n. 7, p. 3995-4008, Apr 2012. https://doi.org/10.1128/JVI.06540-11

WU, F. et al. A new coronavirus associated with human respiratory disease in China. Nature, v. 579, n. 7798, p. 265-269, Mar 2020.https://doi.org/10.1038/s41586-020-2008-3

XAVIER, J. et al. The ongoing COVID-19 epidemic in Minas Gerais, Brazil: insights from epidemiological data and SARS-CoV-2 whole genome sequencing. Emerg Microbes Infect, v. 9, n. 1, p. 1824-1834, Dec 2020. https://doi.org/10.1101/2020.05.05.20091611

ZAKI, A. M. et al. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia.N Engl J Med, v. 367, n. 19, p. 1814-20, Nov 8 2012. https://doi.org/10.1056/NEJMoa1211721

ZHAO, J. et al. Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. Clin Infect Dis, Mar 28 2020.

ZHU, N. et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med, v. 382, n. 8, p. 727-733, Feb 20 2020. https://doi.org/10.1056/NEJMoa2001017

CAPÍTULO II

Manuscript:

SARS-CoV-2 GENOME SEQUECING ANALYSIS IN MINAS GERAIS, BRAZIL.

*****Este capítulo está em formato de manuscrito com algumas alterações estruturais para melhor se adequar ao formato da dissertação. O artigo em questão será submetido à revista **Virology.**

SARS-CoV-2 GENOME SEQUECING ANALYSIS IN MINAS GERAIS, BRAZIL

Giulia Magalhães Ferreira[a], Ingra Morales Claro[b,c], Victória Riquena Grosche[a,e], Darlan Cândido[b,d], Ester Cerdeira Sabino[b,c], Jaqueline Goes de Jesus[b,c], Ana Carolina Gomes Jardim[a,e*]

[a] Institute of Biomedical Science, Federal University of Uberlândia, Uberlândia, Minas Gerais, Brazil

[b] Institute of Tropical Medicine, University of São Paulo Medical School, São Paulo, Brazil.

[c] Department of Infectious and Parasitic Diseases**,** University of São Paulo Medical School, São Paulo, Brazil.

[d] Department of Zoology, University of Oxford, Oxford, UK.

[e] Institute of Bioscience, Humanities and Exact Sciences, São Paulo State University, São José do Rio Preto, São Paulo, Brazil.

Corresponding author**:** Professor Ana Carolina Gomes Jardim, Institute of Biomedical Science, ICBIM, Federal University of Uberlandia, Avenida Amazonas, 4C- Room 216, Umuarama, Uberlândia, Minas Gerais, Brazil, CEP: 38405-302

Tel: +55 (34) 3225-8682

E-mail: jardim@ufu.br

Abstract

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first identified in Wuhan-China, as the causative agent of the coronavirus disease 2019 (COVID-19). Since notified in São Paulo on 26st February 2020, more than 3,000,000 cases and 106,000 deaths were reported in Brazil. In the early epidemic phase, SARS-CoV-2 spread locally, however, over time, this virus was disseminated to other regions of the country. Here, we performed genome sequencing and phylogenetic analysis of SARS-CoV-2 of 20 clinical samples of COVID-19 confirmed cases from 9 cities of MG, in order to evaluate the genetic characterization of circulating viral strains in the state from March to May 2020. Our analyses demonstrated the circulation of B.1 lineage isolates in the investigated locations and nucleotide substitutions were observed into the genomic regions related to important viral structures. Additionally, sequences generated in this study clustered with isolates from Sao Paulo state, suggesting a dissemination route between these 2 states. Alternatively, monophyletic groups of sequences from MG and other states or country were observed, indicating independent events of virus introduction. These results reinforce the need of genomic surveillance for understand the ongoing spread of emerging viral pathogens.

Keyword: SARS-CoV-2, COVID-19, genomic surveillance, Minas Gerais, genome sequencing, B.1 lineage.

Introduction

The World Health Organization (WHO) was informed on 31 December 2019 of the occurrence of unknown etiology respiratory disease cases in Wuhan, China[1]. Chinese authorities isolated and identified this pathogen as a novel coronavirus, the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the causative agent of the coronavirus disease 2019 (COVID-19)[2]. Based on phylogenetic analysis, SARS-Cov-2 have been classified as a Betacoronavirus with a 30 kb genome and divided in two lineages, SARS-CoV-2 A and B, according to the recent proposed lineage nomenclature[3,4,5]. Results from next-generation sequencing analysis have shown that SARS-Cov-2 has homology to others coronaviruses (CoVs) such as SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV)[2,3,6].

SARS-Cov-2 was declared a public health emergency of international concern in January 2020[7] and later in September, COVID-19 caused over 21,000,000 cases and more than 761,000 deaths around the world[8]. The case fatality rates of this disease ranges from 1,2 to 1,6%, although, in over 60s that ratio considerably increases[9,10,11]. While there is no licensed antiviral for SARS-CoV-2, several countries have used non-pharmaceutical interventions (NPIs) in an attempt to control the pandemic[12,13]. However, despite these measures have presented positive effects, studies have demonstrated that SARS-CoV-2 transmission from asymptomatic or pre-symptomatic individuals complicates public health efforts to combat COVID-19[12,14,15].

In Brazil, the first confirmed case of COVID-19 was reported in São Paulo state (SP) on 26 February 2020[16]. Analysis of the first two whole-genome sequences of SARS-CoV-2, isolated from Brazilian patients who recent returned from Italy, demonstrated two independent introductions of the virus into the country and local spread during initial stages of transmission[16]. In Brazil COVID-19 is responsible for causing 3,000,000 cases and more than 106,000 deaths[8]. During the early epidemic phase, SARS-CoV-2 spread locally, however, despite interventions to prevent the virus dissemination over time, large urban centers have become responsible for spreading the virus to other locations. According to recent estimates, over 100 international introductions of virus has been observed in Brazil, and most of Brazilian strains are classified in three clades[12] Clade 1 circulates predominantly in SP state and presents a nucleotide substitution in the spike protein, clade 2 is a widespread lineage, found in a several brazilian states and is

characterized by two nucleotide substitution (ORF 6 and nucleoprotein) and clade 3 was found concentrated in Ceará state[12].

Minas Gerais (MG) ranks as the second most populous state and the fourth largest area in Brazil, as well as is the third position in the Gross Domestic Product (GDP) values. MG shares borders with SP, Bahia, Rio de Janeiro, Goiás and Mato Grosso do Sul. Belo Horizonte, the capital of MG is the major urban and finance center in Latin America[17]. Due to its large population size, economic situation and facilitated access to other economy important states, MG represented a potential region to spread SARS-CoV-2 and potentially aggravated the current pandemic[18]. To date, 222.046 cases and 5.505 deaths of COVID-19 have been reported in MG[19]. According to recent studies, most of sequences analyzed from samples collected in MG were classified as SARS-Cov-2 lineage B.1, which contains sequences from the USA, Australia, China and other countries[18]. Here, we performed genome sequencing analysis of SARS-CoV-2 from 20 clinical samples of COVID-19 confirmed cases in 9 municipalities of MG, in order to evaluate the genetic characterization of circulating viral strains in the state from March to May 2020.

Materials and methods

Ethics statement and samples

Samples used in this study were collected via nasopharyngeal swab and tested positive for SARS-CoV-2 by RT-qPCR, obtained from private medical diagnostic laboratories. Ethical approval for this study was obtained from the National Ethical Review Board with approval number CAAE 30127020.0.0000.0068.

Twenty SARS-CoV-2 samples with RT-qPCR positive results, collected between March 25 and May 27 2020 from patients attended in private laboratories of diagnosis of 9 cities of the state of Minas Gerais (MG), Brazil, were selected to this study. Samples were processed for genome sequencing at the Institute of Tropical Medicine University of São Paulo. Metadata related to the samples of this study included information on samples (date and municipality of collection and cyclethreshold (Ct) of SARS-CoV-2 detection by RT-qPCR) and data from patients (gender and age).

cDNA synthesis and virus multiplex PCR amplification

Viral RNA was used for cDNA transcription using Protoscript II First Strand cDNA synthesis Kit (New England Biolabs, UK) and random hexamers (Thermo Fisher Scientific, USA). Whole genome amplification was performed by multiplex PCR using SARS-CoV-2 primers described previously (https://artic.network/ncov-2019) and Q5 High-Fidelity DNA polymerase (New England Biolabs, UK)[20]. PCR conditions have been previously reported (https://artic.network/ncov-2019). Amplified PCR products were cleaned-up using the 1x AMPure XP beads (Beckman Coulter, United Kingdom) and quantified using fluorimeter with the Qubit dsDNA High Sensitivity assay on the Qubit 3.0 instrument (Life Technologies, USA).

Whole genome sequencing and genome assembly

The 20 samples selected to this study contained sufficient DNA (≥ 1ng\µL) to proceed to library preparation. Amplicons from each sample were normalized and pooled in an equimolar fashion and barcoded using the EXP-NBD104 (1–12) and EXPNBD114 (13–24) Native Barcoding Kits (Oxford Nanopore Technologies, UK), following a previously published protocol[20]. After barcoding ligation, libraries were loaded on a flow cell and sequenced by MinION during 8-24 hours using SQK-LSK109 Kit (ONT, UK). To monitor sequencing in real-time and estimate the depth of coverage (target of 200-fold) across the genome for each barcoded sample (https://artic.network/rampart), RAMPART software from the ARTIC Network (https://artic.network/ncov-2019) was used. After generated reads, fast5 files were basecalled, demultiplexed, and trimmed using Guppy software v2.2.7 (ONT, UK). Minimap2 v2.28.0 was used to obtain the consensus genomes by mapping the fastq files to the reference genome of SARS-CoV-2 isolate Wuhan-Hu 1 (GenBank Accession Number MN908947) and SAMtools was used to converted these files in a sorted BAM file[21]. The quality test and length filtering were performed for each barcode using guppyplex (https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html). Bioedit was used to build a multiple sequence alignment of the resulting dataset[22].

Collation of SARS-CoV-2 global datasets

Our dataset contains 20 sequences of SARS-CoV-2 genomes from 9 MG municipalities produced in this study. We annexed 31 sequences from other regions of MG, available in GISAID, to this data[23] (https://www.gisaid.org), generating a dataset of 51 sequences of the whole genomes from 16 municipalities of MG. The datasets from MG represent approximately 1 sequence for every 157 cases (0.63%) notified up to 27th May 2020 in the state[24] Juiz de Fora is the third city in MG of our dataset with the highest number of SARS-CoV-2 notified cases, including 3,18% of all notified cases in this city during the period of samples collection, representing approximately 1 sequence for every 31 cases[24].

Phylogenetic analysis of SARS-CoV-2 in Minas Gerais

The dataset with 1637 sequences of whole genomes of SARS-CoV-2 globally distributed were generated. In this dataset, 57,73% (945/1637) of the sequences are from Brazilian isolates and 47,27% (692/1637) are sequences of isolates from other countries. Among sequences from Brazil, 51 were generated from samples collected in MG, being 20 of those new sequences generated by this study. Sequences used for analysis had the genome coverage > 65%. MAFFT was used to build a multiple sequence alignment of dataset, which was edited with AliView and used to the phylogenetic reconstruction. A maximum likelihood phylogenetic tree was estimated using Hasegawa-KishinoYano nucleotide substitution model[31] with a gamma distributed rate variation among-site[27] in IQTree v.2[26]. Finally, SARS-CoV-2 lineages were identified using Phylogenetic Assignment of Named Global Outbreak LINeages tool (https://github.com/cov-lineages/pangolin).

Results

Samples from clinically suspected cases of COVID-19 collected at the private laboratory Diagnosis of Brazil (DB) in 9 cities of Minas Gerais (MG) were screened for the detection of SARS-CoV-2 by RT-PCR. Received samples had cyclethreshold (Ct) values that ranged from 9,96 to 25.59 (average of 18.41) (Table 1). Twenty samples (DNA ≥ 1 ηg\µL) were selected for whole genome sequencing using a combination of multiplex PCR amplification (https://artic.network/ncov-2019) and Nanopore sequencing. We obtained an average reference coverage of 85,10% related to the reference genome

NC_045512.2. Selected samples were collected from women (8/20, 40%) and men patients (12/20, 60%) (Table 1), the average age was 53.4 years old. Of the 20 sequenced samples, 35% were from Juiz de Fora, the remaining samples were from Uberlândia (10%), Uberaba (5%), Santos Dumont (5%), Pouso Alegre (10%), Formiga (5%), Extrema(10%), Cambuí (15%) and Barbacena (5%) and the collection dates ranging from 25 March 2020 to 25 May 2020 (Figure 1). As shown in the Figure 1, the samples investigated in this study were obtained in the most southern region of MG state, an area which shares borders with SP. Due to the proximity and the fact that São Paulo is a great economic pole, the movement of people between these locations is quite common.

Phylogenetic trees were reconstructed from the 20 full-length genome sequences generated in this study and additional 1617 sequences deposited in GISAID[23] up to 25 May 2020 as described in the Methods section (Figure2). Sequences clustered according to lineages A and B, representing the Wuhan/WH04/202 and Wuhan-Hu-, respectively, as the recently proposed SARS-CoV-2 lineage nomenclature[5]. Additionally, lineage B is divided into sublineages, and our phylogenetic analysis revealed that sequences from MG grouped to the lineages B.1 and B.1.1 (Figure 2). Pangolin analysis also demonstrated the similarity between isolates analyzed in this study and those identified in countries as China, the USA, Australia, Portugal, United Kingdom and Brazil (Table 1).

Concerning the 20 sequences generated in this study, phylogenetic analysis demonstrated that sequences of isolates from 6 cities (Cambuí [1], Extrema [1], Formiga [1], Juiz de for a [3], Pouso Alegre [2] and Barbacena [1]) were grouped in clusters with isolates from SP (Figure S1), suggesting a route of dissemination of the virus between these two states. Otherwise, some sequences from this work grouped with isolates from other Brazilian states as Santa Catarina (SC; 1), Pará (PA; 2) and Rio Grande do Sul (RS; 1), and with another country, France (1), suggesting independent introductions. Additionally, sequences from Juiz de Fora (MG21) and Santos Dumont (MG19) grouped in the same monophyletic groups, which is sustained by the proximity of these cities (approximately 48km) (Figure 1). A sequence generated in this study from sample of Juiz de Fora grouped with GISAID dataset sequences from the same locality, suggesting a locally circulating of SARS-CoV-2.

The topology of the 20 sequences generated in this study demonstrated that the viral isolates grouped in 3 main monophyletic groups (Figure 3). Sequences from Juiz de Fora (MG42 and MG21), Uberaba (MG6) and Santos Dumont (MG19) (Figure 4A) were

characterized by a nucleotide substitution in the spike protein (A23403G) (Figure 4B). Cluster with sequences from Juiz de Fora (MG22 and MG54), Cambuí (MG10) and Extrema (MG18) (Figure 4C) presented two nucleotide substitutions: in the ORF6 (T27299C) (Figure 4D) and in the nucleoprotein (T29148C) (Figure 4E). Sequences from Cambuí (MG9 and MG10) and from Juiz de Fora (MG22, MG38, MG52 and MG54) demonstrated three nucleotide substitution on nucleocapsid phosphoprotein (G28881A, G28882A and G28883C) (Figure 4F). As described in table 2, some sequences showed nucleotide substitution in viral non-structural proteins (nsps) 2, 3, 4, 7, 8, 12 and in the structural protein S (Table2).

Discussion

This study presents the preliminary situation of COVID-19 epidemic in 9 cities of the most southern region of Minas Gerais (MG). Until 25 May 2020, MG reported 6.962 cases of COVID-19[19]. Due to its proximity to the city of São Paulo and as a result of the intense industrialization that some cities have recently undergone, such as Pouso Alegre and Extrema, the most southern region of MG has a representative role to understand the progress of SARS-CoV-2 in Brazil. In addition, there are currently not many sequences from this region in the GISAID dataset, compromising the understanding of the epidemic. The data presented here was obtained though sequencing of 20 samples of SARS-CoV-2 infection confirmed by RT-qPCR along with others 1617 sequences previously deposited in GISAID[23].

Characterization of circulating viral lineages is realized through genetic analyses and surveillance. SARS-CoV-2 lineage A can be defined by Wuhan/WH04/2020 sequence and lineage B is characterized by Wuhan-Hu-1 sequence, however, lineage B can be divided in sublineages[25,28]. The 20 whole genome sequences generated in this study and further 1617 from GISAID composed the dataset of study and the reconstruction of the phylogenetic tree demonstrated that 20 sequences was grouped to the lineage B.1 and B1.1 clusters. As well as in others studies[18], we demonstrated the prevalence of SARS-CoV-2 lineage B.1.1 circulation in Minas Gerais. The B.1.1 strain also circulates in countries such as Portugal, Australia, United Kingdom and the USA, showing that, at some point of pandemic, international introductions of SARS-CoV-2 occurred in Brazil, as well as Minas Gerais, a well-connected state where these introductions occurred[12].

Our results demonstrated that sequences generated in this work of samples from 6 cities (Cambuí, Extrema, Formiga, Juiz de fora, Pouso Alegre and Barbacena) clustered with sequences of isolates from São Paulo (SP). During the initial of the epidemic in Brazil, COVID-19 cases were mainly reported in Sao Paulo[12], a state that shares border with the most southern regions of MG. Since some isolates from these 2 states were grouped in the same monophilletic groups, it migh suggest a possible transmission route between these municipalities. Our data also demonstrated the clustering of sequences from this work with isolates from other Brazilian states, other country and with an isolated from the same location. All these findings support either, the dissemination of the virus among nearby geographical areas or independent introductions from other locations.

Nucleotide substitutions in the genomic regions of the spike protein, ORF6, nucleoprotein and nucleocapsid phosphoprotein were observed in our analysis. In the beginning of the pandemic in Brazil, viral strains with substitution in the spike protein regions was shown to predominantly circulate in SP state[12], however, substitutions of nucleotides in the ORF6 and nucleoprotein genomic regions were found to be spatially wide spread in Brazil[12]. It suggests that over time, with flexibility of social isolations measures and due to the easy access of the cities studied with SP, the viral strains that mostly circulated in São Paulo have spread to others states.

SARS-CoV-2 contains open reading frames (ORFs) which encode the four main structural proteins (Spike, Envelope, Nucleocapsid and Membrane)[3]. The spike protein interacts with cell host receptor controlling viral tissue tropism[29]. In SARS-CoV-2 case, spike protein attaches to the host receptor Angiotensin-converting enzyme 2 (ACE2)[30]. This protein is absolutely essential in the early stages of SARS-CoV-2 infection, therefore, is important to track differentiations in the genetic sequences that encode this protein.

In summary, we demonstrated the predominance of a circulating lineage in the regions investigated, several nucleotide substitutions in important regions of virus genome and the phylogenetic relationship between the newly generated sequences and sequences from the global dataset Some substitutions were initial found in others states of Brazil, which suggests that over time and with flexibilization of social distance measures, virus has spread to other regions of the country. Despite the samples cover up about 60 days of the MG epidemic, these findings provide very relevant information to better understand the impacts that the genetic variability of SARS-CoV-2 may cause in the future and reinforce

the need for genomic surveillance for understand the ongoing spread of emerging viral pathogens.

REFERENCE

1. WU, F. et al. A new coronavirus associated with human respiratory disease in China. **Nature,** v. 579, n. 7798, p. 265-269, Mar 2020.

2. LU, R. et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. **Lancet,** v. 395, n. 10224, p. 565-574, Feb 22 2020.

3. MALIK, Y. A. Properties of Coronavirus and SARS-CoV-2. **Malays J Pathol,** v. 42, n. 1, p. 3-11, Apr 2020.

4. NEUMAN, B. W. et al. Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. **J Virol,** v. 80, n. 16, p. 7918-28, Aug 2006.

5. RAMBAUT, A. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. **Nat Microbiol**, Jul 15 2020.

6. ZHU, N. et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. **N Engl J Med,** v. 382, n. 8, p. 727-733, Feb 20 2020.

7. 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.

8. 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/20200816-covid-19-sitrep-209.pdf?sfvrsn=5dde1ca2_2

9. WU, J. T. et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. **Nat Med,** v. 26, n. 4, p. 506-510, Apr 2020.

10. VERITY, R. et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. **Lancet Infect Dis,** v. 20, n. 6, p. 669-677, Jun 2020.

11. RUSSELL, T. W. 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. **Euro Surveill,** v. 25, n. 12, Mar 2020.

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

13. SANTOS, I. D. A. et al. Antivirals Against Coronaviruses: Candidate Drugs for SARS-CoV-2 Treatment? **Frontiers in Microbiology,** v. 11, n. 1818, 2020-August-13 2020.

14. ARONS, M. M. et al. Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. **N Engl J Med,** v. 382, n. 22, p. 2081-2090, May 28 2020.

15. MIZUMOTO, K. et al. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. **Euro Surveill,** v. 25, n. 10, Mar 2020.

16. JESUS, J. G. et al. Importation and early local transmission of COVID-19 in Brazil, 2020. **Rev Inst Med Trop Sao Paulo,** v. 62, p. e30, 2020.

17. Instituto Brasileiro de Geografia e Estatística - IBGE. Cidades e Estados: Minas Gerais [Internet]. 2020.Available from: https://www.ibge.gov.br/cidades-eestados/mg.html

18. XAVIER, J. et al. The ongoing COVID-19 epidemic in Minas Gerais, Brazil: insights from epidemiological data and SARS-CoV-2 whole genome sequencing. **Emerg Microbes Infect,** v. 9, n. 1, p. 1824-1834, Dec 2020.

19. Ministério da Saúde. Boletim Epidemiológico Minas Gerais Available from: http://coronavirus.saude.mg.gov.br/images/boletim/09-setembro/16.09.2020-Boletim-Epidemiologico_COVID-19.pdf

20. QUICK, J. et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. **Nat Protoc,** v. 12, n. 6, p. 1261-1276, Jun 2017.

21. LI, H. et al. The Sequence Alignment/Map format and SAMtools. **Bioinformatics,** v. 25, n. 16, p. 2078-9, Aug 15 2009.

22. KATOH, K.; STANDLEY, D. M. MAFFT: iterative refinement and additional methods. **Methods Mol Biol,** v. 1079, p. 131-46, 2014.

23. SHU, Y.; MCCAULEY, J. GISAID: Global initiative on sharing all influenza data - from vision to reality. **Euro Surveill,** v. 22, n. 13, Mar 30 2017.

24. Ministério da Saúde. Boletim Epidemiológico Minas Gerais Available from: http://coronavirus.saude.mg.gov.br/images/boletim/05maio/25052020_Boletim_epidemiologico_COVID-19_MG.pdf

25. KRAEMER, M. U. G. et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. **Science,** v. 368, n. 6490, p. 493-497, May 1 2020.

26. MINH, B. Q. et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. **Mol Biol Evol,** v. 37, n. 5, p. 1530-1534, May 1 2020

27. YANG, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. **J Mol Evol,** v. 39, n. 3, p. 306-14, Sep 1994.

28. TIAN, H. et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. **Science,** v. 368, n. 6491, p. 638-642, May 8 2020.

29. KIRCHDOERFER, R. N. et al. Pre-fusion structure of a human coronavirus spike protein. **Nature,** v. 531, n. 7592, p. 118-21, Mar 3 2016.

30. HUSSAIN, M. et al. Structural variations in human ACE2 may influence its binding with SARS-CoV-2 spike protein. **J Med Virol**, Apr 6 2020.

31. HASEGAWA, M.; KISHINO, H.; YANO, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. **J Mol Evol,** v. 22, n. 2, p. 160-74, 1985.

Table 1. Data from samples of this study.

CADDE ID	Sample	Ct value	Collection date	Age	Gender	State	Municipality	Lineage	Most commom countries
MG 1	Swab	25,59	25/03/20	58	Male	MG	Extrema	B.1.1	USA, Portugal, United Kingdom
MG 4	Swab	9,96	30/03/20	80	Male	MG	Uberlândia	B.1.1	Brazil, China, United Kingdom
MG 6	Swab	20,58	31/03/20	73	Male	MG	Uberaba	B.1	USA, Australia, United Kingdom
MG 7	Swab	17,90	01/04/20	46	Male	MG	Cambuí	B.1	USA, Australia, United Kingdom
MG 8	Swab	20,09	02/04/20	63	Female	MG	Pouso Alegre	B.1.1	USA, Portugal, United Kingdom
MG 9	Swab	18,62	02/04/20	32	Female	MG	Cambuí	B.1	USA, Australia, United Kingdom
MG 10	Swab	19,04	02/04/20	79	Female	MG	Cambuí	B.1	USA, Australia, United Kingdom
MG 13	Swab	19,99	07/04/20	82	Male	MG	Uberlândia	B.1.1	USA, Portugal, United Kingdom
MG 15	Swab	16,87	07/04/20	25	Female	MG	Formiga	B.1.1	USA, Portugal, United Kingdom
MG 18	Swab	18,98	08/04/20	62	Female	MG	Extrema	B.1.1	USA, Portugal, United Kingdom
MG 19	Swab	15,78	09/04/20	75	Male	MG	Santos Dumont	B.1	USA, Australia, United Kingdom
MG 20	Swab	14,89	09/04/20	25	Female	MG	Pouso Alegre	B.1.1	USA, Portugal, United Kingdom
MG 21	Swab	18,67	15/04/20	31	Male	MG	Juiz de Fora	B.1	USA, Australia, United Kingdom
MG 22	Swab	22,09	15/04/20	46	Male	MG	Juiz de Fora	B.1.1	USA, Portugal, United Kingdom
MG 24	Swab	23,43	17/04/20	61	Male	MG	Barbacena	B.1.1	USA, Portugal, United Kingdom
MG 34	Swab	15,55	17/05/20	38	Male	MG	Juiz de Fora	B.1.1	USA, Portugal, United Kingdom
MG 38	Swab	14,61	18/05/20	69	Male	MG	Juiz de Fora	B.1.1	USA, Portugal, United Kingdom
MG 42	Swab	19,99	19/05/20	56	Male	MG	Juiz de Fora	B.1	USA, Australia, United Kingdom
MG 51	Swab	16,89	23/05/20	30	Female	MG	Juiz de Fora	B.1.1	Brazil, China, United Kingdom
MG 54	Swab	17,70	25/05/20	37	Female	MG	Juiz de Fora	B.1.1	USA, Australia, United Kingdom

Table2: Nucleotide substitutions.

Nucleotide Substitution	Product	Samples ID
C920T	nsp2	MG6
C1059T	nsp2	MG6
T5804C	nsp3	MG42
C6286T	nsp3	MG8
C6726T	nsp3	MG54
C8047T	nsp3	MG20
C8266T	nsp3	MG38
C9967T	nsp4	MG18
C12651T	nsp7	MG7, MG8, MG51
C12651T	nsp8	MG10
G14028T	nsp12	MG8
A14271G	nsp12	MG6
C23422T	S protein	MG20
C23683T	S protein	MG10

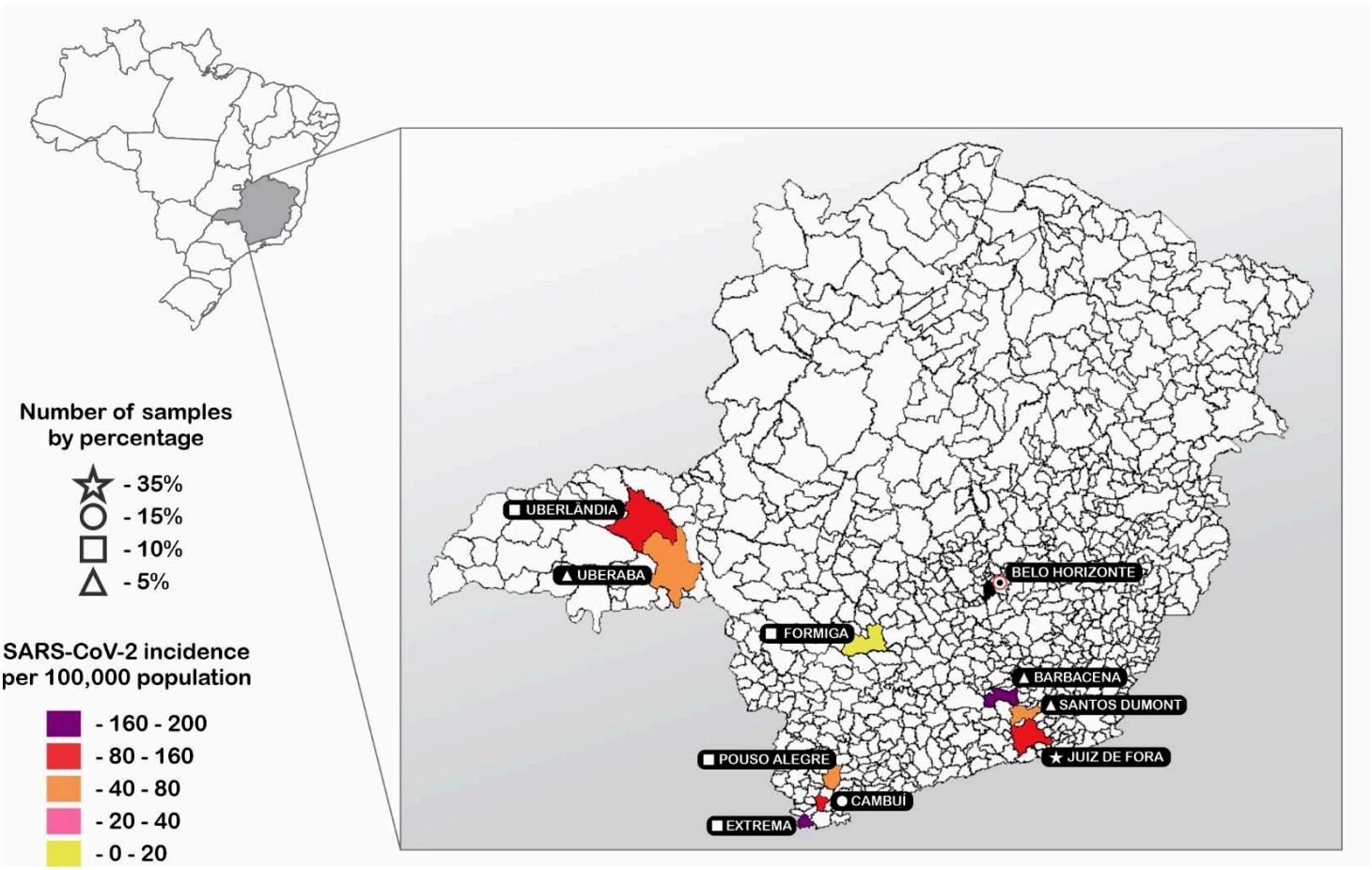


FIGURE 1. Area under investigation. Map of the MG state showing the percentage of new SARS-CoV-2 sequences by municipality and the incidence per 100,000 population.

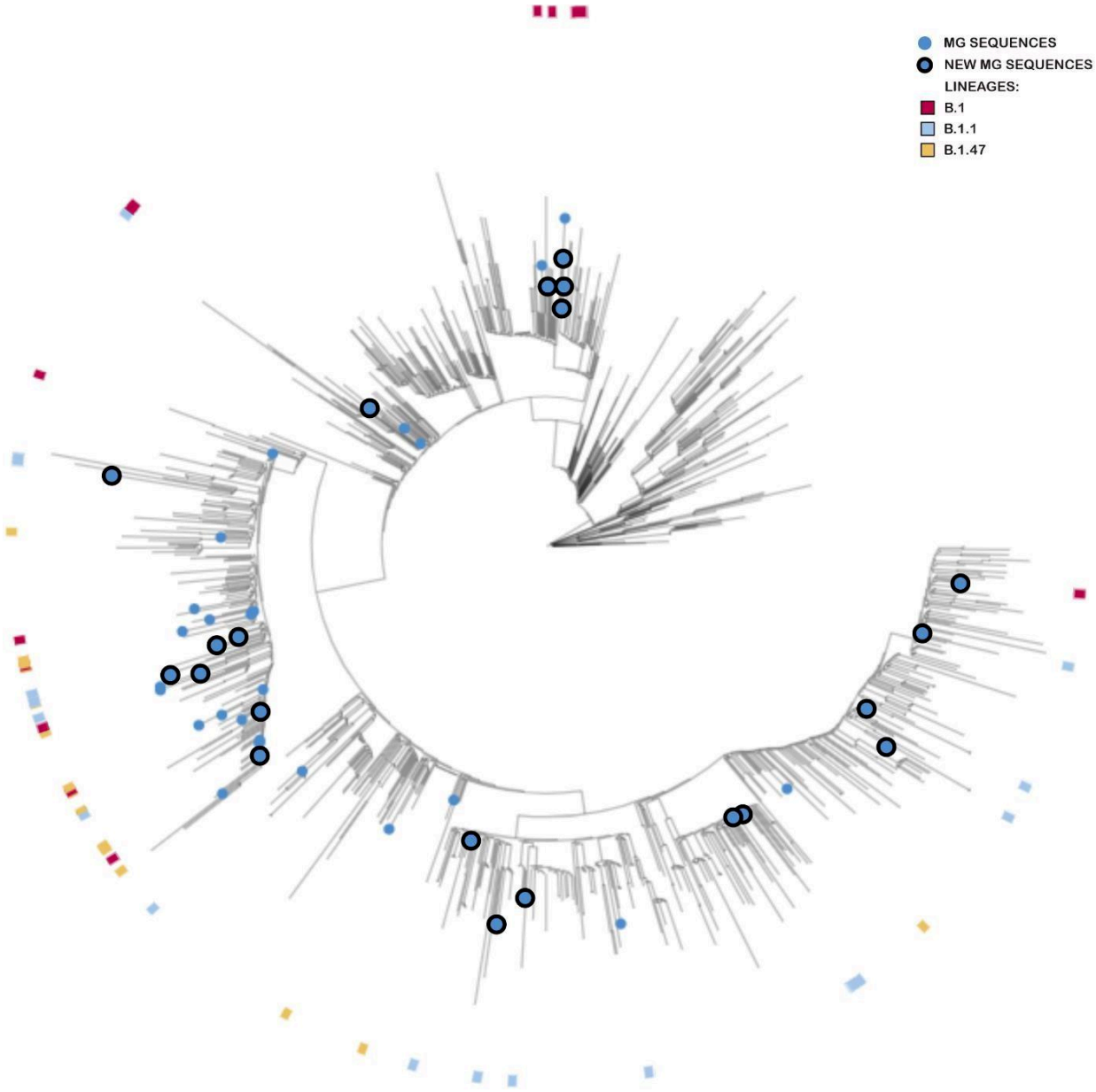


FIGURE 2. Phylogenetic tree of the predominant lineages in MG. All samples from MG belong to lineage B, varying between B1, B1.1 and B.1.47. Blue markers represent samples from Minas Gerais; The red circles mark the SARS-CoV-2 sequences generated in this study; The circle along the tree represents the lineages.

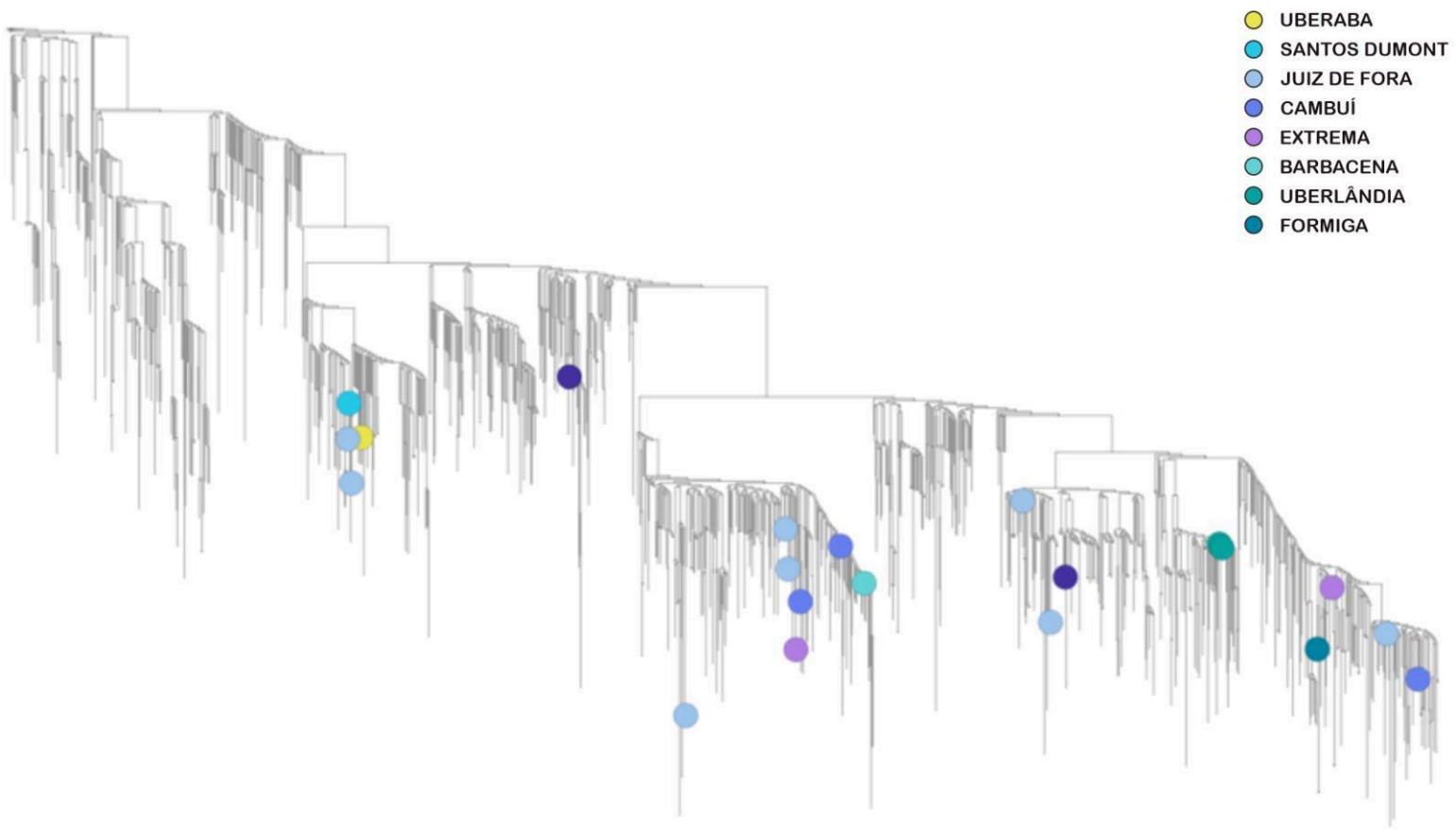


FIGURE 3. Phylogenetic tree. The phylogenetic tree demonstrates sequences grouped into monophylethic groups. The colored circles represent the sequences generated in this study and their respective municipalities.

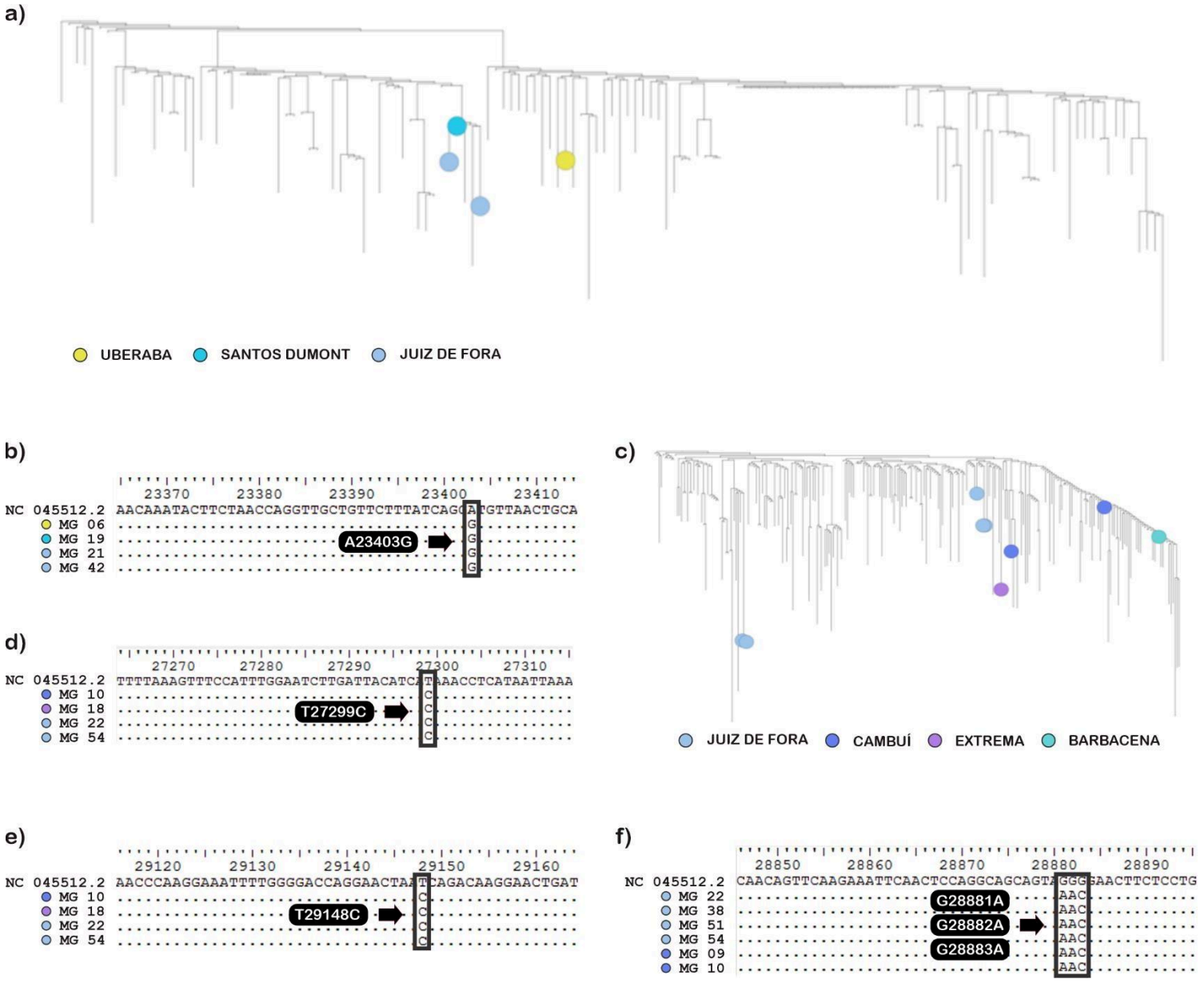
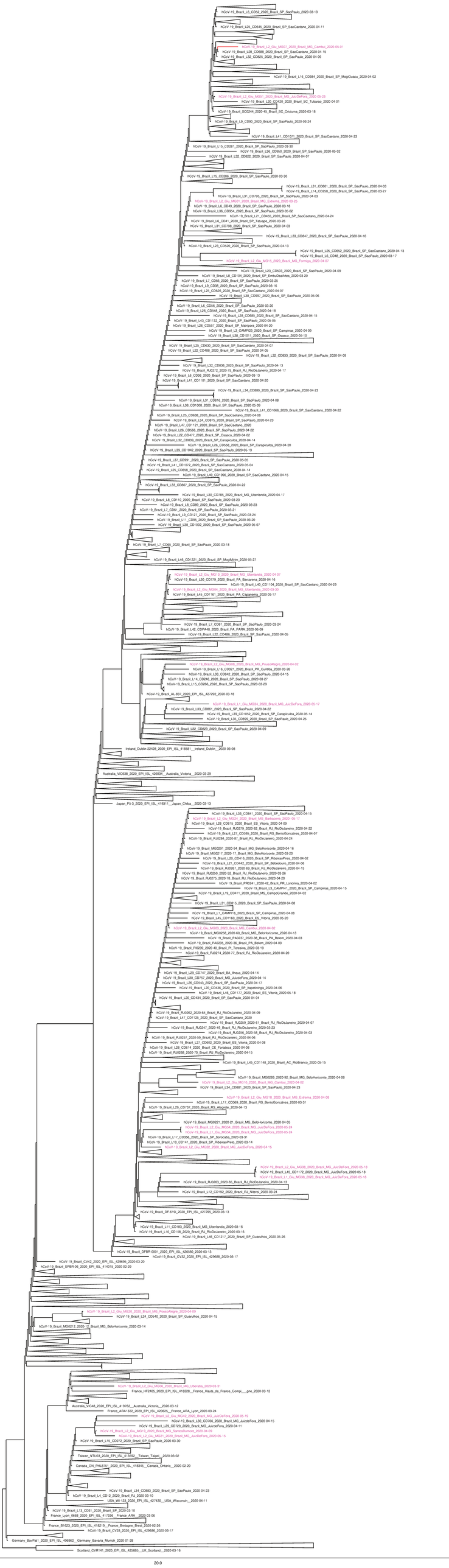


FIGURE 4. Nucleotide substitution. a and c Phylogenetic tree representing the sequences grouped in clade. b, d, e and f alignment of the sequences grouped together with the reference sequence, identifying the nucleotide substitution. The colored circles represent the samples analyzed in this study and their respective cities; The nucleotide substitution regions are marked by the black rectangle; Nucleotide substitutions are indicated by the black text box.

FIGURA SUPLEMENTAR


hCoV-19_Brazil_L6_CD52_2020_Brazil_SP_SaoPaulo_2020-03-19
hCoV-19_Brazil_L25_CD645_2020_Brazil_SP_SaoCaetano_2020-04-11
hCoV-19_Brazil_L2_Glu_MG07_2020_Brazil_MG_Cambui_2020-05-01
hCoV-19_Brazil_L28_CD688_2020_Brazil_SP_SaoCaetano_2020-04-15
hCoV-19_Brazil_L32_CD825_2020_Brazil_SP_SaoPaulo_2020-04-09
hCoV-19_Brazil_L16_CD384_2020_Brazil_SP_MogGuacu_2020-04-02
hCoV-19_Brazil_L2_Glu_MG01_2020_Brazil_MG_JuizDeFora_2020-05-23
hCoV-19_Brazil_L20_CD420_2020_Brazil_SC_Tubarao_2020-04-01
hCoV-19_Brazil_SC0244_2020-40_Brazil_SC_Criciuma_2020-03-18
hCoV-19_Brazil_L9_CD90_2020_Brazil_SP_SaoPaulo_2020-03-24
hCoV-19_Brazil_L41_CD1011_2020_Brazil_SP_SaoCaetano_2020-04-23
hCoV-19_Brazil_L15_CD281_2020_Brazil_SP_SaoPaulo_2020-03-30
hCoV-19_Brazil_L36_CD900_2020_Brazil_SP_SaoPaulo_2020-05-02
hCoV-19_Brazil_L32_CD822_2020_Brazil_SP_SaoPaulo_2020-04-07
hCoV-19_Brazil_L15_CD286_2020_Brazil_SP_SaoPaulo_2020-03-30
hCoV-19_Brazil_L31_CD801_2020_Brazil_SP_SaoPaulo_2020-04-03
hCoV-19_Brazil_L14_CD258_2020_Brazil_SP_SaoPaulo_2020-03-27
hCoV-19_Brazil_L31_CD795_2020_Brazil_SP_SaoPaulo_2020-04-03
hCoV-19_Brazil_L2_Glu_MG07_2020_Brazil_MG_Extrema_2020-05-25
hCoV-19_Brazil_L6_CD93_2020_Brazil_SP_SaoPaulo_2020-03-18
hCoV-19_Brazil_L38_CD954_2020_Brazil_SP_SaoPaulo_2020-05-02
hCoV-19_Brazil_L21_CD455_2020_Brazil_SP_SaoCaetano_2020-04-24
hCoV-19_Brazil_L6_CD41_2020_Brazil_SP_Taiuape_2020-03-26
hCoV-19_Brazil_L31_CD798_2020_Brazil_SP_SaoPaulo_2020-04-03
hCoV-19_Brazil_L33_CD847_2020_Brazil_SP_SaoPaulo_2020-04-16
hCoV-19_Brazil_L23_CD520_2020_Brazil_SP_SaoPaulo_2020-04-13
hCoV-19_Brazil_L25_CD652_2020_Brazil_SP_SaoCaetano_2020-04-13
hCoV-19_Brazil_L6_CD48_2020_Brazil_SP_SaoPaulo_2020-03-17
hCoV-19_Brazil_L2_Glu_MG15_2020_Brazil_MG_Formiga_2020-04-07
hCoV-19_Brazil_L23_CD520_2020_Brazil_SP_SaoPaulo_2020-04-09
hCoV-19_Brazil_L8_CD104_2020_Brazil_SP_EmbuSodrAtes_2020-03-20
hCoV-19_Brazil_L7_CD66_2020_Brazil_SP_SaoPaulo_2020-03-25
hCoV-19_Brazil_L9_CD38_2020_Brazil_SP_SaoPaulo_2020-03-16
hCoV-19_Brazil_L25_CD626_2020_Brazil_SP_SaoCaetano_2020-04-07
hCoV-19_Brazil_L38_CD997_2020_Brazil_SP_SaoPaulo_2020-05-06
hCoV-19_Brazil_L6_CD56_2020_Brazil_SP_SaoPaulo_2020-03-20
hCoV-19_Brazil_L26_CD548_2020_Brazil_SP_SaoPaulo_2020-04-18
hCoV-19_Brazil_L18_CD685_2020_Brazil_SP_SaoCaetano_2020-04-15
hCoV-19_Brazil_L43_CD132_2020_Brazil_SP_SaoPaulo_2020-05-05
hCoV-19_Brazil_L26_CD557_2020_Brazil_SP_Mairipora_2020-04-20
hCoV-19_Brazil_L3_CAMP23_2020_Brazil_SP_Campinas_2020-04-09
hCoV-19_Brazil_L38_CD1011_2020_Brazil_SP_Osasco_2020-05-10
hCoV-19_Brazil_L25_CD630_2020_Brazil_SP_SaoCaetano_2020-04-07
hCoV-19_Brazil_L22_CD488_2020_Brazil_SP_SaoPaulo_2020-04-05
hCoV-19_Brazil_L32_CD833_2020_Brazil_SP_SaoPaulo_2020-04-09
hCoV-19_Brazil_L32_CD836_2020_Brazil_SP_SaoPaulo_2020-04-13
hCoV-19_Brazil_RJ02/2_2020-75_Brazil_RJ_RioDeJaneiro_2020-04-17
hCoV-19_Brazil_L6_CD36_2020_Brazil_SP_SaoPaulo_2020-03-13
hCoV-19_Brazil_L41_CD1101_2020_Brazil_SP_SaoCaetano_2020-04-20
hCoV-19_Brazil_L34_CD880_2020_Brazil_SP_SaoPaulo_2020-04-23
hCoV-19_Brazil_L31_CD816_2020_Brazil_SP_SaoPaulo_2020-04-08
hCoV-19_Brazil_L38_CD1008_2020_Brazil_SP_SaoPaulo_2020-05-09
hCoV-19_Brazil_L41_CD1066_2020_Brazil_SP_SaoCaetano_2020-04-22
hCoV-19_Brazil_L25_CD638_2020_Brazil_SP_SaoCaetano_2020-04-08
hCoV-19_Brazil_L34_CD875_2020_Brazil_SP_SaoPaulo_2020-04-23
hCoV-19_Brazil_L47_CD1121_2020_Brazil_SP_SaoCaetano_2020
hCoV-19_Brazil_L26_CD566_2020_Brazil_SP_SaoPaulo_2020-04-22
hCoV-19_Brazil_L22_CD477_2020_Brazil_SP_Osasco_2020-04-02
hCoV-19_Brazil_L32_CD839_2020_Brazil_SP_Carapicuiba_2020-04-14
hCoV-19_Brazil_L26_CD558_2020_Brazil_SP_Carapicuiba_2020-04-20
hCoV-19_Brazil_L39_CD1042_2020_Brazil_SP_SaoPaulo_2020-05-13
hCoV-19_Brazil_L37_CD991_2020_Brazil_SP_SaoPaulo_2020-05-05
hCoV-19_Brazil_L41_CD1072_2020_Brazil_SP_SaoCaetano_2020-05-04
hCoV-19_Brazil_L25_CD658_2020_Brazil_SP_SaoCaetano_2020-04-14
hCoV-19_Brazil_L45_CD1096_2020_Brazil_SP_SaoCaetano_2020-04-15
hCoV-19_Brazil_L33_CD867_2020_Brazil_SP_SaoPaulo_2020-04-22
hCoV-19_Brazil_L30_CD785_2020_Brazil_MG_Uberlandia_2020-04-17
hCoV-19_Brazil_L8_CD110_2020_Brazil_SP_SaoPaulo_2020-03-23
hCoV-19_Brazil_L6_CD89_2020_Brazil_SP_SaoPaulo_2020-03-23
hCoV-19_Brazil_L7_CD81_2020_Brazil_SP_SaoPaulo_2020-03-21
hCoV-19_Brazil_L9_CD127_2020_Brazil_SP_SaoPaulo_2020-03-24
hCoV-19_Brazil_L11_CD99_2020_Brazil_SP_SaoPaulo_2020-03-20
hCoV-19_Brazil_L38_CD1002_2020_Brazil_SP_SaoPaulo_2020-05-07
hCoV-19_Brazil_L7_CD83_2020_Brazil_SP_SaoPaulo_2020-03-18
hCoV-19_Brazil_L46_CD1221_2020_Brazil_SP_MogiMirim_2020-05-27
hCoV-19_Brazil_L2_Glu_MG13_2020_Brazil_MG_Uberlandia_2020-04-07
hCoV-19_Brazil_L30_CD779_2020_Brazil_PA_Barcarena_2020-04-16
hCoV-19_Brazil_L40_CD1104_2020_Brazil_SP_SaoCaetano_2020-04-29
hCoV-19_Brazil_L2_Glu_MG04_2020_Brazil_MG_Uberlandia_2020-03-30
hCoV-19_Brazil_L45_CD1161_2020_Brazil_PA_Capanema_2020-05-17
hCoV-19_Brazil_L7_CD81_2020_Brazil_SP_SaoPaulo_2020-03-24
hCoV-19_Brazil_L42_CDPA48_2020_Brazil_PA_PARA_2020-06-09
hCoV-19_Brazil_L22_CD496_2020_Brazil_SP_SaoPaulo_2020-04-05
hCoV-19_Brazil_L2_Glu_MG08_2020_Brazil_MG_PousoAlegre_2020-04-02
hCoV-19_Brazil_L16_CD321_2020_Brazil_PR_Curitiba_2020-03-26
hCoV-19_Brazil_L33_CD842_2020_Brazil_SP_SaoPaulo_2020-04-15
hCoV-19_Brazil_L14_CD246_2020_Brazil_SP_SaoPaulo_2020-03-27
hCoV-19_Brazil_L15_CD266_2020_Brazil_SP_SaoPaulo_2020-03-29
hCoV-19_Brazil_AL-93/_2020_EPI_ISL_427292_2020-03-18
hCoV-19_Brazil_L1_Glu_MG04_2020_Brazil_MG_JuizDeFora_2020-05-17
hCoV-19_Brazil_L33_CD861_2020_Brazil_SP_SaoPaulo_2020-04-22
hCoV-19_Brazil_L39_CD1052_2020_Brazil_SP_Carapicuiba_2020-05-14
hCoV-19_Brazil_L35_CD859_2020_Brazil_SP_SaoPaulo_2020-04-25
hCoV-19_Brazil_L32_CD829_2020_Brazil_SP_SaoPaulo_2020-04-09
Ireland_Dublin-22428_2020_EPI_ISL_418581__Ireland_Dublin__2020-03-08
Australia_VIC638_2020_EPI_ISL_426934__Australia_Victoria__2020-03-29
Japan_P5-3_2020_EPI_ISL_419311__Japan_Chiba__2020-03-13
hCoV-19_Brazil_L33_CD841_2020_Brazil_SP_SaoPaulo_2020-04-15
hCoV-19_Brazil_L2_Glu_MG26_2020_Brazil_MG_Barbacena_2020-05-17
hCoV-19_Brazil_L28_CD615_2020_Brazil_ES_Vitoria_2020-04-09
hCoV-19_Brazil_RJ0279_2020-82_Brazil_RJ_RioDeJaneiro_2020-04-22
hCoV-19_Brazil_L27_CD595_2020_Brazil_RS_BentoGoncalves_2020-04-07
hCoV-19_Brazil_RJ0284_2020-87_Brazil_RJ_RioDeJaneiro_2020-04-24
hCoV-19_Brazil_MG0291_2020-94_Brazil_MG_BeloHorizonte_2020-04-16
hCoV-19_Brazil_MG0017_2020-17_Brazil_MG_BeloHorizonte_2020-03-20
hCoV-19_Brazil_L20_CD416_2020_Brazil_SP_RibeiraoPres_2020-04-02
hCoV-19_Brazil_L21_CD442_2020_Brazil_SP_Bebedouro_2020-04-06
hCoV-19_Brazil_RJ0287_2020-89_Brazil_RJ_RioDeJaneiro_2020-04-15
hCoV-19_Brazil_RJ0250_2020-52_Brazil_RJ_RioDeJaneiro_2020-03-26
hCoV-19_Brazil_RJ0275_2020-78_Brazil_RJ_RioDeJaneiro_2020-04-20
hCoV-19_Brazil_PR0241_2020-42_Brazil_PR_Londrina_2020-04-02
hCoV-19_Brazil_L3_CAMPH41_2020_Brazil_SP_Campinas_2020-04-15
hCoV-19_Brazil_L19_CD411_2020_Brazil_MS_CamoDrande_2020-04-02
hCoV-19_Brazil_L31_CD815_2020_Brazil_SP_SaoPaulo_2020-04-08
hCoV-19_Brazil_L1_CAMPI16_2020_Brazil_SP_Campinas_2020-04-08
hCoV-19_Brazil_L45_CD1160_2020_Brazil_ES_Vitoria_2020-05-20
hCoV-19_Brazil_L2_Glu_MG09_2020_Brazil_MG_Cambui_2020-04-02
hCoV-19_Brazil_MG0258_2020-60_Brazil_MG_BeloHorizonte_2020-04-13
hCoV-19_Brazil_PA0237_2020-38_Brazil_PA_Belem_2020-04-03
hCoV-19_Brazil_PA0235_2020-36_Brazil_PA_Belem_2020-04-03
hCoV-19_Brazil_PI0239_2020-40_Brazil_PI_Teresina_2020-03-19
hCoV-19_Brazil_RJ0274_2020-77_Brazil_RJ_RioDeJaneiro_2020-04-20
hCoV-19_Brazil_L29_CD747_2020_Brazil_BA_Ilheus_2020-04-14
hCoV-19_Brazil_L30_CD737_2020_Brazil_MG_JuizdeFora_2020-04-14
hCoV-19_Brazil_L26_CD543_2020_Brazil_SP_SaoPaulo_2020-04-17
hCoV-19_Brazil_L20_CD436_2020_Brazil_SP_Itapetininga_2020-04-06
hCoV-19_Brazil_L46_CD1177_2020_Brazil_ES_Vitoria_2020-05-18
hCoV-19_Brazil_L20_CD434_2020_Brazil_SP_SaoPaulo_2020-04-04
hCoV-19_Brazil_RJ0262_2020-64_Brazil_RJ_RioDeJaneiro_2020-04-09
hCoV-19_Brazil_L47_CD1125_2020_Brazil_SP_SaoCaetano_2020
hCoV-19_Brazil_RJ0259_2020-61_Brazil_RJ_RioDeJaneiro_2020-04-07
hCoV-19_Brazil_RJ0247_2020-49_Brazil_RJ_RioDeJaneiro_2020-03-23
hCoV-19_Brazil_RJ0256_2020-58_Brazil_RJ_RioDeJaneiro_2020-04-03
hCoV-19_Brazil_RJ0257_2020-59_Brazil_RJ_RioDeJaneiro_2020-04-06
hCoV-19_Brazil_L27_CD682_2020_Brazil_ES_Vitoria_2020-04-08
hCoV-19_Brazil_L28_CD614_2020_Brazil_CE_Fortaleza_2020-04-08
hCoV-19_Brazil_RJ0266_2020-70_Brazil_RJ_RioDeJaneiro_2020-04-15
hCoV-19_Brazil_L45_CD1148_2020_Brazil_AC_RioBranco_2020-05-15
hCoV-19_Brazil_MG0289_2020-92_Brazil_MG_BeloHorizonte_2020-04-08
hCoV-19_Brazil_L2_Glu_MG10_2020_Brazil_MG_Cambui_2020-04-02
hCoV-19_Brazil_L34_CD881_2020_Brazil_SP_SaoPaulo_2020-04-23
hCoV-19_Brazil_L2_Glu_MG18_2020_Brazil_MG_Extrema_2020-04-08
hCoV-19_Brazil_L17_CD369_2020_Brazil_RS_BentoGoncalves_2020-03-31
hCoV-19_Brazil_L29_CD737_2020_Brazil_RS_Alegrete_2020-04-13
hCoV-19_Brazil_MG0221_2020-21_Brazil_MG_BeloHorizonte_2020-04-05
hCoV-19_Brazil_L2_Glu_MG04_2020_Brazil_MG_JuizDeFora_2020-05-24
hCoV-19_Brazil_L1_Glu_MG04_2020_Brazil_MG_JuizDeFora_2020-05-24
hCoV-19_Brazil_L17_CD356_2020_Brazil_SP_Sorocaba_2020-03-31
hCoV-19_Brazil_L10_CD141_2020_Brazil_SP_RibeiraoPreto_2020-03-14
hCoV-19_Brazil_L2_Glu_MG22_2020_Brazil_MG_JuizDeFora_2020-04-15
hCoV-19_Brazil_L2_Glu_MG38_2020_Brazil_MG_JuizDeFora_2020-05-18
hCoV-19_Brazil_L45_CD1172_2020_Brazil_MG_JuizDeFora_2020-05-18
hCoV-19_Brazil_L1_Glu_MG38_2020_Brazil_MG_JuizDeFora_2020-05-18
hCoV-19_Brazil_RJ0263_2020-65_Brazil_RJ_RioDeJaneiro_2020-04-13
hCoV-19_Brazil_L12_CD192_2020_Brazil_RJ_Niteroi_2020-03-24
hCoV-19_Brazil_DF-619/_2020_EPI_ISL_427295_2020-03-13
hCoV-19_Brazil_L11_CD183_2020_Brazil_MG_Uberlandia_2020-03-16
hCoV-19_Brazil_L10_CD158_2020_Brazil_RJ_RioDeJaneiro_2020-03-16
hCoV-19_Brazil_L46_CD1217_2020_Brazil_SP_Guarulhos_2020-05-26
hCoV-19_Brazil_DFBR-0001-_2020_EPI_ISL_426080_2020-03-13
hCoV-19_Brazil_CVH2_2020_EPI_ISL_429695_2020-03-20
hCoV-19_Brazil_CV32_2020_EPI_ISL_429688_2020-03-17
hCoV-19_Brazil_SPBR-06_2020_EPI_ISL_414015_2020-02-29
hCoV-19_Brazil_L2_Glu_MG05_2020_Brazil_MG_PousoAlegre_2020-04-09
hCoV-19_Brazil_L24_CD540_2020_Brazil_SP_Guarulhos_2020-04-15
hCoV-19_Brazil_MG0012_2020-12_Brazil_MG_BeloHorizonte_2020-03-14
hCoV-19_Brazil_L2_Glu_MG06_2020_Brazil_MG_Uberaba_2020-03-31
France_HF2405_2020_EPI_ISL_418228__France_Hauts_de_France_Compi___gne_2020-03-12
Australia_VIC48_2020_EPI_ISL_413162__Australia_Victoria__2020-03-12
France_ARA1322_2020_EPI_ISL_420625__France_ARA_Lyon_2020-03-24
hCoV-19_Brazil_L2_Glu_MG42_2020_Brazil_MG_JuizDeFora_2020-05-19
hCoV-19_Brazil_L30_CD766_2020_Brazil_MG_JuizdeFora_2020-04-15
hCoV-19_Brazil_L29_CD720_2020_Brazil_MG_JuizdeFora_2020-04-11
hCoV-19_Brazil_L2_Glu_MG16_2020_Brazil_MG_SantosDumont_2020-04-05
hCoV-19_Brazil_L2_Glu_MG31_2020_Brazil_MG_JuizDeFora_2020-05-15
Taiwan_NTU03_2020_EPI_ISL_413292__Taiwan_Taipei__2020-03-02
Canada_ON_PHL81/1_2020_EPI_ISL_418345__Canada_Ontario__2020-02-29
hCoV-19_Brazil_L34_CD883_2020_Brazil_SP_SaoPaulo_2020-04-23
hCoV-19_Brazil_L4_CD12_2020_Brazil_RJ_2020-03-10
USA_WI-123_2020_EPI_ISL_427430__USA_Wisconsin__2020-04-11
hCoV-19_Brazil_L13_CD31_2020_Brazil_SP_2020-03-10
France_Lyon_0668_2020_EPI_ISL_417356__France_ARA__2020-03-06
France_B1623_2020_EPI_ISL_418219__France_Bretagne_Brest_2020-02-26
hCoV-19_Brazil_CV29_2020_EPI_ISL_429686_2020-03-17
Germany_BavPat1_2020_EPI_ISL_406862__Germany_Bavaria_Munich_2020-01-28
Scotland_CVR141_2020_EPI_ISL_425685__UK_Scotland__2020-03-18

ANEXOS

Artigos submetidos e/ou publicados

Cite as: D. S. Candido *et al.*, *Science* 10.1126/science.abd2161 (2020).

Evolution and epidemic spread of SARS-CoV-2 in Brazil

Darlan S. Candido[1,2*], Ingra M. Claro[2,3*], Jaqueline G. de Jesus[2,3*], William M. Souza[4*], Filipe R. R. Moreira[5*], Simon Dellicour[6,7*], Thomas A. Mellan[8*], Louis du Plessis[1], Rafael H. M. Pereira[9], Flavia C. S. Sales[2,3], Erika R. Manuli[2,3], Julien Thézé[10], Luiz Almeida[11], Mariane T. Menezes[5], Carolina M. Voloch[5], Marcilio J. Fumagalli[4], Thaís M. Coletti[2,3], Camila A. M. da Silva[2,3], Mariana S. Ramundo[2,3], Mariene R. Amorim[12], Henrique H. Hoeltgebaum[13], Swapnil Mishra[8], Mandev S. Gill[7], Luiz M. Carvalho[14], Lewis F. Buss[2], Carlos A. Prete Jr[15], Jordan Ashworth[16], Helder I. Nakaya[17], Pedro S. Peixoto[18], Oliver J. Brady[19,20], Samuel M. Nicholls[21], Amilcar Tanuri[5], Átila D. Rossi[5], Carlos K.V. Braga[9], Alexandra L. Gerber[11], Ana Paula de C. Guimarães[11], Nelson Gaburo Jr[22], Cecila Salete Alencar[23], Alessandro C.S. Ferreira[24], Cristiano X. Lima[25,26], José Eduardo Levi[27], Celso Granato[28], Giulia M. Ferreira[29], Ronaldo S. Francisco Jr[11], Fabiana Granja[12,30], Marcia T. Garcia[31], Maria Luiza Moretti[31], Mauricio W. Perroud Jr[32], Terezinha M. P. P. Castiñeiras[33], Carolina S. Lazari[34], Sarah C. Hill[1,35], Andreza Aruska de Souza Santos[36], Camila L. Simeoni[12], Julia Forato[12], Andrei C. Sposito[37], Angelica Z. Schreiber[38], Magnun N. N. Santos[38], Camila Zolini de Sá[39], Renan P. Souza[39], Luciana C. Resende-Moreira[40], Mauro M. Teixeira[41], Josy Hubner[42], Patricia A. F. Leme[43], Rennan G Moreira[44], Maurício L. Nogueira[45], Brazil-UK Centre for Arbovirus Discovery, Diagnosis, Genomics and Epidemiology (CADDE) Genomic Network, Neil M Ferguson[8], Silvia F. Costa[2,3], José Luiz Proenca-Modena[12], Ana Tereza R. Vasconcelos[11], Samir Bhatt[8], Philippe Lemey[7], Chieh-Hsi Wu[46], Andrew Rambaut[47], Nick J. Loman[21], Renato S. Aguiar[39], Oliver G. Pybus[1], Ester C. Sabino[2,3]†, Nuno Rodrigues Faria[1,2,8]†

[1]Department of Zoology, University of Oxford, Oxford, UK. [2]Instituto de Medicina Tropical, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil. [3]Departamento de Moléstias Infecciosas e Parasitárias, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil. [4]Centro de Pesquisa em Virologia, Faculdade de Medicina de Ribeirão Preto, Ribeirão Preto, Brazil. [5]Departamento de Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil. [6]Spatial Epidemiology Lab, Université Libre de Bruxelles, Brussels, Belgium. [7] Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium. [8]MRC Centre for Global Infectious Disease Analysis, J-IDEA, Imperial College London, London, UK. [9]Institute for Applied Economic Research, Brasília, Brazil. [10]Université Clermont Auvergne, INRAE, VetAgro Sup, UMR EPIA, Saint-Genès-Champanelle, France. [11]Laboratório de Bioinformática, Laboratório Nacional de Computação Científica, Petrópolis, Brazil. [12]Departamento de Genética, Evolução, Microbiologia e Imunologia, Instituto de Biologia and Experimental Medicine Research Cluster (EMRC), Universidade Estadual de Campinas, Campinas, Brazil. [13]Department of Mathematics, Imperial College London, London, UK. [14]Escola de Matemática Aplicada (EMAp), Fundação Getúlio Vargas, Rio de Janeiro, Brazil. [15]Department of Electronic Systems Engineering, University of São Paulo, São Paulo, Brazil. [16]Usher Institute, University of Edinburgh, Edinburgh, UK. [17]Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil. [18]Departamento de Matemática Aplicada, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil. [19]Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK. [20]Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London, UK. [21]Institute for Microbiology and Infection, University of Birmingham, Birmingham, UK. [22]DB Diagnósticos do Brasil, São Paulo, Brazil. [23]LIM 03 Laboratório de Medicina Laboratorial, Hospital das Clínicas Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil. [24]Instituto Hermes Pardini, Belo Horizonte, Brazil. [25]Departamento de Cirurgia, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil. [26]Simile Instituto de Imunologia Aplicada Ltda, Belo Horizonte, Brazil. [27]Laboratório DASA, São Paulo, Brazil. [28]Laboratório Fleury, São Paulo, Brazil. [29]Laboratório de Virologia, Instituto de Ciências Biomédicas, Universidade Federal de Uberlândia, Uberlândia, Brazil. [30]Centro de Estudos da Biodiversidade, Universidade Federal de Roraima, Boa Vista, Brazil. [31]Divisão de Doenças Infecciosas, Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, Brazil. [32]Hospital Estadual Sumaré, Universidade Estadual de Campinas, Campinas, Brazil. [33]Departamento de Doenças Infecciosas e Parasitárias, Faculdade de Medicina, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil. [34]Divisão de Laboratório Central do Hospital das Clínicas, da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil. [35]Department of Pathobiology and Population Sciences, Royal Veterinary College, Hatfield, UK. [36]University of Oxford, Latin American Centre, Oxford School of Global and Area Studies, Oxford, UK. [37]Departamento de Clínica Médica, Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, Brazil. [38]Departamento de Patologia Clínica, Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, Brazil. [39]Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil. [40]Departamento de Botânica, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil. [41]Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil. [42]Departamento de Biologia Celular, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil. [43]Centro de Saúde da Comunidade, Universidade Estadual de Campinas, Campinas, Brazil. [44]Centro de Laboratórios Multiusuários, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil. [45]Laboratório de Pesquisas em Virologia, Faculdade de Medicina de São José do Rio Preto, São José do Rio Preto, São Paulo, Brazil. [46]Mathematical Sciences, University of Southampton, Southampton, UK. [47]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK.

***These authors contributed equally to this work.**

†Corresponding author. Email: sabinoec@usp.br (E.C.S.); nuno.faria@zoo.ox.ac.uk (N.R.F.)

Brazil currently has one of the fastest growing SARS-CoV-2 epidemics in the world. Owing to limited available data, assessments of the impact of non-pharmaceutical interventions (NPIs) on virus spread remain challenging. Using a mobility-driven transmission model, we show that NPIs reduced the reproduction number from >3 to 1–1.6 in São Paulo and Rio de Janeiro. Sequencing of 427 new genomes and analysis of a geographically representative genomic dataset identified >100 international virus introductions in Brazil. We estimate that most (76%) of the Brazilian strains fell in three clades that were introduced from Europe between 22 February–11 March 2020. During the early epidemic phase, we found that SARS-CoV-2 spread mostly locally and within-state borders. After this period, despite sharp decreases in air travel, we estimated multiple exportations from large urban centers that coincided with a 25% increase in average travelled distances in national flights. This study sheds new light on the epidemic transmission and evolutionary trajectories of SARS-CoV-2 lineages in Brazil and provides evidence that current interventions remain insufficient to keep virus transmission under control in the country.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel betacoronavirus with a 30-kb genome that was first reported in December 2019 in Wuhan, China (*1*, *2*). SARS-CoV-2 was declared a public health emergency of international concern on 30 January 2020. As of 12 July 2020, coronavirus disease 2019 (COVID-19) has caused over 12.5 million cases and 561 thousand deaths globally (*3*). The virus can be classified into two main phylogenetic lineages, namely A and B, that spread from Wuhan before strict travel restrictions were enacted (*4*, *5*) and now co-circulate around the world (*6*). The case fatality ratio of SARS-CoV-2 infection has been estimated between 1.2 and 1.6% (*7–9*) with substantially higher ratios in those aged above 60 years (*8*). Some estimates suggest that 18-56% of SARS-CoV-2 transmission is from asymptomatic or pre-symptomatic individuals (*10–13*), complicating epidemiological assessments and public health efforts to curb the pandemic.

Challenges of real-time assessment of transmission

While the SARS-CoV-2 epidemics in several countries including China, Italy, and Spain have been brought under control through non-pharmaceutical interventions (NPIs) (*3*), the number of SARS-CoV-2 cases and deaths in Brazil continues to increase (*14*) (Fig. 1A). As of 12 July 2020, Brazil has now reported 1,800,827 SARS-CoV-2 cases, the second largest number in the world, and 70,398 deaths. Over a third of the cases (34%) in Brazil are concentrated in the southeast region which includes São Paulo city (Fig. 1B), the world's fourth largest conurbation, where the first case in Latin America was reported on 25 February 2020 (*15*). Diagnostic assays for SARS-CoV-2 molecular detection were widely distributed across the regional reference centres of the national public health laboratory network from 21 February 2020 onwards (*16*, *17*). However, several factors, including delays in reporting, changes in notification, and heterogeneous access to testing across populations, obfuscate the real-time assessment of virus transmission using SARS-CoV-2 case counts (*15*). Consequently, a more accurate measure of SARS-CoV-2 transmission in Brazil is provided by reported deaths due to severe acute respiratory infections (SARI), provided by the Sistema Único de Saúde (SUS) (*18*). Changes in the opportunity for SARS-CoV-2 transmission are strongly associated with changes in average mobility (*18–20*), and can typically be measured by calculating the effective reproduction number, R, defined as the average number of secondary infections caused by an infected person. R >1 indicates a growing epidemic while R <1 is needed to achieve a decrease in transmission.

We used a Bayesian semi-mechanistic model (*21*, *22*) to analyze SARI mortality statistics and human mobility data to estimate daily changes in R in São Paulo city (12,2 million inhabitants) and Rio de Janeiro city (6,7 million inhabitants), the largest urban metropoles in Brazil (Fig. 1, C and D). NPIs in Brazil consisted of school closures implemented between 12 and 23 March 2020 across the country's 27 federal units/states, and store closures implemented between 13 and 23 March 2020. In São Paulo city, schools started closing on 16 March and stores closed four days later. At the start of the epidemics, we found R >*3* in São Paulo and Rio de Janeiro, and that concurrent with the timing of state mandated NPIs, R values fell close to 1.

Mobility driven reproduction number changes

Analysis of the reproduction number after NPI implementation highlights several notable mobility-driven features. There was a period immediately following NPIs, between 21 and 31 March 2020, when R was consistently <1 in São Paulo city (Fig. 1C). However, after this initial decrease, the R value for São Paulo rises >1 and increases through time, a trend associated with increased population mobility. This can be seen in the Google transit stations index, which rises from -60% to -52%, and by a decrease in the social isolation index from 54% to 47%. By 4 May 2020, we estimate R = 1.3 (BCI 95%: 1.0-1.6) in both São Paulo and Rio de Janeiro cities (table S1). However, we note that there were instances in the previous 7 days when the 95% credible intervals for R included values below 1, drawing attention to the fluctuations and uncertainty in the estimated reproduction number for both cities.

Early sharing of genomic sequences, including the first SARS-CoV-2 genome, Wuhan-Hu-1, released on 10 January (*23*), has enabled unprecedented global levels of molecular testing for an emerging virus (*24*, *25*). However, despite the thousands of virus genomes deposited on public access databases, there is a lack of consistent sampling structure, and limited data from Brazil (*26–28*), which hampers accurate reconstructions of virus movement and transmission using phylogenetic analyses. To investigate how SARS-CoV-2 became established in the country, and to quantify the impact of NPIs on virus spatiotemporal spread, we tested a total of 26,732 samples from public and private laboratories using real-time PCR assays and found 7,944 (29%) to be positive for SARS-CoV-2. We then focused our sequencing efforts on generating a large and spatially representative genomic dataset with curated metadata in order to maximise the association between the number of sequences and the number of SARS-CoV-2 confirmed cases per state.

Spatially representative sequencing efforts

We generated 427 new SARS-CoV-2 genomes with >75% genome coverage from Brazilian samples collected between 5 March and 30 April 2020 (figs. S1 to S3 and data S1). For each state, the time between the date of the first reported case and the collection date of the first sequence analyzed in that state was only 4.5 days on average (Fig. 2A). For eight federal states, genomes were obtained from samples collected up to 6 days before the first case notifications. The genomes generated here were collected in 85 municipalities across 18 of 27 federal units spanning all regions in Brazil (Fig. 2A and fig. S2). Sequenced genomes were obtained from samples collected on average 4 days (median, range: 0 to 29 days) after onset of symptoms and were generated in 3 laboratories using harmonized sequencing and bioinformatic protocols (table S2). When we include 63 additional available sequences from Brazil deposited in GISAID (*29*) (see data S1 and S2), we find the data set is representative of the spatial heterogeneity of the Brazilian epidemic. Specifically, the number of genomes per state strongly correlates with SARI SARS-CoV-2 confirmed cases and SARI cases with unknown aetiology per state (n = 490 sequences from 21 states, Spearman's correlation, ρ = 0.83; Fig. 2A). This correlation varied from 0.70 to 0.83 when considering SARI cases and deaths caused by SARS-CoV-2, and SARI cases and deaths from unknown aetiology (fig. S4). Most (n = 485/490) Brazilian sequences belong to SARS-CoV-2 lineage B, with only 5 strains belonging to lineage A (2 from Amazonas, 1 from Rio Grande do Sul, 1 from Minas Gerais and 1 from Rio de Janeiro; data S1 and fig. S5 show detailed lineage information for each sequence). Moreover, we used an *in-silico* assessment of diagnostic assay specificity for Brazilian strains (n = 490) to identify potential mismatches in some assays targeting Brazilian strains. We find that the forward primers of the Chinese CDC and Hong Kong University nucleoprotein-targeting RT-qPCR may be less appropriate for use in Brazil than other diagnostic assays, for which few or no mismatches were identified (fig. S6 and table S3). The impact of these mismatches on the sensitivity of these assays should be confirmed experimentally. If sensitivity is affected, the use of duplex RT-qPCR assays that concurrently target different genomic regions may help the detection of viruses with variants in primer or probe binding regions.

Phylogenetic analyses and international introductions

We estimated maximum likelihood and molecular clock phylogenies for a global dataset with a total of 1,182 genomes sampled from 24 Dec 2019 to 30 Apr 2020 (root-to-tip genetic distance correlation with sampling dates, r^2 = 0.53; Fig. 3A and fig. S7). We inferred a median evolutionary rate of 1.13×10^{-3} (95% BCI: $1.03–1.23 \times 10^{-3}$) substitutions per site per year (s/s/y), using an exponential growth coalescent model, equating to 33 changes per year on average across the virus genome. This is within the range of evolutionary rates estimated for other human coronaviruses (*30–33*). We estimate the date of the common ancestor (TMRCA) of the SARS-CoV-2 pandemic to around mid-Nov 2019 (median = 19 Nov 2019, 95% BCI: 26 Oct 2019–6 Dec 2019), in line with recent findings (*34*, *35*).

Phylogenetic analysis revealed that the majority of the Brazilian genomes (76%, n = 370/490) fell into three clades hereafter named as Clade 1 (n = 186/490, 38% of Brazilian strains), Clade 2 (n = 166, 34%) and Clade 3 (n = 18/490, 4%; Fig. 3A and figs. S8 and S9), which were largely in agreement with those identified in a phylogenetic analysis using 13,833 global genomes. The most recent common ancestors of the three main Brazilian clades (Clades 1 to 3) were dated from 28 February (21 Feb to 4 Mar 2020) (*Clade 1*), 22 February (17 to 24 Feb 2020) (Clade 2) to 11 March (9 to 12 Mar 2020) (Clade 3) (Fig. 3A and fig. S10). This indicates that community-driven transmission was already established in Brazil by early March, suggesting that international travel restrictions initiated after this period would have had limited impact. Brazilian Clade 1 is characterized by a nucleotide substitution in the spike protein (G25088T, numbering relative to GenBank reference NC_045512.2) and circulates predominantly in São Paulo state (n = 159, 85.4%; figs. S9 and S11). Clade 2 is defined by two nucleotide substitutions in ORF6 (T27299C) and nucleoprotein (T29148C); this is the most spatially widespread lineage, with sequences from a total of 16 states in Brazil. Clade *3* is concentrated in Ceará state (n = 16, 89%) and falls in a global cluster with sequences mainly from Europe. In the Amazon region, where the epidemic is expanding rapidly (*14, 22*), we find evidence for multiple national and international introductions, with 37% (n = 7/19) of sequences

from Pará and Amazonas states clustering in Clade 1 and 32% (n = 6/19) in Clade 2.

Time-measured phylogeographic analyses revealed at least 102 (95% BCI: 95–109) international introductions of SARS-CoV-2 in Brazil (Fig. 3A and figs. S8 and S12). This represents an underestimate of the real number of introductions, as we have sequenced, on average, only 1 out of 200 confirmed cases. Most of these estimated introductions were directed to internationally well-connected states (*36*) such as São Paulo (36% of all imports), Minas Gerais (24%), Ceará (10%) and Rio de Janeiro (8%) (fig. S12). We further assessed the contribution of international vs. national virus lineage movement events through time (Fig. 3B). In the first phase of the epidemic, we find an increasing number of international introductions until 10 Mar 2020 (Fig. 2B). Limited available travel history data (*15*) suggests that these early cases were predominantly acquired from Italy (26%, n = 70 of 266 unambiguously identified country of infection) and the USA (28%, n = 76 of 266). After this initial phase, we find that the estimated number of international imports decreased concomitantly with the decline in the number of international passengers travelling to Brazil (Fig. 3, B and C, and S13). In contrast, despite the declines in the number of passengers travelling on national flights (Fig. 3C), we detected an increase in virus lineage movement events between Brazilian regions at least until early April 2020.

Modelling spatiotemporal spread within Brazil

To better understand virus spread across spatiotemporal scales within Brazil, we use a continuous phylogeographic model that maps phylogenetic nodes to their inferred origin locations (*37*) (Fig. 4). We distinguish branches that remain within a state versus those that cross a state to infer the proportion of within versus between state observed virus movement.

We estimate that during the first epidemic phase, SARS-CoV-2 spread mostly locally and within-state borders. In contrast, the second phase was characterized by long-distance movement events and the ignition of the epidemic outside the southeast region of Brazil (Fig. 4A). Throughout the epidemic, we find that within-state virus lineage movement was, on average, 5.1-fold more frequent than between-state movement. Moreover, our data suggests that within-state virus spread, and to a lesser extent, between-state virus spread, decreased after the implementation of NPIs (Fig. 4B). However, it is useful to note that the more limited sampling after April 6 2020 (see fig. S2) decreases inferred virus lineage movement toward present (Figs. 3B and 4B).

Interestingly, we find that the average route length travelled by passenger increased by 25% during the second phase of the epidemic (Fig. 4C), despite a concomitant reduction in the number of passengers flying within Brazil (Fig. 3C). The increase in the average route length post-NPI implementation results from a larger reduction in the number of air passengers flying on shorter distance journeys compared to those flying longer distance journeys. For example, we find an 8.8-fold reduction in the number of passengers flying in flight legs < 1000 km, compared to a 4.4-fold reduction in those flying >2000 km (fig. S15). These findings emphasize the roles of within and between-state mobility as a key driver of both local and inter-regional virus spread, with highly populated and well-connected urban conurbations in the southeast region acting as main sources of virus exports within the country (fig. S12).

Discussion

We provide a comprehensive analysis of SARS-CoV-2 spread in Brazil that shows the importance of community and nation-wide measures to control the COVID-19 epidemic Brazil. Although NPIs initially reduced virus transmission and spread, the continued increase in the number of cases and deaths in Brazil highlights the urgent need to prevent future virus transmission by implementing rapid and accessible diagnostic screening, contact tracing, quarantining of new cases and coordinated social and physical distancing measures across the country (*38*). With the recent relaxation of NPIs in Brazil and elsewhere, continued molecular, immunological and genomic surveillance are required for real-time data-driven decisions. Our analysis shows how changes in mobility may impact global and local transmission of SARS-CoV-2, and demonstrates how combining genomic and mobility data can complement traditional surveillance approaches.

REFERENCES AND NOTES

1. K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, R. F. Garry, The proximal origin of SARS-CoV-2. *Nat. Med.* **26**, 450–452 (2020). doi:10.1038/s41591-020-0820-9 Medline
2. F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E. C. Holmes, Y.-Z. Zhang, A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020). doi:10.1038/s41586-020-2008-3 Medline
3. World Health Organization, *Coronavirus Disease (COVID-2019) Situation Reports* (2020); www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.
4. H. Tian, Y. Liu, Y. Li, C.-H. Wu, B. Chen, M. U. G. Kraemer, B. Li, J. Cai, B. Xu, Q. Yang, B. Wang, P. Yang, Y. Cui, Y. Song, P. Zheng, Q. Wang, O. N. Bjornstad, R. Yang, B. T. Grenfell, O. G. Pybus, C. Dye, An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* **368**, 638–642 (2020). doi:10.1126/science.abb6105 Medline
5. M. U. G. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, L. du Plessis, N. R. Faria, R. Li, W. P. Hanage, J. S. Brownstein, M. Layan, A. Vespignani, H. Tian, C. Dye, O. G. Pybus, S. V. Scarpino; Open COVID-19 Data Working Group, The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020). doi:10.1126/science.abb4218 Medline
6. A. Rambaut, E. C. Holmes, Á. O'Toole, V. Hill, J. T. McCrone, C. Ruis, L. du Plessis, O. G. Pybus, A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* (2020). doi:10.1038/s41564-020-0770-5 Medline
7. T. W. Russell, J. Hellewell, C. I. Jarvis, K. van Zandvoort, S. Abbott, R. Ratnayake, S. Flasche, R. M. Eggo, W. J. Edmunds, A. J. Kucharski; Cmmid Covid-Working Group,

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. *Euro Surveill.* **25**, 2000256 (2020). doi:10.2807/1560-7917.ES.2020.25.12.2000256 Medline

8. R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. T. Walker, H. Fu, A. Dighe, J. T. Griffin, M. Baguelin, S. Bhatia, A. Boonyasiri, A. Cori, Z. Cucunubá, R. FitzJohn, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. van Elsland, E. Volz, H. Wang, Y. Wang, X. Xi, C. A. Donnelly, A. C. Ghani, N. M. Ferguson, Estimates of the severity of coronavirus disease 2019: A model-based analysis. *Lancet Infect. Dis.* **20**, 669–677 (2020). doi:10.1016/S1473-3099(20)30243-7 Medline
9. J. T. Wu, K. Leung, M. Bushman, N. Kishore, R. Niehus, P. M. de Salazar, B. J. Cowling, M. Lipsitch, G. M. Leung, Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat. Med.* **26**, 506–510 (2020). doi:10.1038/s41591-020-0822-7 Medline
10. M. M. Arons, K. M. Hatfield, S. C. Reddy, A. Kimball, A. James, J. R. Jacobs, J. Taylor, K. Spicer, A. C. Bardossy, L. P. Oakley, S. Tanwar, J. W. Dyal, J. Harney, Z. Chisty, J. M. Bell, M. Methner, P. Paul, C. M. Carlson, H. P. McLaughlin, N. Thornburg, S. Tong, A. Tamin, Y. Tao, A. Uehara, J. Harcourt, S. Clark, C. Brostrom-Smith, L. C. Page, M. Kay, J. Lewis, P. Montgomery, N. D. Stone, T. A. Clark, M. A. Honein, J. S. Duchin, J. A. Jernigan; Public Health–Seattle and King County and CDC COVID-19 Investigation Team, Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. *N. Engl. J. Med.* **382**, 2081–2090 (2020). doi:10.1056/NEJMoa2008457 Medline
11. L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, C. Fraser, Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936 (2020). doi:10.1126/science.abb6936 Medline
12. E. Lavezzo, E. Franchin, C. Ciavarella, G. Cuomo-Dannenburg, L. Barzon, C. Del Vecchio, L. Rossi, R. Manganelli, A. Loregian, N. Navarin, D. Abate, M. Sciro, S. Merigliano, E. De Canale, M. C. Vanuzzo, V. Besutti, F. Saluzzo, F. Onelia, M. Pacenti, S. Parisi, G. Carretta, D. Donato, L. Flor, S. Cocchio, G. Masi, A. Sperduti, L. Cattarino, R. Salvador, M. Nicoletti, F. Caldart, G. Castelli, E. Nieddu, B. Labella, L. Fava, M. Drigo, K. A. M. Gaythorpe, A. R. Brazzale, S. Toppo, M. Trevisan, V. Baldo, C. A. Donnelly, N. M. Ferguson, I. Dorigatti, A. Crisanti, Suppression of COVID-19 outbreak in the municipality of Vo, Italy. *Nature* Medline (2020). doi:10.1038/s41586-020-2488-1
13. K. Mizumoto, K. Kagaya, A. Zarebski, G. Chowell, Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Euro Surveill.* **25**, 2000180 (2020). doi:10.2807/1560-7917.ES.2020.25.10.2000180 Medline
14. Brazilian Ministry of Health, *Painel de Casos de Doença Pelo Coronavírus 2019 (COVID-19) No Brasil Pelo Ministério da Saúde* (2020); http://covid.saude.gov.br.
15. W. M. de Souza, L. Fletcher Buss, D. da Silva Candido, J. P. Carrera, S. Li, A. Zarebski, M. Vincenti-Gonzalez, J. Messina, F. C. da Silva Sales, P. dos Santos Andrade, C. A. Prete Jr., V. H. Nascimento, F. Ghilardi, R. H. Moraes Pereira, A. A. de Souza Santos, L. Abade, B. Gutierrez, M. U. G. Kraemer, R. Santana Aguiar, N. Alexander, P. Mayaud, O. J. Brady, I. O. M. de Souza, N. Gouveia, G. Li, A. Tami, S. Barbosa Oliveira, V. B. Gomes Porto, F. Ganem, W. Ferreira Almeida, F. Fontana Sutile Tardetti Fantinato, E. Marques Macario, W. Kleber Oliveira, O. Pybus, C.-H. Wu, J. Croda, E. Cerdeira Sabino, N. R. Faria, Epidemiological and clinical characteristics of the early phase of the COVID-19 epidemic in Brazil. medRxiv 10.1101/2020.04.25.20077396 [Preprint]. 29 April 2020; https://doi.org/10.1101/2020.04.25.20077396 .
16. J. Croda, W. K. Oliveira, R. L. Frutuoso, L. H. Mandetta, D. C. Baia-da-Silva, J. D. Brito-Sousa, W. M. Monteiro, M. V. G. Lacerda, COVID-19 in Brazil: Advantages of a socialized unified health system and preparation to contain cases. *Rev. Soc. Bras. Med. Trop.* **53**, e20200167 (2020). doi:10.1590/0037-8682-0167-2020 Medline
17. J. Croda, L. Garcia, Immediate Health Surveillance Response to COVID-19 Epidemic [in Portuguese]. *Epidemiol. Ser. Saúde* **29**, e2020002 (2020). doi:10.5123/S1679-49742020000100021 Medline
18. S. B. Oliveira, V. Bertollo Gomes Porto, F. Ganem, F. Macedo Mendes, M. Almiron, W. Kleber de Oliveira, F. Fontana Sutile Tardetti Fantinato, W. Aparecida Ferreira de Almeida, A. Pereira de Macedo Borges, H. Natan Batista Pinheiro, R. dos Santos Oliveira, J. R. Andrews, N. R. Faria, M. Barreto Lopes, W. Araujo, F. A. Diaz-Quijano, H. I. Nakaya, J. Croda, Monitoring social distancing and SARS-CoV-2 transmission in Brazil using cell phone mobility data. medRxiv 2020.04.30.20082172 [Preprint] (5 May 2020); https://doi.org/10.1101/2020.04.30.20082172.
19. S. M. Kissler, Reductions in commuting mobility predict geographic differences in SARS-CoV-2 prevalence in New York City (Harvard DASH Repository, 2020); https://dash.harvard.edu/bitstream/handle/1/42665370/Kissler_etal_NYC_mobility.pdf?sequence=1&isAllowed=y.
20. H. J. T. Unwin, S. Mishra, V. C. Bradley, A. Gandy, M. Vollmer, T. Mellan, H. Coupland, K. Ainslie, C. Whittaker, J. Ish-Horowicz, S. Filippi, X. Xi, M. Monod, O. Ratmann, M. Hutchinson, F. Valka, H. Zhu, I. Hawryluk, P. Milton, M. Baguelin, A. Boonyasiri, N. Brazeau, L. Cattarino, G. Charles, L. V. Cooper, Z. Cucunuba, G. Cuomo-Dannenburg, B. Djaafara, I. Dorigatti, O. J. Eales, J. Eaton, S. van Elsland, R. FitzJohn, K. Gaythorpe, W. Green, T. Hallett, W. Hinsley, N. Imai, B. Jeffrey, E. Knock, D. Laydon, J. Lees, G. Nedjati-Gilani, P. Nouvellet, L. Okell, A. Ower, K. V. Parag, I. Siveroni, H. A. Thompson, R. Verity, P. Walker, C. Walters, Y. Wang, O. J. Watson, L. Whittles, A. Ghani, N. M. Ferguson, S. Riley, C. A. Donnelly, S. Bhatt, S. Flaxman, *Report 23: State-Level Tracking of COVID-19 in the United States (21-05-2020)* (Imperial College London, 2020); https://doi.org/10.25561/79231.
21. S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J. W. Eaton, M. Monod, A. C. Ghani, C. A. Donnelly, S. Riley, M. A. C. Vollmer, N. M. Ferguson, L. C. Okell, S. Bhatt; Imperial College COVID-19 Response Team, Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* (2020). doi:10.1038/s41586-020-2405-7 Medline
22. T. A. Mellan, H. H. Hoeltgebaum, S. Mishra, C. Whittaker, R. P. Schnekenberg, A. Gandy, H. J. T. Unwin, M. A. C. Vollmer, H. Coupland, I. Hawryluk, N. Rodrigues Faria, J. Vesga, H. Zhu, M. Hutchinson, O. Ratmann, M. Monod, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, N. Brazeau, G. Charles, L. V. Cooper, Z. Cucunuba, G. Cuomo-Dannenburg, A. Dighe, B. Djaafara, J. Eaton, S. L. van Elsland, R. FitzJohn, K. Fraser, K. Gaythorpe, W. Green, S. Hayes, N. Imai, B. Jeffrey, E. Knock, D. Laydon, J. Lees, T. Mangal, A. Mousa, G. Nedjati-Gilani, P. Nouvellet, D. Olivera, K. V. Parag, M. Pickles, H. A. Thompson, R. Verity, C. Walters, H. Wang, Y. Wang, O. J. Watson, L. Whittles, X. Xi, L. Okell, I. Dorigatti, P. Walker, A. Ghani, S. Riley, N. M. Ferguson, C. A. Donnelly, S. Flaxman, S. Bhatt, *Report 21: Estimating COVID-19 Cases and Reproduction Number in Brazil* (2020); https://doi.org/10.25561/78872.
23. Y.-Z. Zhang, E. C. Holmes, Novel 2019 coronavirus genome, Virological (2020); https://virological.org/t/novel-2019-coronavirus-genome/319.
24. V. M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D. K. W. Chu, T. Bleicker, S. Brünink, J. Schneider, M. L. Schmidt, D. G. J. C. Mulders, B. L. Haagmans, B. van der Veer, S. van den Brink, L. Wijsman, G. Goderski, J.-L. Romette, J. Ellis, M. Zambon, M. Peiris, H. Goossens, C. Reusken, M. P. G. Koopmans, C. Drosten, Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill.* **25**, 2000045 (2020). doi:10.2807/1560-7917.ES.2020.25.3.2000045 Medline
25. T. Thi Nhu Thao, F. Labroussaa, N. Ebert, P. V'kovski, H. Stalder, J. Portmann, J. Kelly, S. Steiner, M. Holwerda, A. Kratzel, M. Gultom, K. Schmied, L. Laloli, L. Hüsser, M. Wider, S. Pfaender, D. Hirt, V. Cippà, S. Crespo-Pomar, S. Schröder, D. Muth, D. Niemeyer, V. M. Corman, M. A. Müller, C. Drosten, R. Dijkman, J. Jores, V. Thiel, Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* **582**, 561–565 (2020). doi:10.1038/s41586-020-2294-9 Medline
26. P. C. Resende, E. Delatorre, T. Gräf, D. Mir, F. do Couto Motta, L. Reis Appolinario, A. C. Dias da Paixão, M. Ogrzewalska, B. Caetano, M. Cordeiro dos Santos, J. de Almeida Ferreira, E. Costa Santos Junior, S. Patroca da Silva, S. Bianchini Fernandes, L. A. Vianna, L. da Costa Souza, J. F. G. Ferro, V. B. Nardy, J. Croda, W. K. Oliveira, A. Abreu, G. Bello, M. M. Siqueira, Genomic surveillance of SARS-CoV-2 reveals community transmission of a major lineage during the early pandemic phase in Brazil. bioRxiv 020.06.17.158006 [Preprint] (2020); https://doi.org/10.1101/2020.06.17.158006.
27. J. Xavier, M. Giovanetti, T. Adelino, V. Fonseca, A. V. Barbosa da Costa, A. Aparecida Ribeiro, K. Nascimento Felicio, C. Guerra Duarte, M. V. Ferreira Silva, A. Salgado, M. Teixeira Lima, R. de Jesus, A. Fabri, C. Franco Soares Zoboli, T.

Gutemberg Souza Santos, F. Iani, A. M. Bispo de Filippis, M. Agudo Mendonca Teixeira de Siqueira, A. L. de Abreu, V. de Azevedo, D. Brock Ramalho, C. F. Campelo de Albuquerque, T. de Oliveira, E. C. Holmes, J. Lourenco, L. C. Junior Alcantara, M. Aparecida Assuncao Oliveira, The ongoing COVID-19 epidemic in Minas Gerais, Brazil: insights from epidemiological data and SARS-CoV-2 whole genome sequencing. medRxiv 2020.2005.2005.20091611 [Preprint] (2020); https://doi.org/10.1101/2020.05.05.20091611.

28. V. A. Nascimento, A. L. G. Corado, F. O. Nascimento, A. K. A. Costa, D. C. G. Duarte, M. S. Jesus, S. L. B. Luz, L. M. F. Goncalves, C. F. Costa, E. Delatorre, F. G. Naveca, Genomic and phylogenetic characterization of an imported case of SARS-CoV-2 in Amazonas State, Brazil. *Memoirs of the Oswaldo Cruz Institute* 10.1590/0074-02760200310 (2020).
29. Y. Shu, J. McCauley, GISAID: Global initiative on sharing all influenza data – from vision to reality. *Euro. Surveill.* **22**, 30494 (2017) doi:10.2807/1560-7917.ES.2017.22.13.30494 Medline
30. M. Cotten, S. J. Watson, P. Kellam, A. A. Al-Rabeeah, H. Q. Makhdoom, A. Assiri, J. A. Al-Tawfiq, R. F. Alhakeem, H. Madani, F. A. AlRabiah, S. Al Hajjar, W. N. Al-nassir, A. Albarrak, H. Flemban, H. H. Balkhy, S. Alsubaie, A. L. Palser, A. Gall, R. Bashford-Rogers, A. Rambaut, A. I. Zumla, Z. A. Memish, Transmission and evolution of the Middle East respiratory syndrome coronavirus in Saudi Arabia: A descriptive genomic study. *Lancet* **382**, 1993–2002 (2013). doi:10.1016/S0140-6736(13)61887-5 Medline
31. M. Cotten, S. J. Watson, A. I. Zumla, H. Q. Makhdoom, A. L. Palser, S. H. Ong, A. A. Al Rabeeah, R. F. Alhakeem, A. Assiri, J. A. Al-Tawfiq, A. Albarrak, M. Barry, A. Shibl, F. A. Alrabiah, S. Hajjar, H. H. Balkhy, H. Flemban, A. Rambaut, P. Kellam, Z. A. Memish, Spread, circulation, and evolution of the Middle East respiratory syndrome coronavirus. *mBio* **5**, e01062-13 (2014). doi:10.1128/mBio.01062-13 Medline
32. G. Dudas, L. M. Carvalho, A. Rambaut, T. Bedford, MERS-CoV spillover at the camel-human interface. *eLife* **7**, e31257 (2018). doi:10.7554/eLife.31257 Medline
33. Z. Zhao, H. Li, X. Wu, Y. Zhong, K. Zhang, Y.-P. Zhang, E. Boerwinkle, Y.-X. Fu, Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC Evol. Biol.* **4**, 21 (2004). doi:10.1186/1471-2148-4-21 Medline
34. S. Duchene, L. Featherstone, M. Haritopoulou-Sinanidou, A. Rambaut, P. Lemey, G. Baele, Temporal signal and the phylodynamic threshold of SARS-CoV-2. bioRxiv 2020.05.04.077735 [Preprint] (2020); https://doi.org/10.1101/2020.05.04.077735.
35. J. Lu, L. du Plessis, Z. Liu, V. Hill, M. Kang, H. Lin, J. Sun, S. François, M. U. G. Kraemer, N. R. Faria, J. T. McCrone, J. Peng, Q. Xiong, R. Yuan, L. Zeng, P. Zhou, C. Liang, L. Yi, J. Liu, J. Xiao, J. Hu, T. Liu, W. Ma, W. Li, J. Su, H. Zheng, B. Peng, S. Fang, W. Su, K. Li, R. Sun, R. Bai, X. Tang, M. Liang, J. Quick, T. Song, A. Rambaut, N. Loman, J. Raghwani, O. G. Pybus, C. Ke, Genomic Epidemiology of SARS-CoV-2 in Guangdong Province, China. *Cell* **181**, 997–1003.e9 (2020). doi:10.1016/j.cell.2020.04.023 Medline
36. D. D. S. Candido, A. Watts, L. Abade, M. U. G. Kraemer, O. G. Pybus, J. Croda, W. de Oliveira, K. Khan, E. C. Sabino, N. R. Faria, Routes for COVID-19 importation in Brazil. *J. Travel Med.* **27**, taaa042 (2020). doi:10.1093/jtm/taaa042 Medline
37. S. Dellicour, K. Durkin, S. L. Hong, B. Vanmechelen, J. Martí-Carreras, M. S. Gill, C. Meex, S. Bontems, E. André, M. Gilbert, C. Walker, N. De Maio, N. R. Faria, J. Hadfield, M.-P. Hayette, V. Bours, T. Wawina-Bokalanga, M. Artesi, G. Baele, P. Maes, A phylodynamic workflow to rapidly gain insights into the dispersal history and dynamics of SARS-CoV-2 lineages. bioRxiv 2020.05.05.078758 [Preprint] (2020); https://doi.org/10.1101/2020.05.05.078758.
38. World Health Organization, Coronavirus disease 2019 (COVID-19): Situation report –72 (WHO, 2020); https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200401-sitrep-72-covid-19.pdf?sfvrsn = 3dd8971b_2 .
39. Centre for Genomic Pathogen Surveillance, Imperial College London, Report of 427 novel genomes from Brazil and the associated metadata, Microreact (2020); https://microreact.org/project/rKjKLMrjdPVHKR1erUzKyi.
40. Data and code for: D. S. Candido, I. M. Claro, J. G. de Jesus, W. M. Souza, F. R. R. Moreira, S. Dellicour, T. A. Mellan, L. du Plessis, R. H. M. Pereira, F. C. S. Sales, E. R. Manuli, J. Thézé, L. Almeida, M. T. Menezes, C. M. Voloch, M. J. Fumagalli, T. M. Coletti, C. A. M. da Silva, M. S. Ramundo, M. R. Amorim, H. Hoeltgebaum, S. Mishra, M. S. Gill, L. M. Carvalho, L. F. Buss, C. A. Prete Jr., J. Ashworth, H. I. Nakaya, P. S. Peixoto, O. J. Brady, S. M. Nicholls, A. Tanuri, Á. D. Rossi, C. K. V. Braga, A. L. Gerber, A. P. de C. Guimarães, N. Gaburo Jr., C. Salete Alencar, A. C. S. Ferreira, C. X. Lima, J. E. Levi, C. Granato, G. M. Ferreira, R. S. Francisco Jr., F. Granja, M. T. Garcia, M. L. Moretti, M. W. Perroud Jr., T. M. P. P. Castiñeiras, C. S. Lazari, S. C. Hill, A. A. de Souza Santos, C. L. Simeoni, J. Forato, A. C. Sposito, A. Z. Schreiber, M. N. N. Santos, C. Zolini de Sá, R. P. Souza, L. C. Resende-Moreira, M. M. Teixeira, J. Hubner, P. A. F. Leme, R. G. Moreira, M. L. Nogueira, CADDE-Genomic-Network, N. M. Ferguson, S. F. Costa, J. L. Proenca-Modena, A. T. R. Vasconcelos, S. Bhatt, P. Lemey, C.-H. Wu, A. Rambaut, N. J. Loman, R. S. Aguiar, O. G. Pybus, E. C. Sabino, N. Rodrigues Faria, Evolution and epidemic spread of SARS-CoV-2 in Brazil, Dryad (2020); https://doi.org/10.5061/dryad.rxwdbrv5z.
41. A. Aktay, S. Bavadekar, G. Cossoul, J. Davis, D. Desfontaines, A. Fabrikant, E. Gabrilovich, K. Gadepalli, B. Gipson, M. Guevara, C. Kamath, M. Kansal, A. Lange, C. Mandayam, A. Oplinger, C. Pluntke, T. Roessler, A. Schlosberg, T. Shekel, S. Vispute, M. Vu, G. Wellenius, B. Williams, R. J. Wilson, Google COVID-19 community mobility reports: Anonymization process description (version 1.0). arXiv:2004.04145 [cs.CR] (8 April 2020).
42. InLoco, Mapa Brasileiro da COVID-19 (2020); https://mapabrasileirodacovid.inloco.com.br/pt/.
43. P. S. Peixoto, D. Marcondes, C. Peixoto, S. M. Oliva, Modeling future spread of infections via mobile geolocation data and population dynamics. An application to COVID-19 in Brazil. *PLoS One* **15**, e0235732 (2020). doi:10.1371/journal.pone.0235732 Medline
44. Y. Liu, A. A. Gayle, A. Wilder-Smith, J. Rocklöv, The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel Med.* **27**, taaa021 (2020). doi:10.1093/jtm/taaa021 Medline
45. J. J. Waggoner, V. Stittleburg, R. Pond, Y. Saklawi, M. K. Sahoo, A. Babiker, L. Hussaini, C. S. Kraft, B. A. Pinsky, E. J. Anderson, N. Rouphael, Triplex real-time RT-PCR for severe acute respiratory syndrome coronavirus 2. *Emerg. Infect. Dis.* **26**, 1633–1635 (2020). doi:10.3201/eid2607.201285 Medline
46. J. Quick, N. D. Grubaugh, S. T. Pullan, I. M. Claro, A. D. Smith, K. Gangavarapu, G. Oliveira, R. Robles-Sikisaka, T. F. Rogers, N. A. Beutler, D. R. Burton, L. L. Lewis-Ximenez, J. G. de Jesus, M. Giovanetti, S. C. Hill, A. Black, T. Bedford, M. W. Carroll, M. Nunes, L. C. Alcantara Jr, E. C. Sabino, S. A. Baylis, N. R. Faria, M. Loose, J. T. Simpson, O. G. Pybus, K. G. Andersen, N. J. Loman, Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **12**, 1261–1276 (2017). doi:10.1038/nprot.2017.066 Medline
47. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009). doi:10.1093/bioinformatics/btp352 Medline
48. I. Milne, M. Bayer, L. Cardle, P. Shaw, G. Stephen, F. Wright, D. Marshall, Tablet: Next generation sequence assembly visualization. *Bioinformatics* **26**, 401–402 (2010). doi:10.1093/bioinformatics/btp666 Medline
49. K. Katoh, D. M. Standley, MAFFT: Iterative refinement and additional methods. *Methods Mol. Biol.* **1079**, 131–146 (2014). doi:10.1007/978-1-62703-646-7_8 Medline
50. M. Hasegawa, H. Kishino, T. Yano, Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985). doi:10.1007/BF02101694 Medline
51. Z. Yang, Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J. Mol. Evol.* **39**, 306–314 (1994). doi:10.1007/BF00160154 Medline
52. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, R. Lanfear, IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020). doi:10.1093/molbev/msaa015 Medline
53. A. Rambaut, T. T. Lam, L. Max Carvalho, O. G. Pybus, Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016). doi:10.1093/ve/vew007 Medline
54. J. Singer, R. Gifford, M. Cotten, D. Robertson, CoV-GLUE: A web application for tracking SARS-CoV-2 genomic variation (2020); https://doi.org/10.20944/preprints202006.0225.v1.
55. J. G. Jesus, C. Sacchi, D. D. S. Candido, I. M. Claro, F. C. S. Sales, E. R. Manuli, D. B.

B. D. Silva, T. M. Paiva, M. A. B. Pinho, K. C. O. Santos, S. C. Hill, R. S. Aguiar, F. Romero, F. C. P. D. Santos, C. R. Gonçalves, M. D. C. Timenetsky, J. Quick, J. H. R. Croda, W. Oliveira, A. Rambaut, O. G. Pybus, N. J. Loman, E. C. Sabino, N. R. Faria, Importation and early local transmission of COVID-19 in Brazil, 2020. *Rev. Inst. Med. Trop. São Paulo* **62**, e30 (2020). doi:10.1590/s1678-9946202062030 Medline

56. Centers for Disease Control and Prevention, *Research Use Only 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Primers and Probes* (2020); https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel-primer-probes.html.
57. J. Northill, I. Mackay, Wuhan coronavirus (2019-nCoV) real-time RT-PCR N gene 2020 (Wuhan-N) V.1 (2020); https://www.protocols.io/view/wuhan-coronavirus-2019-ncov-real-time-rt-pcr-n-gen-ba86ihze?version_warning = no.
58. N. Nao, Shirato, K., Katano, H., Matsuyama, S., Takeda, M., Detection of second case of 2019-nCoV infection in Japan (corrected version) (2020); https://www.niid.go.jp/niid/images/vir3/nCoV/method-niid-20200123-2_erratum.pdf.
59. Thailand Ministry of Public Health, Diagnostic detection of novel coronavirus 2019 by real time RT-PCR (2020); https://www.who.int/docs/default-source/coronaviruse/conventional-rt-pcr-followed-by-sequencing-for-detection-of-ncov-rirl-nat-inst-health-t.pdf?sfvrsn = 42271c6d_4.
60. Chinese National Institute for Viral Disease Control and Prevention, Specific primers and probes for detection 2019 novel coronavirus (2020); http://ivdc.chinacdc.cn/kyjz/202001/t20200121_211337.html.
61. HKU Med, LKS Faculty of Medicine, School of Public Health, Detection of, 2019 novel coronavirus (2019-nCoV) in suspected human cases by RT-PCR (2020); https://www.who.int/docs/default-source/coronaviruse/peiris-protocol-16-1-20.pdf?sfvrsn = af1aac73_4.
62. T. C. Bruen, H. Philippe, D. Bryant, A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665–2681 (2006). doi:10.1534/genetics.105.048975 Medline
63. D. H. Huson, D. Bryant, Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006). doi:10.1093/molbev/msj030 Medline
64. D. P. Martin, B. Murrell, M. Golden, A. Khoosal, B. Muhire, RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, vev003 (2015). doi:10.1093/ve/vev003 Medline
65. M. A. Suchard, P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, A. Rambaut, Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, vey016 (2018). doi:10.1093/ve/vey016 Medline
66. D. L. Ayres, A. Darling, D. J. Zwickl, P. Beerli, M. T. Holder, P. O. Lewis, J. P. Huelsenbeck, F. Ronquist, D. L. Swofford, M. P. Cummings, A. Rambaut, M. A. Suchard, BEAGLE: An application programming interface and high-performance computing library for statistical phylogenetics. *Syst. Biol.* **61**, 170–173 (2012). doi:10.1093/sysbio/syr100 Medline
67. M. S. Gill, P. Lemey, N. R. Faria, A. Rambaut, B. Shapiro, M. A. Suchard, Improving Bayesian population dynamics inference: A coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30**, 713–724 (2013). doi:10.1093/molbev/mss265 Medline
68. M. A. R. Ferreira, M. A. Suchard, Bayesian analysis of elapsed times in continuous-time Markov chains. *Can. J. Stat.* **36**, 355–368 (2008). doi:10.1002/cjs.5550360302
69. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018). doi:10.1093/sysbio/syy032 Medline
70. P. Lemey, A. Rambaut, A. J. Drummond, M. A. Suchard, Bayesian phylogeography finds its roots. *PLOS Comput. Biol.* **5**, e1000520 (2009). doi:10.1371/journal.pcbi.1000520 Medline
71. N. R. Faria, A. Rambaut, M. A. Suchard, G. Baele, T. Bedford, M. J. Ward, A. J. Tatem, J. D. Sousa, N. Arinaminpathy, J. Pépin, D. Posada, M. Peeters, O. G. Pybus, P. Lemey, HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science* **346**, 56–61 (2014). doi:10.1126/science.1256739 Medline
72. J. D. O'Brien, V. N. Minin, M. A. Suchard, Learning to count: Robust estimates for labeled distances between molecular sequences. *Mol. Biol. Evol.* **26**, 801–814 (2009). doi:10.1093/molbev/msp003 Medline
73. V. N. Minin, M. A. Suchard, Fast, accurate and simulation-free stochastic mapping. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363**, 3985–3995 (2008). doi:10.1098/rstb.2008.0176 Medline
74. V. N. Minin, M. A. Suchard, Counting labeled transitions in continuous-time Markov models of evolution. *J. Math. Biol.* **56**, 391–412 (2008). doi:10.1007/s00285-007-0120-8 Medline
75. P. Lemey, A. Rambaut, J. J. Welch, M. A. Suchard, Phylogeography takes a relaxed random walk in continuous space and time. *Mol. Biol. Evol.* **27**, 1877–1885 (2010). doi:10.1093/molbev/msq067 Medline
76. S. Dellicour, G. Baele, G. Dudas, N. R. Faria, O. G. Pybus, M. A. Suchard, A. Rambaut, P. Lemey, Phylodynamic assessment of intervention strategies for the West African Ebola virus outbreak. *Nat. Commun.* **9**, 2222 (2018). doi:10.1038/s41467-018-03763-2 Medline
77. S. Dellicour, R. Rose, N. R. Faria, P. Lemey, O. G. Pybus, SERAPHIM: Studying environmental rasters and phylogenetically informed movements. *Bioinformatics* **32**, 3204–3206 (2016). doi:10.1093/bioinformatics/btw384 Medline

ACKNOWLEDGMENTS

A full list acknowledging those involved in the diagnostic and generation of new sequences as part of the CADDE-Genomic-Network can be found in the Supplementary Material. We thank GISAID database for supporting rapid and transparent sharing of genomic data during the COVID-19 pandemic. A full list acknowledging the authors submitting data used in this study can be found in data S2. We thank Paola Resende (FIOCRUZ), Talita Adelino (FUNED), Claudio Sacchi (IAL), Valdinete Nascimento (FIOCRUZ Amazonia) and their colleagues for submitting Brazilian data to GISAID. We thank Adriano Pinter (SUCEN), Nelson Gouveia (USP) and Izabel Marcílio de Souza (HCFM-USP) for fruitful discussions. We thank Lucy Matkin and Josh Quick for logistic support. We thank the UNICAMP Task Force against Covid-19 for support in generating genomes from Campinas. The analysis of openly available epidemiological data from https://covid.saude.gov.br/ has benefited from the COVID-19 surveillance efforts by the Secretaria de Vigilância em Saúde, Ministry of Health, Brazil. **Funding**: This project was supported by the Medical Research Council-São Paulo Research Foundation (FAPESP) CADDE partnership award (MR/S0195/1 and FAPESP 18/14389-0) (http://caddecentre.org/). FAPESP further supports IMC (2018/17176-8 and 2019/12000-1), JGJ (2018/17176-8 and 2019/12000-1, 18/14389-0), FCSS (2018/25468-9), WMS (2017/13981-0, 2019/24251-9), MF (2018/09383-3), TMC (2019/07544-2), CAMS (2019/21301-5), HIN (2018/14933-2), PSP (16/18445-7), MLN (20/04836-0), JLM (2020/04558-0 and 2016/00194-8). NRF is supported by a Wellcome Trust and Royal Society Sir Henry Dale Fellowship (204311/Z/16/Z). DSC is supported by the Clarendon Fund and by the Department of Zoology, University of Oxford. SD is supported by the *Fonds National de la Recherche Scientifique* (FNRS, Belgium). JT and PL supported by European Union's Horizon 2020 project MOOD (874850). This project was supported by CNPq (MTM and MLN; ATRV: 303170/2017-4; RSA: 312688/2017-2 and 439119/2018-9; RPS: 310627/2018-4; and WMS: 408338/2018-0), FAPERJ (ATRV: E-26/202.826/2018 and RSA: 202.922/2018). MSR is supported by FMUSP. CAP, GMF, JH, MRA are supported by CAPES. OJB is supported by Sir Henry Wellcome Fellowship funded by the Wellcome Trust (206471/Z/17/Z). RPS is supported by FAPEMIG (APQ-00475-20). MMT is supported by Instituto Nacional de Ciância e Tecnologia em Dengue (INCT Dengue 465425/2014-3). ATRV is supported by FINEP (01.16.0078.00). PL and NJL are supported by Wellcome Trust ARTIC network (Collaborators Award 206298/Z/17/Z). PL and AR are supported by the European Research Council (grant agreement no. 725422 -ReservoirDOCS). OGP, NRF and LdP are supported by the Oxford Martin School. This work received funding from the UK Medical Research Council, under a concordat with the UK Department for International Development. We additionally acknowledge support from Community Jameel and the NIHR Health Protection Research Unit in Modelling Methodology. **Author Contributions:** Conception: DSC, IMC, JGJ, ECS, NRF; Acquisition: DSC, IMC, JGJ, WMS, FRRM, RHMP, FCSS, ERM, MTM, CMV, MJF, TMC, CAMS, MSR, MRA, JA, HN, PSP, AT, ADR, CKVB, ALG, APG, NG, CSA, ACSF, CXL, JEL, CG, GMF, RSF, FG, MTG, MLM, MWP, TMPPC, CSL, AASS, CLS, JF,

ACS, AZS, MNNS, CZS, RPS, LCRM, MMT, JH, PAFL, RGM, MLN, SFC, JLPM, ATRV, RSA, ECS, NRF; Analysis: DSC, IMC, JGJ, WMS, FRRM, SD, TAM, LP, RHMP, JT, LA, CMV, HH, SM, MSG, LMC, LFB, CAP, OJB, SMN, SCH, JLPM, ATRV, SB, OGP, PL, CHW, RSA, NRF; Interpretation: DSC, IMC, JGJ, WMS, FRRM, SD, TAM, LP, RHMP, SCH, AASS, NMF, ATRV, SB, PL, CHW, AR, RSA, OGP, ECS, NRF; Drafting: DSC, IMC, JGJ, WMS, FRRM, SD, TAM, RSA, OGP, ECS, NRF; Revising: All authors have read and approved the final version of the manuscript.
Funding: WMS, MLN, NMF, JLPM, ATRV, NJL, RSA, OGP, ECS and NRF.
Competing interests: Authors declare no competing interests. **Data and materials availability:** The 427 SARS-CoV newly generated genomes from this study can be found on GISAID under the accession IDs: EPI_ISL_470568-470655 and EPI_ISL_476152-476490. An interactive visualization of the temporal, geographic and mutational patterns in our data can be found at https://microreact.org/project/rKjKLMrjdPVHKR1erUzKyi (*39*). Reads have been deposited to accession numbers PRJEB39487 (IMT-USP and UNICAMP) and PRJNA640656 (UFRJ-LNCC). All data, code, and materials used in the analysis are available on DRYAD (*40*). The IRB protocol number is CAAE 30127020.0.0000.0068 as currently described in Material and Methods.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/cgi/content/full/science.abd2161/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 to S3
List of Members of the CADDE Genomic Network
References (*41*–*77*)
Data S1 and S2
MDAR Reproducibility Checklist

10 June 2020; accepted 16 July 2020
Published online 23 July 2020
10.1126/science.abd2161

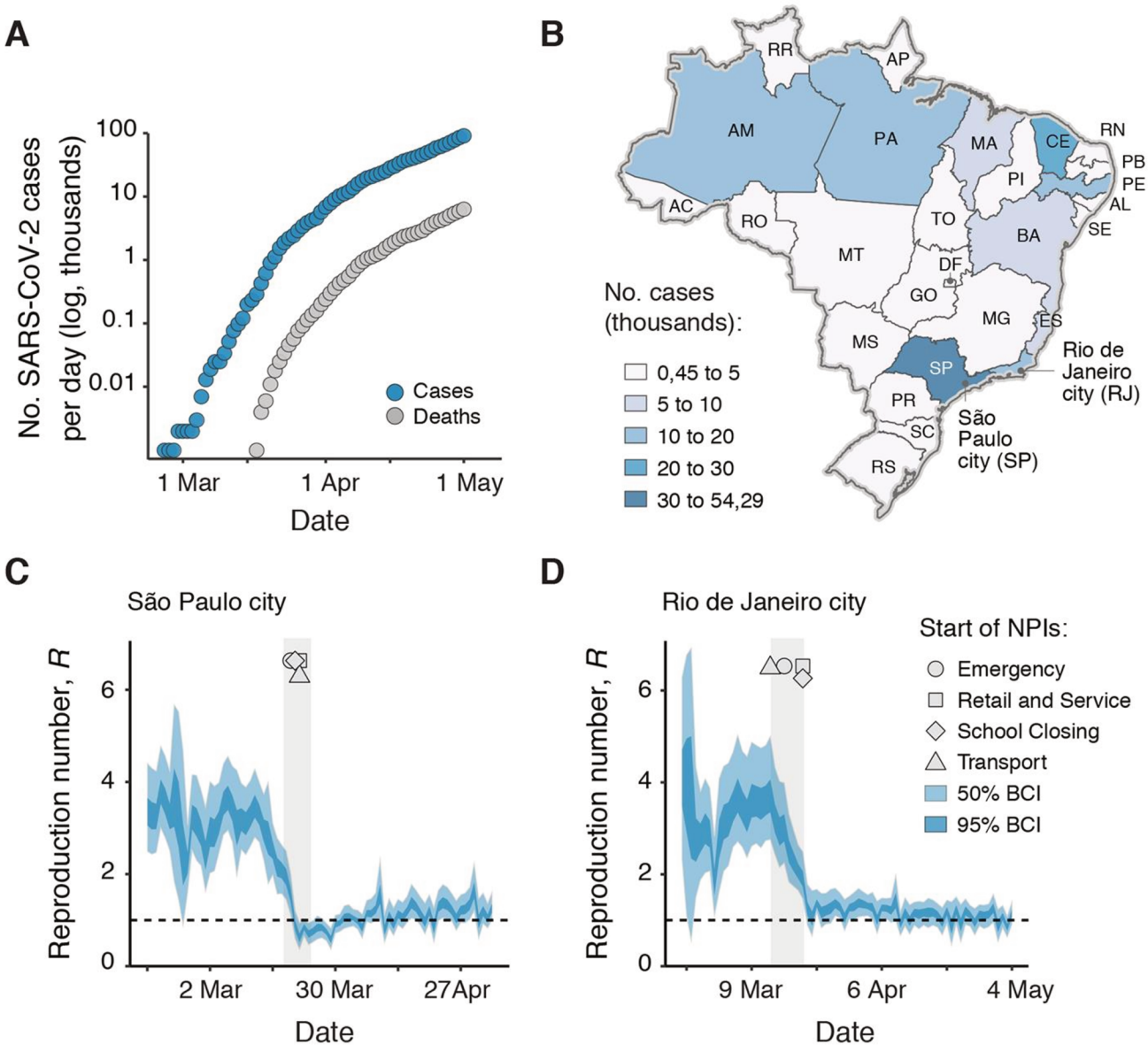


Fig. 1. SARS-CoV-2 epidemiology and epidemic spread in Brazil. (**A**) Cumulative number of SARS-CoV-2 reported cases (blue) and deaths (grey) in Brazil. (**B**) States are colored according to the number of cumulative confirmed cases by 30 April 2020. (**C** and **D**) Reproduction number (*R*) over time for the cities of São Paulo (C) and Rio de Janeiro (D). *R* were estimated using a Bayesian approach incorporating daily number of deaths and four variables related to mobility data (a social isolation index from Brazilian geolocation company *InLoco*, and Google mobility indices for time spent in transit stations, parks, and the average between groceries and pharmacies, retail and recreational, and workspaces). Dashed horizontal line indicates $R = 1$. Grey area and geometric symbols show the times at which NPIs interventions were implemented. Bayesian credible intervals (BCIs, 50 and 95%) are shown as shaded areas. The 2-letter ISO 3166-1 codes for the 27 federal units in Brazil are provided in Supplementary Information.

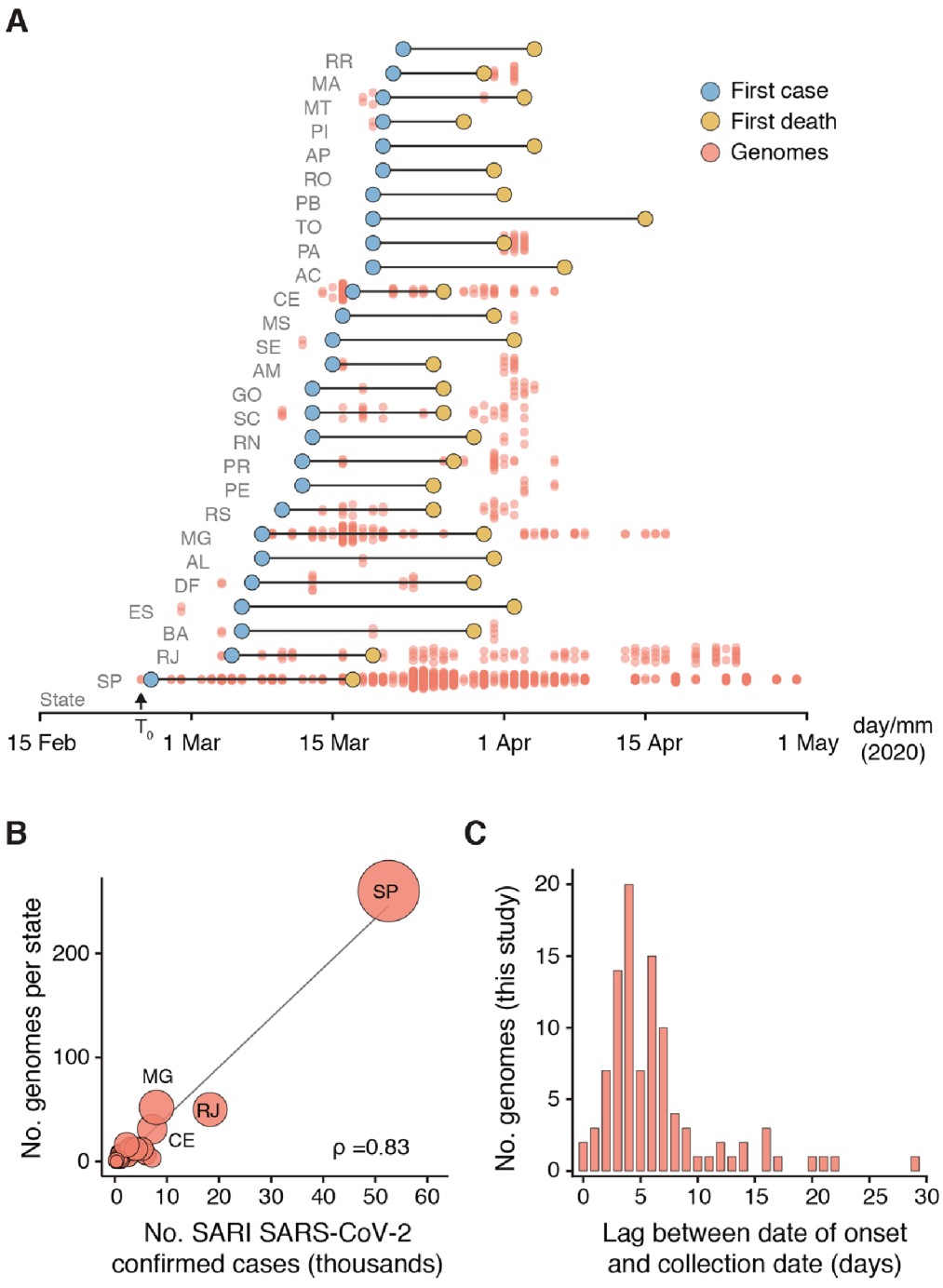


Fig. 2. Spatially representative genomic sampling. (**A**) Dumbbell plot showing the time intervals between date of collection of sampled genomes, notification of first cases and first deaths in each state. Red lines indicate the lag between the date of collection of first genome sequence and first reported case. The key for the 2-letter ISO 3166-1 codes for Brazilian federal units (or states) are provided in Supplementary Information. (**B**) Spearman's rank (ρ) correlation between the number of SARI SARS-CoV-2 confirmed and SARI cases with unknown aetiology against number of sequences for each of the 21 Brazilian states included in this study (see also fig. S4). Circle sizes are proportional to the number of sequences for each federal unit. (**C**) Interval between the date of symptom onset and date of sample collection for the sequences generated in this study.

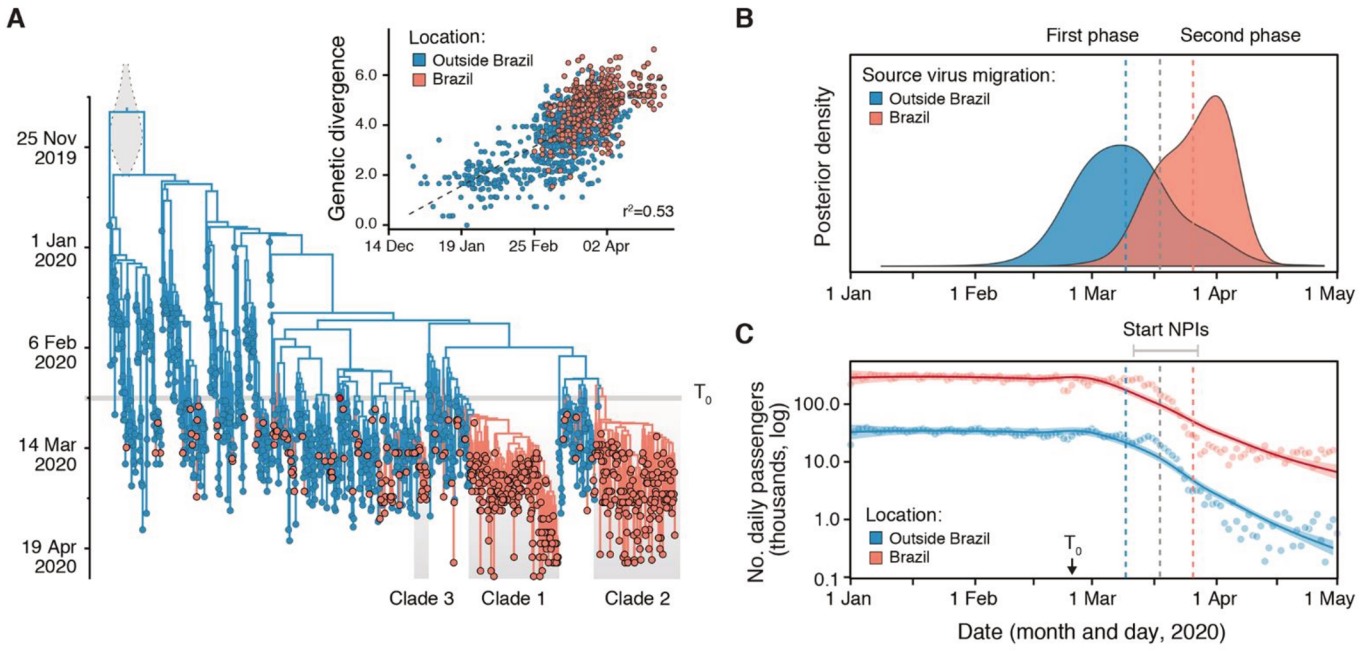


Fig. 3. Evolution and spread of SARS-CoV-2 in Brazil. (**A**) Time-resolved maximum clade credibility phylogeny of 1182 SARS-CoV-2 sequences, 490 from Brazil (red) and 692 from outside Brazil (blue). The largest Brazilian clades are highlighted by grey boxes (*Clade 1*, *Clade 2* and *Clade 3*). The panel A inset shows a root-to-tip regression of genetic divergence against dates of sample collection. (**B**) Dynamics of SARS-CoV-2 import events in Brazil. Dates of international and national (between federal states) migration events were estimated from virus genomes using a phylogeographic approach. The first phase was dominated by virus migrations from outside Brazil while the second phase is marked by virus spread within Brazil. Dashed vertical lines correspond to the mean posterior estimate for migration events from outside Brazil (blue) and within Brazil (red). (**C**) Locally estimated scatterplot smoothing of the daily number of international (blue) and national (red) air passengers in Brazil in 2020. T_0 = date of first reported case in Brazil (25 February 2020).

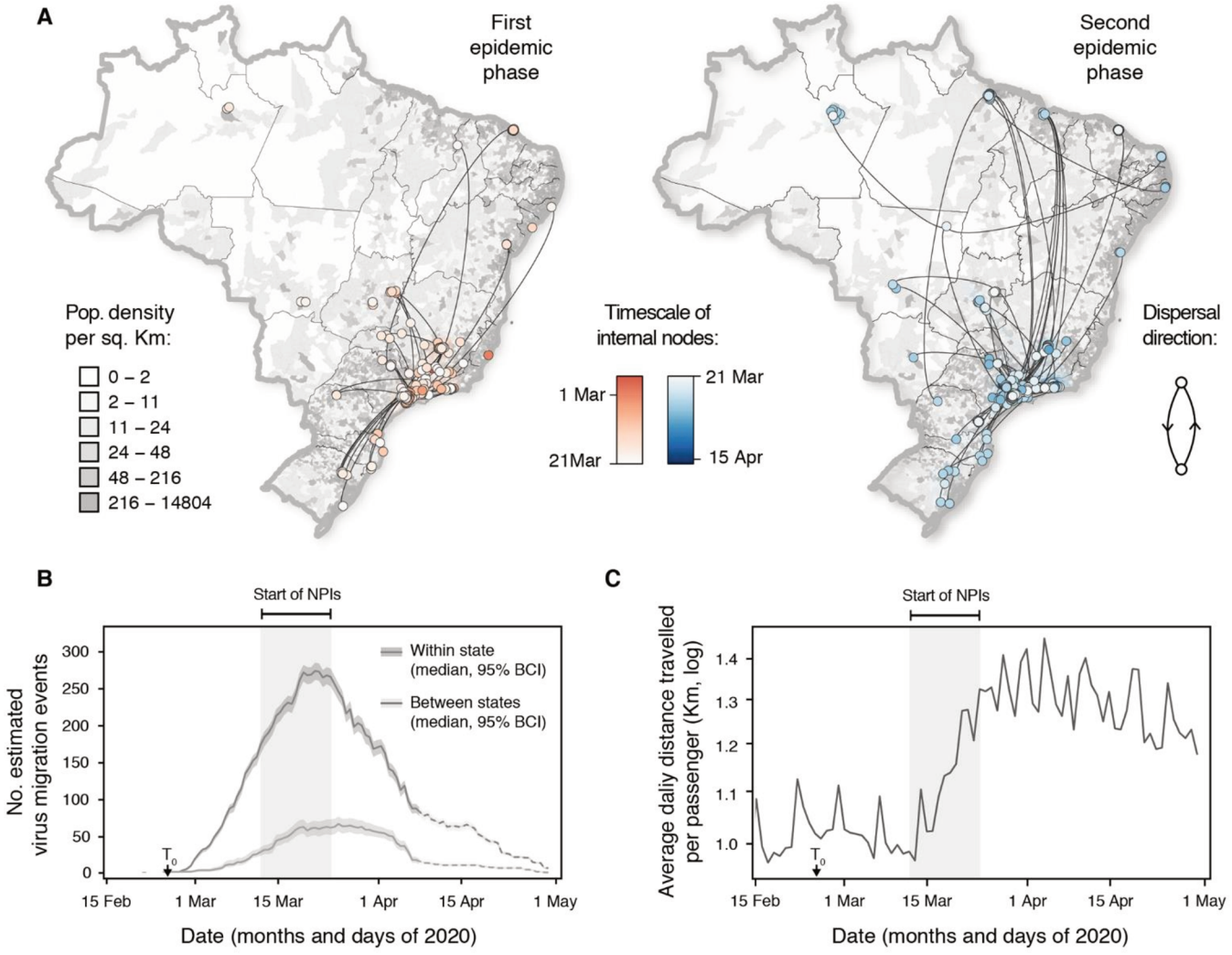


Fig. 4. Spread of SARS-CoV-2 in Brazil. (**A**) Spatiotemporal reconstruction of the spread of Brazilian SARS-CoV-2 clusters containing >2 sequences during the first (left) and the second epidemic phase (right) epidemic phase (Fig. 3B). Circles represent nodes of the MCC phylogeny and are colored according to their inferred time of occurrence. Shaded areas represent the 80% highest posterior density (HPD) interval and depict the uncertainty of the phylogeographic estimates for each node. Solid curved lines denote the links between nodes and the directionality of movement. Sequences belonging to clusters with <3 sequences were also plotted on the map with no lines connecting them. Background population density for each municipality was obtained from the Brazilian Institute of Geography (https://www.ibge.gov.br/). See fig. S14 for details of virus spread in the Southeast region. (**B**) Estimated number of within state (or within a given federal unit) and between-state (or between federal units) virus migrations over time. Dashed lines indicate estimates obtained during period of limited sampling (fig. S2). (**C**) Average distance in kilometres travelled by an air passenger per day in Brazil. Number of daily air passengers is shown in Fig. 3B. Light grey boxes indicate starting dates of NPIs across Brazil.

Rapid remote characterization of RAS associated to inhibitors of HCV NS5B from dried blood spot of samples from São Paulo state, Brazil

1.1 Author names

Victória Riquena Grosche[1,2], Kazeem Adeboyejo[3,4], Diego Pandeló José[5], Giulia Magalhães Ferreira [2], Jacqueline Farinha Shimizu[1,2], Márcia Maria Costa Nunes Soares[6], Barnabas J. King[3,4], Alexander W. Tarr[3,4,7], Jonathan K. Ball[3,4,7], C. Patrick McClure [3,4,7], Ana Carolina Gomes Jardim[1,2]

1.2 Affiliation

[1] Institute of Biomedical Science, Federal University of Uberlandia, Uberlandia, MG, Brazil

[2] Institute of Bioscience, Language and Exact Sciences, São Paulo State University, São José do Rio Preto, SP, Brazil

[3] NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, UK

[4] School of Life Sciences, University of Nottingham, UK

[5] Federal University of Triângulo Mineiro, Campus Universitário de Iturama, Iturama, MG, Brazil

[6] Instituto Adolfo Lutz, Brazilian Ministry of Health, São Paulo, Brazil

[7] MRC/EPSRC Nottingham Molecular Pathology Node, University of Nottingham, UK

1.3 Corresponding author

Professor Ana Carolina Gomes Jardim, Institute of Biomedical Science, ICBIM, Federal University of Uberlandia, Avenida Amazonas, 4C- Room 216, Umuarama, Uberlândia, Minas Gerais, Brazil, CEP: 38405-302

Tel: +55 (34) 3225-8682

E-mail: jardim@ufu.br

1.4 Keyword

HCV; Dried Blood Spot; NS5B; Direct-Action Antivirals; Resistance-Associated Substitutions; Genotyping.

2. Abstract

Hepacivirus C (HCV) is responsible for more than 180 million infections worldwide. Therapy is based on the administration of interferon (INF), ribavirin (RBV) or more recently Direct-Acting Antivirals (DAAs). However, amino acid substitutions associated with resistance (RAS) have been extensively described and can contribute to treatment failure but requires infrastructure to determine, not always locally available. Dried blood spots (DBS) sampling is an alternative specimen collection method, recommended by World Health Organization (WHO) testing guidelines, which embeds drops of blood onto filter paper, that can then be transported by posting to a centralized laboratory, where testing can take place. Here, we assessed feasibility of genotypic analysis of HCV from DBS in a cohort of 80 patients from Sao Paulo state Brazil, by international ambient shipment to a partnering institution in the UK. HCV RNA was detected on DBS specimens in 66 out 80 subjects with chronic hepatitis C, representing 83 % success rate. We performed genetic and phylogenetic analysis of the palm domain of NS5B region, compared the response to the treatment among patients and characterized the RAS identified in these sequences. The assays allowed determination of HCV genotype, 1a, 1b, 2a, 2c and 3a, and identification of RAS C316N/Y, Q309R and V321I in HCV 1b infected samples. One patient that presented RAS (V321I) was classified as non-responder. Concerning therapy outcome, 75 % of the patients who used INF+RBV as a previous protocol of treatment did not respond to treatment with DAAs, and 25 % were end-of-treatment responders. Based on this data, therapy with INF plus RBV may contribute for non-response to a second therapeutic protocol with DAAs and combination of RAS C316N and Q309R does not necessarily imply in resistance to treatment in this cohort of patients, however, the study of viral variants circulating in the region is important for a better understanding of HCV variability and resistance to the therapy. In this context, the feasibility of carrying out genotyping and RAS phenotyping analysis by using DBS card for the potential of informing future treatment interventions could be relevant to overcome the limitations of processing samples in several location worldwide.

3. Introduction

Hepatitis C is an infectious disease caused by the hepacivirus C (HCV), a globally prevalent pathogen which has infected nearly 180 million people worldwilde (1,2). Currently, over 50% of HCV infected people are unaware of their clinical condition and infection can result in either, acute or chronic hepatitis, ranging in severity from a mild to severe lifelong illness (3–5). Persistent HCV infection may initiate a chronic inflammatory disease process that can progress to liver fibrosis, cirrhosis, hepatocellular carcinoma, and death (3,6).

HCV is an enveloped, icosahedral, positive sense single-stranded RNA virus belonging to the Flaviviridae family, genus *Hepacivirus* (7,8). The viral genome is composed of approximately 9,600 nucleotides and represents an uninterrupted open reading frame (ORF) encoding a polyprotein precursor of approximately 3,000 amino acids (9,10). The HCV polyprotein is co- and post-translationally processed by a combination of cellular and viral proteases into three structural proteins (Core, E1, and E2) and seven non-structural (NS) proteins (P7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B). Phylogenetic analysis of globally strain diversity resulted in the distinction of eight HCV genotypes (1 to 8) and 86 subtypes (a, b, c etc.), based on the nucleotide sequences of the Core/E1 and NS5B regions (2,3,9,11–13). NS5B is the viral RNA-dependent RNA-polymerase (RdRp) responsible for HCV RNA replication (9,14). The HCV RdRp enzyme lacks proof-reading activity and is error-prone, leading to the emergence of intrahost diversity of viral populations circulating in the blood of an individual as a mosaic of highly related sequences (9). The variability and dynamism of viral infection in the host organism are related to the progression of antiviral treatments, which can result in selective pressure for the emergence and spread of drug-resistant viral strains (9,15).The initial therapies used against HCV were interferons-based (IFN), resulting in several collateral effects. In 2015, Food and Drug Administration (FDA) approved the administration of IFN-free direct-acting antivirals (DAAs) (3,16), classified as NS3/4A protease inhibitors, NS5A viral replication complex inhibitors and NS5B polymerase inhibitors (NI - nucleoside inhibitor, or NNI - non-nucleoside inhibitor). Treatment with DAAs presents high levels of sustained virologic response (SVR) rates, typically > 90%, contributing to both the control of HCV epidemics and also improving patients' quality of life, notably during treatment with reduced side effects compared to IFN (3). However, under treatment, resistance-associated substitutions (RAS) might occur in to the genomic regions related to the viral proteins targeted by anti-HCV drugs, which may confer resistance to the DAA classes (3).

The three-dimensional structure of the viral NS5B polymerase contains 3 domains: Palm, Fingers, and Thumb (17–19). Structural analyzes of this protein and its active site in the palm domain demonstrated that, close to the catalytic triad of HCV NS5B polymerase, occurred substitutions that interfere with the effectiveness of antivirals. The most common RAS that interferes in the activity of NS5B inhibitors are L159F, D244N, S282T/R, Q309R, D310N, and A333E (20–22). Due to this, supervising the viral evolution and identification of RAS in the NS5B of HCV is a relevant approach. Especially in Brazil, since it is estimated that around two million people are infected with HCV (23,24), but there still are difficulties in detecting new ones and in monitoring the treatment (23,25).

Identification and genotyping of RAS can be a challenge in many locations. The use of dried blood spots (DBS) may facilitate the monitoring of it in remote settings, with limited access to laboratory services, reaching vulnerable populations (26–29). In that way, DBS sampling is an alternative

specimen collection method, which embeds drops of blood onto filter paper, that can then be transported by posting to a centralized laboratory, where testing can take place (30). Also, the use of this method may reduce costs associated with sample collection, preparation, storage and transportation, as it avoids the need for plasma separation within a specified time period, and cold chain storage (26). World Health Organization (WHO) testing guidelines recommend the use of DBS specimens as an option for HCV testing and analysis as well, especially in settings where there are no facilities (31). Here we assessed the suitability of dried blood spots for genetic and phylogenetic analyses the HCV NS5B palm domain variability and treatment response in a cohort of chronically infected patients from Sao Paulo state, Brazil.

4. Methods

4.1 Clinical Samples:

The studied population was composed of 80 patients chronically infected with Hepacivirus C (HCV), which 66 have been successfully amplified using dried blood spots (DBS) sampling. Samples were collected in eight specialized centres of the northwest region of the state of São Paulo, Brazil, in the cities of São José do Rio Preto, Catanduva, Votuporanga, Fernandópolis, Jales, and Santa Fé do Sul (Figure 1A), between 2010 and 2017. Patients data, including age, gender, location, and clinical situation, were obtained during sample collection. This study was approved by the Ethical Clearance Committee on Human Rights Related to Research Involving Human Subjects from Institute Adolfo Lutz, as per resolution 196/96 National Health Council.

All samples were obtained for routine diagnostic investigation of HCV in the Laboratory of Molecular Biology of HCV, HBV, and HIV Viruses, Adolfo Lutz Institute in the city of São José do Rio Preto, São Paulo, Brazil. Thirty microliters of serum were spotted onto a DBS Protein Saver^{TM} 903 Card (Whatman Cardiff) air-dried at room temperature for 1 hour before packaging for shipment to University of Nottingham, UK, or storage at -70 °C, for extraction and sequence analysis. Samples varied in virus titer from 1.1 x 10^5 to 1.1 x 10^7 infectious units (IU)/ml.

4.2 NS5B Amplification and Sequencing:

The entire 12mm punch from each DBS was collected into 300 µl PBS and incubated with agitation at room temperature for 10 minutes. The punching tool was cleaned by punching a blank card between samples. Viral RNA extraction was performed with 140 µL of PBS spot eluate using a viral RNA mini extraction kit (QIAGEN) according to manufacturer´s protocol. Extracted HCV RNA was eluted into 50 µl elution buffer (QIAGEN) and stored at -70 °C. cDNA was obtained from 20uL of extracted RNA using RNA to cDNA EcoDry Premix, Random Hexamers kit (Takara), according to manufacturer´s protocol. Reactions were thoroughly mixed then incubated at 42 °C for 1 hour and 70 °C for 10 mins then stored at -20 °C.

PCR and nested PCR reaction were performed using HotStarTaq (QIAGEN, Hilden, Germany), according to manufacturer´s instruction. Primers qHCV_NCRf (5'-GCGCAACCGGTGAGTACA-3') and qHCV_NCRr (5'-ACTCGCAAGCACCCTATCAG-3')(32) were used to amplify the 5' non coding region;

and Sn755 (5'-TATGAYACCCGCTGYTTTGACTC-3') and Asn1121 (5'-GCNGARTAYCTVGTCATAGCCTC-3')(33) for NS5b. Reactions were performed using the following conditions: 94 ºC/ 15min enzyme activation; 55 cycles of denaturation 95 °C/ 20s, annealing 55 °C/ 20s and extension 72 °C/ 20s; and final extension 72 °C/ 10min. Expected amplicon is 388 bp, confirmed by electrophoresis. PCR products were diluted 1 in 10 with water and shipped direct for Sanger sequencing (Source BioScience, Nottingham, UK) (Figure 1B).

4.3 Sequence Analysis, RAS identification and Phylogeny Reconstruction:

The NS5B sequences generated were analyzed with the BLAST algorithm (*National Center for Biotechnology Information* - NCBI)(34) and compared with HCV sequences available in the GenBank and Los Alamos databases (*Los Alamos National Lab* - LANL)(35), for confirmation of the sequenced region and genotypes. For the analysis of viral genetic diversity, the sequences were grouped in datasets according to the respective genotypes, and aligned using the program Clustal W, BioEdit (version 7.0.5.3) (36,37).

After alignment, nucleotide sequences were translated using BioEdit software, and then, NS5B inhibitors RAS screening was performed. The translated region comprised a fragment between amino acids 240 and 342 of the polymerase, where the active site of the enzyme in the palm domain was localized, between positions 287 and 371 (38). The investigated RAS were selected based on the substitutions described in the literature in the NS5B palm domain genomic region (Table 1)**.**

Maximum-likelihood phylogeny reconstruction with bootstrap evaluation was conducted in MEGA X program, using the GTR nucleotide substitution model and gamma (Γ) distribution of rate variability among sites (GTR + gamma)(39). All the substitutions in the NS5B and the RAS were identified using Geno2Pheno, a web service for rapidly identifying drug resistance in HCV samples (40).

5. Results

5.1 Clinical Characteristics of the Patients:

Studied population characteristics are shown in table 2. The patients were chronically infected with Hepacivirus C (HCV) genotypes 1a, 1b, 2a 2c, or 3a, submitted or not to treatment for 12 weeks with DAAs. Considering the distribution of patients according to gender, men were the most prevalent group, comprising 56 % of the samples. Also, the group in middle age, between 40s and 50s, was the most prevalent, comprising, respectively, 29 and 23 % of those infected, and the most prevalent infection route was the transmission of the virus through inhalable or injectable drugs and medication (23%). Regarding the degree of liver damage, infection in 4 patients progressed to cirrhosis, more than half of the group presented some degree of fibrosis and 20% of all patients were co-infected with HIV and/or HBV.

About the treatment profile of the patients, 23 % of the analyzed samples were from naive patients, which means they had not undergone previous treatment against HCV. Patients undergoing therapy with DAAs presented three distinct patterns of response to treatment, classified as: i) the sustained

virological response (SVR), meaning that blood tests of the patient continue to show no detectable RNA 12 weeks or more after treatment; ii) end-of-treatment virological responders (ETR), when there is no detectable hepatitis C virus in the blood at the completion of HCV treatment but it reappears in the blood test after the treatment finished; iii) or non-responders (NR), when HCV RNA is detectable during and after the treatment. Nineteen percent of the patients were previously treated with interferon and ribavirin, and 58 % of them submitted to a second protocol of treatment with direct-action antivirals (DAAs). Among patients who used sofosbuvir and daclatasvir, or simeprevir, as the second therapeutic approach, 75 % did not respond to treatment and 25 % were ETR, representing 0 % of SVR. Therefore, therapy with DAAs followed by a previous IFN-based protocol demonstrated reduced SVR. Patients who received DAAs as the first therapeutic approach represented 39 % of the patients, with an SVR rate of 96 % (Table 2).

5.2 HCV NS5B Sequences Show Important RAS Presence in Individuals Inside Analysed Population:

To determine the feasibility of carrying out genotyping and resistance-associated substitutions (RAS) phenotyping analysis in a different country for the potential of informing future treatment interventions, we took HCV positive samples collected in São Paulo State, Brazil, and conveyed them by dried blood spot (DBS) card to the University of Nottingham, United Kingdom, for analysis (Figure 2). The samples had previously not been genotyped, although viral load was available (Table 2). Samples varied in virus titer from 1.1 x 10^5 to 1.1 x 10^7 infectious units (IU)/ml. HCV RNA was recovered from 66 of the 80 samples, representing 83 % success rate.

The 66 recovered samples were amplified and sequenced, as described previously. The NS5B sequences (nt 8256-8644) were evaluated for quality and similarity using the BLAST and LANL algorithms (Table 2). Following a simple workflow with commercially available reagents and freely available online tools we determined that it was feasible to have genotype and phenotype data within 7 – 10 days of blood sample collection (Figure 2). The NS5B amplicon sequences were subjected to online HCV genotyping tools (LANL HCV database, Oxford HCV subtyping tool, ViRP database and Geno2Pheno) and showed 100% agreement in genotype prediction across all 66 samples (Table 2). Genotype 1a was the most prevalent in this group (53 %) followed by genotypes 1b (25 %), 3a (18 %), 2a (2 %) and 2c (2 %).

NS5B region of the study samples was evaluated for RAS by subjecting the sequences to the Geno2Pheno software tool. The amino acid sequence in this analysis overlaps 240-342 fragment. This segment was well characterized and was used for the analysis (Table 1). About HCV 1a sequences, both RAS C316Y and Q309R were observed in the sequences of patients (Table 3). Among genotype 1b sequences, patients 51, 52 and 53 presented RAS Q309R and C316N, previously described for the amplified region (Table 1); and patients 35 and 45 presented RAS V321I (Table 3). Regarding the genotypes 2a, 2c and 3a, no RAS was identified among the sequences. Other substitutions not yet described as RAS are listed on table 2.

5.3 Phylogenetic Analysis Disclose Proximity Among Sequences Harboring the Identified RAS:

The tree from the phylogenetic reconstruction with the 66 analyzed sequences can be seen in figure 3. Genotypes 1a, 1b, 2a, 2c and 3a are grouped in different monophyletic branches strongly sustained by the bootstrap values of 99%, 99%, 100%, 99%, and 100%, respectively, confirming previous genotyping analyzes (Figure 3). The HCV 1b genotype sequences 51, 52 and 53, which presented RAS C316N and Q309R, grouped in the same monophyletic branch, with 94 of bootstrap support value. Additionally, sequences 52 and 53 grouped in a clade strongly sustained by the highest bootstrap value (Figure 3). Also, the HCV 1b sequences 35 and 45 presented RAS V321I, and grouped in a clade sustained by the bootstrap value of 80%.

6. Discussion

The worldwide prevalence of hepacivirus C (HCV) infection and its relation to chronic liver disease made it the subject of constant research. The lack of a vaccine to prevent infections has resulted in a significant advance in antiviral treatment against HCV in recent years. The substitution of interferon-based therapy by direct-acting antivirals (DAAs) has significantly improved the quality of life of patients since it interferes mainly in the viral replicative cycle, generating reduced side effects (41).

The new findings on HCV viral replication allowed a better understanding of non-structural 5B (NS5B) activity, resulting in the polymerase inhibitors implementation. The DAAs of this class act mainly in the NS5B palm domain, between the amino acids at positions 287 and 371, which is the most well-conserved domain and contains the catalytic residues. Therefore, resistance associated substitutions (RAS) to these antivirals are concentrated in this HCV region (20). Until now, there are few studies in Brazil associating the effectiveness of DAA therapeutic regimens and the role of RAS in a flaw treatment outcome (42). Monitoring the identification and circulation of the variants in Brazilian population, and worldwide, is relevant to measure the possible impact in therapeutic failure with DAAs (43). One reason for the lack of data on the circulation of RAS in several countries is the difficulty in testing patients infected with HCV (23,25). The dried blood spots (DBS) sampling is, however, an reliable alternative specimen collection method that may facilitate the monitoring of ongoing RAS into vulnerable populations (26–29). In this study, we highlight that 83 % of the samples had HCV RNA recovered, and all the subsequent analyses came from them.

According to the data, patients who previously undergone therapy with interferon (INF) and ribavirin (RBV) were submitted to a second therapeutic approach, using DAAs (sofosbuvir and daclatasvir, or simeprevir). These patients did not achieve sustained virological response (SVR), suggesting that previous therapy with INF and RBV could interfere in the effectiveness of subsequent therapy with DAAs. Additionally, patients who received DAAs as the first therapeutic approach presented SVR rate of 96%. Treatments with INF and RBV present many side effects and, the RBV mechanism action is still unclear. It is known that RBV is a purine analog, and it is involved in several cellular pathways acting synergistically with INF(41). Although, within the antiviral action of RBV, an important mechanism is the inhibition of HCV replication, which interferes with both the NS5B polymerase and the NS5A multifunctional protein (44). The action of RBV as a purine analog, associated with the lack of corrective activity of NS5B, results in high rates of mutations in the viral variants of chronically infected patients undergoing this therapeutic protocol (45). It could potentially cause the failure of

the first treatment and could also result in a bottleneck effect in the viral population. In that way, previous treatment with INF and RBV could select resistant variants, which could explain why patients did not respond to the treatment DAAs as a second therapeutic approach. Jardim and coworkers suggested in a study that evaluates intrahost viral diversity in HCV chronically infected patients that the composition of quasispecies population at the beginning of the treatment, followed by an increase in some predominant quasispecies after treatment of non-responders (NR) and end-of-treatment responders (ETR) patients, represented an advantage for the virus remains in the organism (46).

Treatment resistance is associated with several factors. The resistance mechanism associated with RAS is still not well described, therefore the presence of an isolated substitution may not confer resistance to the treatment. It is known that the combination of RAS L159F and L320F or C316N, for example, directly interferes in the action of sofosbuvir, targeting the palm domain of NS5B (20,47). Furthermore, a study by Uchida and colleagues demonstrated that patients with previous treatment with RBV had a lower frequency of C316N replacement than those who did not treat with this drug (48).

The RAS C316N / Y, Q309R and V321I are also associated with DAA resistant variants. The spread of these mutations in the northwest region of São Paulo, Brazil, is worrisome, as it may be associated with an increase in the number of individuals resistant to the available treatment. Therefore, surveillance of circulating variants in the infected population should be done regularly. A recent study by Costa et al. evaluated more than 100 HCV-infected patients in Rio de Janeiro, Brazil, proving the presence of amino acid substitutions in positions 159 and 316 of the NS5B of genotype 1b. Although the majority of Brazilian patients with HCV are susceptible to therapeutic regimens with DAAs, the presence of these RAS in the variants of this region opens the possibility for the implementation of resistant viral variants (43).

In the present study, both RAS C316N and Q309R located in the NS5B palm region were observed in the sequences from patients 51, 52, and 53. These sequences were grouped in the same monophyletic branch, showing the high similarity between the viral variants of these patients. Despite these patients having presented SVR after therapy with DAAs, the identification of the RAS in these samples is an important indicator that substitutions that are circulating in the region. Mutations in the amino acid C316 have been frequent in Asian variants, moreover, they have been reported in many analyzes in Brazil (49). Additionally, a second monophyletic branch was presented in the reconstruction of the topology of the phylogenetic tree, grouping sequences 52 and 53, demonstrating that these are closer genetically than to sequence 51. Interestingly, patients 52 and 53 were examined and diagnosed in Votuporanga, unlike patient 51 who was examined in São José do Rio Preto. Besides that, the RAS V321I was present in only two sequences, 35 and 45, and both also grouped in a separated clade. One of them is from a patient diagnosed in Fernandópolis, and the other, from Jales, is the only non-responder patient that presents a target RAS in the hole studied population.

The routes of transmission of the virus within the population studied were similar to the national standard notified by the Brazilian Ministry of Health, since in its latest epidemiological bulletin on Hepatitis C it was disclosed that the highest percentage of probable source of infection was related to the use of drugs, with 13 % of cases (50). In this study, the amount related to the use of inhalable

or injectable drugs medicaments comprised 23 % of the reported cases, this being the main route of transmission of the virus. The transmission of HCV through the sharing of needles, syringes, and other material related to drug abuse leads to risk factors in the country, mainly affecting truck drivers, people who have been in prison, and homeless people (51). This demonstrates the importance of monitoring viral variants within the population, especially those who had contact with the virus through this route of transmission, and the DBS method, by its facilities, is a favorable approach to reach this population. Patients who did not present the probable source or mechanism of infection, or had it as "ignored", comprised 30 % of cases. Despite still being a high number, it is much smaller than the reality in Brazil in which the lack of information to characterize the transmission routes accounted for 54 % of the reported cases (50).

The use of DBS sampling in testing on HCV RNA, despite of been recommended by the World Health Organization (WHO) testing guidelines, especially in difficult situations (31), still not so common. There are few studies that prioritize low- and middle-income countries and people in a vulnerable situation, as homeless and with historic of drug abuse, for example. Nguyen and collaborators evaluate HCV RNA and HCV antigen testing on DBS in HIV/HCV co-infected people who inject drugs in Vietnam. The group were tested positive for HCV RNA in serum, and HCV RNA was detected on DBS specimens in 79 out 86 subjects with chronic hepatitis C. According to their data, detection of HCV RNA on DBS showed a higher analytical sensitivity compared to quantification of HCVcAg on the same method. Finally, clinical management of HCV requires accurate diagnosis, genotyping and resistance phenotyping. In some countries limited facilities are available to genotype and perform resistance associated substitution (RAS) analysis. Therefore, the feasibility of carrying out genotyping and RAS phenotyping analysis in a different country by using dried blood spot (DBS) card for the potential of informing future treatment interventions could be relevant to cover a higher number of samples analyzed and might be the only possibility for some countries.

7. Author statements

7.1 Authors and contributors

VRG performed phylogenetic analysis, drafted the manuscript and illustrations; KA carried out the experimental work, performed genetic and phylogenetic analysis; GMF drafted the manuscript; MCNS provided samples and associated clinical data; JFS prepared DSS cards; CPM, DPJ and ACGJ designed the experiments, analysed data and critically revised the manuscript; BJK, AWR and JKB analysed the data and drafted the manuscript. All the authors read and approved the final manuscript.

7.2 Conflicts of interest

The authors declare that they have no conflicts of interest.

7.3 Funding information

Funding was provided by the Royal Society – Newton Advanced Fellowship (grant reference), and CAPES (Coordination for the Improvement of Higher Education Personnel – Code 001). VGR received fellowship from the pro-rectory of undergraduate studies program of the Federal University of Uberlândia. ACGJ received productivity fellowship (311219/2019-5) from the CNPq (National Counsel of Technological and Scientific Development).

7.4 Acknowledgements

We would like to thank the support of fellowships and the samples provided by the Adolfo Lutz Institute.

8. References

1. Paraskevis D, Stylianou DC, Hezka J, Stern Z, Oikonomopoulou M, Mamais I, et al. HCV Phylogeography of the General Population and High-Risk Groups in Cyprus Identifies the Island as a Global Sink for and Source of Infection. Sci Rep [Internet]. 2019 Dec 11;9(1):10077. Available from: http://www.nature.com/articles/s41598-019-46552-7

2. Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, Stapleton JT, et al. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: Updated criteria and genotype assignment web resource. Hepatology [Internet]. 2014 Jan;59(1):318–27. Available from: http://doi.wiley.com/10.1002/hep.26744

3. Manns MP, Buti M, Gane E, Pawlotsky J-M, Razavi H, Terrault N, et al. Hepatitis C virus infection. Nat Rev Dis Prim [Internet]. 2017;3:17006. Available from: http://www.nature.com/articles/nrdp20176

4. Messina JP, Humphreys I, Flaxman A, Brown A, Cooke GS, Pybus OG, et al. Global distribution and prevalence of hepatitis C virus genotypes. Hepatology [Internet]. 2015 Jan;61(1):77–87. Available from: http://doi.wiley.com/10.1002/hep.27259

5. Mayberry J, Lee WM. The Revolution in Treatment of Hepatitis C. Med Clin North Am [Internet]. 2019 Jan;103(1):43–55. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0025712518300981

6. Stanaway JD, Flaxman AD, Naghavi M, Fitzmaurice C, Vos T, Abubakar I, et al. The global burden of viral hepatitis from 1990 to 2013: findings from the Global Burden of Disease Study 2013. Lancet [Internet]. 2016 Sep;388(10049):1081–8. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0140673616305797

7. Kaito M, Watanabe S, Tsukiyama-Kohara K, Yamaguchi K, Kobayashi Y, Konishi M, et al. Hepatitis C virus particle detected by immunoelectron microscopic study. J Gen Virol [Internet]. 1994 Jul 1;75(7):1755–60. Available from: https://www.microbiologyresearch.org/content/journal/jgv/10.1099/0022-1317-75-7-1755

8. Pol S, Lagaye S. The remarkable history of the hepatitis C virus. Genes Immun [Internet]. 2019 May 25;20(5):436–46. Available from: http://www.nature.com/articles/s41435-019-0066-z

9. Tsukiyama-Kohara K, Kohara M. Hepatitis C virus: Viral quasispecies and genotypes. Int J Mol Sci. 2018;19(1):1–8.

10. Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science [Internet]. 1989 Apr 21;244(4902):359–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2523562

11. Hedskog C, Parhy B, Chang S, Zeuzem S, Moreno C, Shafran SD, et al. Identification of 19 novel hepatitis C virus subtypes-further expanding HCV classification. Open Forum Infect Dis. 2019;6(3):1–9.

12. Borgia SM, Hedskog C, Parhy B, Hyland RH, Stamm LM, Brainard DM, et al. Identification of a novel hepatitis C virus genotype from Punjab, India: Expanding classification of hepatitis C virus into 8 genotypes. J Infect Dis. 2018;218(11):1722–9.

13. Smith, Donald B.; Bukh, Jens; Kuiken, Carla; Muerhoff, A. Scott; Rice, Charles M.; Stapleton, Jack T.; Simmonds P. International Committee on Taxonomy of Viruses (ICTV). HCV Classification. 2019.

14. Moradpour D, Penin F, Rice CM. Replication of hepatitis C virus. Nat Rev Microbiol [Internet]. 2007 Jun 8;5(6):453–63. Available from: http://www.nature.com/articles/nrmicro1645

15. Perales C, Quer J, Gregori J, Esteban JI, Domingo E. Resistance of hepatitis C virus to inhibitors: Complexity and clinical implications. Viruses. 2015;7(11):5746–66.

16. Manns MP, von Hahn T. Novel therapies for hepatitis C — one pill fits all? Nat Rev Drug Discov [Internet]. 2013 Aug 28;12(8):595–610. Available from: http://www.nature.com/articles/nrd4050

17. Eltahla A, Luciani F, White P, Lloyd A, Bull R. Inhibitors of the Hepatitis C Virus Polymerase; Mode of Action and Resistance. Viruses [Internet]. 2015 Sep 29;7(10):5206–24. Available from: http://www.mdpi.com/1999-4915/7/10/2868

18. Membreno FE, Lawitz EJ. The HCV NS5B Nucleoside and Non-Nucleoside Inhibitors. Clin Liver Dis [Internet]. 2011 Aug;15(3):611–26. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1089326111000304

19. Choi KH. Viral Polymerases. In 2012. p. 267–304. Available from: http://link.springer.com/10.1007/978-1-4614-0980-9_12

20. Donaldson EF, Harrington PR, Rear JJO, Naeger LK. Clinical Evidence and Bioinformatics Characterization of Potential Hepatitis C Virus Resistance Pathways for Sofosbuvir. 2014;56–65.

21. Castilho1 MCB, /+ ANM, Horbach2 IS, Perez3 R de M, Figueiredo3 FAF, Pinto4 P de TA, et al. Castilho et al 2011. 2011;106(December):968–75.

22. Noble CF, Malta F, Lisboa-Neto G, Gomes-Gouvêa MS, Leite AGB, de Castro VFD, et al. Natural occurrence of NS5B inhibitor resistance-associated variants in Brazilian patients infected with HCV or HCV and HIV. Arch Virol. 2017;162(1):165–9.

23. Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Boletim Epidemiológico Hepatites Virais 2020. 2020; Available from: http://www.aids.gov.br/pt-br/pub/2020/boletim-epidemiologico-hepatites-virais-2020

24. Ferreira PRA, Brandão-Mello CE, Estes C, Júnior FLG, Coelho HSM, Razavi H, et al. Disease burden of chronic hepatitis C in Brazil. Brazilian J Infect Dis [Internet]. 2015 Jul;19(4):363–8. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1413867015000951

25. Portari-Filho LH, Álvares-da-Silva MR, Gonzalez A, Ferreira AP, Villela-Nogueira CA, Mendes-Correa MC, et al. How are HCV-infected patients being identified in Brazil: a multicenter study. Brazilian J Infect Dis [Internet]. 2019 Jan;23(1):34–9. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1413867018306408

26. Lange B, Roberts T, Cohn J, Greenman J, Camp J, Ishizaki A, et al. Diagnostic accuracy of detection and quantification of HBV-DNA and HCV-RNA using dried blood spot (DBS) samples - a systematic review and meta-analysis. BMC Infect Dis. 2017;17(Suppl 1).

27. Komas NP, Vickos U, Hübschen JM, Béré A, Manirakiza A, Muller CP, et al. Cross-sectional study of hepatitis B virus infection in rural communities, Central African Republic. BMC Infect Dis [Internet]. 2013 Dec 24;13(1):286. Available from: https://bmcinfectdis.biomedcentral.com/articles/10.1186/1471-2334-13-286

28. Nobari RF, Meshkati M, Ataei B, Yazdani MR, Heidari K, Kassaian N, et al. Identification of Patients with Hepatitis C Virus Infection in Persons with Background of Intravenous Drug Use: The First Community Announcement-based Study From Iran. Int J Prev Med [Internet]. 2012 Mar;3(Suppl 1):S170-5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22826761

29. Lopes CLR, Teles S a, Espírito-Santo MP, Lampe E, Rodrigues FP, Motta-Castro ARC, et al. Prevalence, risk factors and genotypes of hepatitis C virus infection among drug users, Central-Western Brazil. Rev Saude Publica [Internet]. 2009;43 Suppl 1:43–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19669064

30. Snijdewind IJM, van Kampen JJA, Fraaij PLA, van der Ende ME, Osterhaus ADME, Gruters RA. Current and future applications of dried blood spots in viral disease management. Antiviral Res [Internet]. 2012 Mar;93(3):309–21. Available from: https://linkinghub.elsevier.com/retrieve/pii/S016635421100547X

31. WHO. Guidelines on Hepatitis B and C Testing [Internet]. Vol. 66, Who. 2017. 1–170 p. Available from: http://apps.who.int/iris/bitstream/10665/254621/1/9789241549981-eng.pdf?ua=1%0Ahttp://www.ncbi.nlm.nih.gov/pubmed/28742301%0Ahttp://www.ncbi.nlm.nih.gov/pubmed/28742301

32. Vercauteren K, Brown RJP, Mesalam AA, Doerrbecker J, Bhuju S, Geffers R, et al. Targeting a host-cell entry factor barricades antiviral-resistant HCV variants from on-therapy breakthrough in human-liver mice. Gut. 2016;65(12):2029–34.

33. Lortholary O, Jeantils V, Roulot D, Grando V, Gault E, Morice Y, et al. Phylogenetic analyses confirm the high prevalence of hepatitis C virus (HCV) type 4 in the Seine-Saint-Denis district (France) and indicate seven different HCV-4 subtypes linked to two different epidemiological patterns. J Gen Virol [Internet]. 2001 May 1;82(5):1001–12. Available from: http://jgv.microbiologyresearch.org/content/journal/jgv/10.1099/0022-1317-82-5-1001

34. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res [Internet]. 2008 May 19;36(Web

Server):W5–9. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkn201

35. Kuiken C, Thurmond J, Dimitrijevic M, Yoon H. The LANL hemorrhagic fever virus database, a new platform for analyzing biothreat viruses. Nucleic Acids Res [Internet]. 2012 Jan 1;40(D1):D587–92. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkr898

36. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics [Internet]. 2007 Nov 1;23(21):2947–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17846036

37. Phillips JL, Gnanakaran S. BioEdit: An important software for molecular biology. Proteins Struct Funct Bioinforma [Internet]. 2015;83(1):46–65. Available from: http://www.academia.edu/2350041/BioEdit_An_important_software_for_molecular_biology

38. Mosley RT, Edwards TE, Murakami E, Lam AM, Grice RL, Du J, et al. Structure of Hepatitis C Virus Polymerase in Complex with Primer-Template RNA. J Virol. 2012;86(12):6503–11.

39. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Battistuzzi FU, editor. Mol Biol Evol [Internet]. 2018 Jun 1;35(6):1547–9. Available from: https://academic.oup.com/mbe/article/35/6/1547/4990887

40. Döring M, Büch J, Friedrich G, Pironti A, Kalaghatgi P, Knops E, et al. geno2pheno[ngs-freq]: a genotypic interpretation system for identifying viral drug resistance using next-generation sequencing data. Nucleic Acids Res [Internet]. 2018 Jul 2;46(W1):W271–7. Available from: https://academic.oup.com/nar/article/46/W1/W271/4990638

41. Puchades Renau L, Berenguer M. Introduction to hepatitis C virus infection: Overview and history of hepatitis C virus therapies. Hemodial Int. 2018;22:S8–21.

42. Sette-Jr H, Cheinquer H, Wolff FH, de Araujo A, Coelho-Borges S, Soares SRP, et al. Treatment of Chronic HCV Infection with the New Direct Acting Antivirals (DAA): First Report of a Real World Experience in Southern Brazil. Ann Hepatol [Internet]. 2017 Sep;16(5):727–33. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1665268119304314

43. Costa VD, Brandão-Mello CE, Nunes EP, Dos Santos Silva PGC, de Souza Rodrigues LLLX, Lampe E, et al. Treatment of chronic HCV infection with DAAs in Rio de Janeiro/Brazil: SVR rates and baseline resistance analyses in NS5A and NS5B genes. PLoS One. 2019;14(5):e0216327.

44. Te HS, Randall G, Jensen DM. Mechanism of action of ribavirin in the treatment of chronic hepatitis C. Gastroenterol Hepatol. 2007;3(3):218–25.

45. Asahina Y, Izumi N, Enomoto N, Uchihara M, Kurosaki M, Onuki Y, et al. Mutagenic effects of ribavirin and response to interferon/ribavirin combination therapy in

chronic hepatitis C. J Hepatol [Internet]. 2005 Oct;43(4):623–9. Available from: https://linkinghub.elsevier.com/retrieve/pii/S016882780500382X

46. Jardim ACG, Yamasaki LHT, de Queiróz ATL, Bittar C, Pinho JRR, Carareto CMA, et al. Quasispecies of hepatitis C virus genotype 1 and treatment outcome with Peginterferon and Ribavirin☆. Infect Genet Evol [Internet]. 2009 Jul;9(4):689–98. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1567134808002244

47. Hang JQ, Yang Y, Harris SF, Leveque V, Whittington HJ, Rajyaguru S, et al. Slow Binding Inhibition and Mechanism of Resistance of Non-nucleoside Polymerase Inhibitors of Hepatitis C Virus. J Biol Chem [Internet]. 2009 Jun 5;284(23):15517–29. Available from: http://www.jbc.org/lookup/doi/10.1074/jbc.M808889200

48. Uchida Y, Nakamura S, Kouyama J, Naiki K, Motoya D, Sugawara K, et al. Significance of NS5B Substitutions in Genotype 1b Hepatitis C Virus Evaluated by Bioinformatics Analysis. Sci Rep [Internet]. 2018;(May):1–12. Available from: http://dx.doi.org/10.1038/s41598-018-27291-7

49. Peres-Da-Silva A, Brandão-Mello CE, Lampe E. Prevalence of sofosbuvir resistance-associated variants in Brazilian and worldwide NS5B sequences of genotype-1 HCV. Antivir Ther. 2017;22(5):447–51.

50. BRASIL. Ministério da Saúde. Hepatites virais - Boletim Epidemiológico. Ministério da Saúde. 2018;49(31):1–69.

51. Teles SA, Gir E, Martins RMB, dos Santos Carneiro MA, de Matos MA, Caetano KAA. Emergent predictors of hepatitis C infection among non-injection drug users. J Infect Public Health [Internet]. 2018 Jul;11(4):526–9. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1876034117302836

52. Svarovskaia ES, Dvory-Sobol H, Parkin N, Hebner C, Gontcharova V, Martin R, et al. Infrequent Development of Resistance in Genotype 1–6 Hepatitis C Virus–Infected Subjects Treated With Sofosbuvir in Phase 2 and 3 Clinical Trials. Clin Infect Dis [Internet]. 2014 Dec 15;59(12):1666–74. Available from: https://academic.oup.com/cid/article/59/12/1666/2895511

53. Tong X, Le Pogam S, Li L, Haines K, Piso K, Baronas V, et al. In Vivo Emergence of a Novel Mutant L159F/L320F in the NS5B Polymerase Confers Low-Level Resistance to the HCV Polymerase Inhibitors Mericitabine and Sofosbuvir. J Infect Dis [Internet]. 2014 Mar 1;209(5):668–75. Available from: https://academic.oup.com/jid/article-lookup/doi/10.1093/infdis/jit562

54. Lontok E, Harrington P, Howe A, Kieffer T, Lennerstrand J, Lenz O, et al. Hepatitis C virus drug resistance-associated substitutions: State of the art summary. Hepatology. 2015;62(5):1623–32.

55. Svarovskaia ES, Gane E, Dvory-Sobol H, Martin R, Doehle B, Hedskog C, et al. L159F and V321A Sofosbuvir-Associated Hepatitis C Virus NS5B Substitutions. J Infect Dis [Internet]. 2016 Apr 15;213(8):1240–7. Available from: https://academic.oup.com/jid/article-lookup/doi/10.1093/infdis/jiv564

56. HAMANO K, SAKAMOTO N, ENOMOTO N, IZUMI N, ASAHINA Y, KUROSAKI M, et al. Mutations in the NS5B region of the hepatitis C virus genome correlate with clinical outcomes of interferon-alpha plus ribavirin combination therapy. J Gastroenterol Hepatol [Internet]. 2005 Sep;20(9):1401–9. Available from: http://doi.wiley.com/10.1111/j.1440-1746.2005.04024.x

57. Shi ST, Herlihy KJ, Graham JP, Fuhrman SA, Doan C, Parge H, et al. In Vitro Resistance Study of AG-021541, a Novel Nonnucleoside Inhibitor of the Hepatitis C Virus RNA-Dependent RNA Polymerase. Antimicrob Agents Chemother [Internet]. 2008 Feb;52(2):675–83. Available from: https://aac.asm.org/content/52/2/675

58. McCown MF, Rajyaguru S, Kular S, Cammack N, Nájera I. GT-1a or GT-1b Subtype-Specific Resistance Profiles for Hepatitis C Virus Inhibitors Telaprevir and HCV-796. Antimicrob Agents Chemother [Internet]. 2009 May;53(5):2129–32. Available from: https://aac.asm.org/content/53/5/2129

9. Figures and tables

9.1 Figure legends:

Figure 1: (A) Location of the samples collection distributed by percentage according to the city in the northwest region of the state of São Paulo (SP). (B) Schematic representation of the HCV genome, its polyprotein and the primers used in the fragment amplification.

Figure 2: Remote genotyping workflow: duration between conveyed samples by dried blood spot (DBS) card until final results.

Figure 3: Phylogenetic tree not rooted with 66 sequences of 308 nucleotides from the NS5B region from samples from patients infected with HCV genotypes 1a, 1b, 2a, 2c and 3a. Red highlights indicate the clade of sequences 51, 52, and 53, which contains RAS C316N and Q309R, and the clade of sequences 35 and 45, with V321I. Bootstrap values obtained with 1000 replicates. The figure shows the bootstraps with values above 60.

9.2 Table legends:

Table 1: Substitutions in the NS5B polymerase palm region related to resistance to direct-acting antivirals.

RAS	GENOTYPE	REFERENCE
L 159 F	1a, 1b, 3a	Donaldson et al. (2014); Svarovskaia et al. (2014); Tong et al. (2014); Lontok et al. (2015); Noble et al. (2017) (20,22,52–54)
D 244 N	3a	Asahina et al. (2005); Castilho et al. (2011)(21,45)
S 282 T/R	1a, 1b	Donaldson et al. (2014); Tong et al. (2014); Svarovskaia et al. (2016)(20,53,55)
Q 309 R	1a, 1b, 3a	Hamano et al. (2005); Castilho et al. (2011)(21,56)
D 310 N	1b, 3a	Asahina et al. (2005); Castilho et al. (2011)(21,45)
C 316 H/N/Y	1a, 1b	Shi et al. (2008), McCown et al. (2009); Castilho et al. (2011); Donaldson et al. (2014); Lontok et al. (2015); Noble et al. (2017)(20–22,54,57,58)
L 320 F	1a, 1b, 3a	Tong et al. (2014); Svarovskaia et al. (2016)(53,55)
V 321 A/I	1b, 3a	Donaldson et al. (2014); Svarovskaia et al. (2014); Lontok et al. (2015); Noble et al. (2017); Hezode et al. (2018)(22,52,54)
A 333 E	1a	Hamano et al. (2005); Castilho et al. (2011)(21,56)

Table 2: Clinical information of the studied population. About the cities, CATAN = Catanduva; FERPO = Fernandópolis; JALES = Jales; STSUL = Santa Fé do Sul; SJRPR = São José do Rio Preto; and VOTUP = Votuporanga. Referring to gender, F means Female and M, Male. Regarding the institutions, SAE = “Specialized Care Service”; HEC = “Emílio Carlos Hospital”; AMHV = “Municipal Ambulatory of Viral Hepatitis”; CADIP = “Care Center for Infectocontagious and Parasitic Diseases”. On medication, PEGINF = Pegylated Interferon; RBV = Ribavirina; SFB = Sofosbuvir; SMV = Simeprevir; and DCV = Daclatasvir. Clinical Response, NR = Non-Responders; ETR = End-of-Treatment Virological Responders; SVR = Sustained Virological Response; and ABD = Abandonment. Notes: (1) Biopsy performed to assess liver function. The METAVIR scale is used, which is exclusive for patients with chronic hepatitis C. The letter "F" informs the degree of fibrosis or cirrhosis and the letter "A", the necroinflammatory activity. F0 - liver in perfect condition; F1 - with minimal fibrosis; F2 - the existence of moderate fibrosis; advanced fibrosis; F4 - cirrhosis or very advanced fibrosis. A0 - minimal or nonexistent inflammatory activity, progressing to A3 - high inflammatory activity, faster progression to fibrosis.

Sample	Viral Load (UI/mL)	PCR DBS	LANL / Blastn NCBI	RAS	NS5B Reference	Institution/City		Age at the time of sampling	Gender	Probable Infection Rote	Co-infection (HIV, HBV)	Beginning of Treatment	Medication	Retreatment	Second Medication	Biopsy[1]	Clinical Response
1	281959	POS	1b(96%)	I262V, S300T	EU781827	SAE	JALES	38	F	Ignored	-	Naive				F2/F3	-
2	935461	POS	3a(95%)	K304R, N307G, A335S	D17763	SAE	JALES	44	M	Sexual	HIV	Naive				-	-
3	713209	POS	1b(96%)	A252V, I262V, S335A	EU781827	CADIP	FERPO	42	M	Ignored	-	09/09/2015	INF + RBV	24/05/2017	SFB + DCV	F2/F3	NR
4	4831469	POS	1a(98%)	R300Q, D310G	H77	CADIP	FERPO	38	F	Ignored	-	-	-	-	-	F2/F3	-
5	277962	POS	1a(97%)	-	H77	SAE	JALES	48	F	Sexual	-	Naive				F2/F3	-
6	2788440	POS	3a(95%)	N307G, A335T	D17763	AMHV	SJRPR	56	F	Mani-Pedi	-	Naive				F2/F3	-
7	1477204	POS	1a(97%)	C335S	HQ850279	HEC	CATAN	48	F	Ignored	-	-	-	-	-	-	-
8	563520	POS	1a(97%)	R300Q, Q309R*, V329I	H77	HEC	CATAN	76	F	Blood Transf.	-	-	-	-	-	-	-
9	481619	POS	1a(97%)	V251I, N273S, R300Q, Q309R*	H77	SAE	VOTUP	80	F	Blood Transf.	-	Naive				-	-
10	2973469	POS	1a(98%)	K254R, R300Q, Q309R*, C311Y	H77	AMHV	SJRPR	40	F	Glass Syringe	-	Naive				F2/F3	-
11	5301299	POS	1a(98%)	N291K, R300Q, Q309R*, S335N	H77	CADIP	FERPO	55	F	Ignored	-	-	-	-	-	F3 /F4	-
12	688201	POS	1a(99%)	K254R, R300Q, Q330R	H77	CADIP	FERPO	30	M	Ignored	HIV + HBV	21/11/2012	INF + RBV	06/11/2017	SFB + DCV	F1	NR
13	275985	POS	1b(97%)	A246V, R254K, S335N	EU781827	AMHV	SJRPR	52	M	Blood Transf.	-	10/10/2015	INF + RBV	-	-	F1/A1	-
14	5339278	POS	1b(97%)	-	EU781827	SAE	VOTUP	46	F	Parenteral	-	22/06/2017	SFB + DCV	-	-	F2/F3	-
15	962557	POS	3a(96%)	K304R, N307G, D327A	D17763	AMHV	SJRPR	51	M	Tattoo	-	ABD	-	-	-	-	-
16	1477204	POS	3a(97%)	K304R, N307G	D17763	SAE	VOTUP	58	F	Parenteral	-	04/08/2017	SFB + DCV + RBV	-	-	F2/F3	-
17	2316089	POS	3a(95%)	K304R, N307G	D17763	SAE	VOTUP	74	F	Hemodialysis	-	22/06/2017	SFB + DCV	-	-	F3/F4	-
18	11542707	POS	1a(98%)	R300Q	H77	CADIP	FERPO	38	M	Ignored	HIV + HBV	01/08/2007	INF + RBV	-	-	-	-
19	575718	POS	3a(95%)	K304R, N307G, D330S	D17763	SAE	JALES	43	M	Ignored	-	ABD	-	-	-	-	-
20	570663	POS	1a(96%)	Q309R*	H77	AMHV	SJRPR	46	M	Inhalable Drugs	-	Naive				Cirrhosis	-
21	841176	POS	1b(97%)	I262V, K307R	EU781827	AMHV	SJRPR	67	M	Cutting Material	-	23/06/2017	SFB + DCV + RBV	-	-	F3	-
22	1729132	POS	1a(97%)	R300Q, S326N	H77	AMHV	SJRPR	48	F	Surgery	-	12/06/2015	SFB + DCV + RBV	16/08/2017	SFB + SMV	F2/A1	NR
23	266308	POS	1b(97%)	-	EU781827	SAE	STSUL	63	M	Ignored	-	Naive				-	-
24	2345450	POS	1a(98%)	K254R, R300Q	H77	AMHV	SJRPR	45	M	Inject. Drugs	HIV	09/09/2017	SFB + DCV + RBV	-	-	-	-
25	3554424	POS	1a(99%)	-	H77	HEC	CATAN	51	M	Inject. Drugs	-	-	-	-	-	-	-
26	655482	POS	1b(97%)	-	EU781827	AMHV	SJRPR	46	M	Inject. Drugs	-	01/12/2016	SFB + DCV + RBV	ABD	-	Cirrhosis	-
27	547426	POS	1a(98%)	-	EU781827	HEC	CATAN	63	M	Ignored	-	-	-	-	-	-	-
28	384472	POS	1a(98%)	R300Q	H77	CADIP	FERPO	47	M	Ignored	-	-	-	-	-	F2/F3	-
29	163965	POS	1b(97%)	-	EU781827	HEC	CATAN	68	F	Ignored	-	-	-	-	-	-	-
30	11463009	POS	3a(97%)	Q273H, K304R, N307G	D17763	HEC	CATAN	32	F	Inject. Drugs	-	-	-	-	-	-	-
31	141762	POS	1a(98%)	K254R, R300Q, Q309R*, A327E	H77	SAE	SJRPR	44	M	Sexual	HIV	Naive				-	-

Sample	Viral Load (UI/mL)	PCR DBS	LANL / Blastn NCBI	RAS	NS5B Reference	Institution/City		Age at the time of sampling	Gender	Probable Infection Rote	Co-infection (HIV, HBV)	Beginning of Treatment	Medication	Retreatment	Second Medication	Biopsy[1]	Clinical Response
32	384472	POS	1a(97%)	R300Q, Q309R*, S335N	H77	CADIP	FERPO	46	F	Ignored	-	-	-	-	-	F2/F3	-
33	1840388	POS	1a(98%)	K254R, R300Q, Q309R	H77	CADIP	FERPO	43	F	Ignored	-	18/08/2017	SFB + DCV	-	-	F2/F3	-
34	118393	POS	1a(98%)	R300Q, Q309R*	H77	SAE	JALES	42	M	Sexual	-	Naive				F4	-
35	463563	POS	1b(95%)	R254K, V321I*, S335N	EU781827	SAE	JALES	56	M	Sexual	-	30/06/2014	INF + RBV	20/07/2017	SFB + DCV	A2/F0	NR
36	412055	POS	1b(98%)	I262V	EU781827	SAE	JALES	64	F	Ignored	-	05/03/2015	INF + RBV	07/09/2017	SFB + DCV	F2/F3	ETR
37	10920318	POS	1a(97%)	R300Q, Q309R*	H77	AMHV	SJRPR	54	M	Inhalable Drugs	HIV	ABD	-	-	-	F2/A1	-
38	1328875	POS	3a(97%)	N244D, K304R, N307G	D17763	HEC	CATAN	45	F	Surgery	-	-	-	-	-	-	-
39	271902	POS	3a(97%)	K304R, N307G	D17763	HEC	CATAN	66	F	Surgery	-	-	-	-	-	-	-
40	125141	POS	2c(94%)	K300R, V329A	D50409	HEC	CATAN	71	M	Inject. Medication	-	-	-	-	-	-	-
41	966202	POS	1a(97%)	K254R, R300Q, Q309R*, S335N	H77	HEC	CATAN	54	M	Ignored	-	-	-	-	-	-	-
42	6631071	POS	3a(95%)	N244D, K304R, N307G	D17763	HEC	CATAN	32	M	Sexual	-	-	-	-	-	-	-
43	1231359	POS	2a(99%)	R241Q, H250R, I309V	D00944	SAE	JALES	71	M	Ignored	-	Naive				F2/F3	-
44	7408469	POS	1a(97%)	Q309R*	H77	SAE	JALES	54	M	Inject. Drugs	-	Naive				F2/F3	-
45	1319700	POS	1b(95%)	R254K, V321I*, S335N	EU781827	CADIP	FERPO	69	F	Ignored	-	-	-	-	-	F3/F4	-
46	2297201	POS	1a(98%)	R300Q, Q309R*	H77	AMHV	SJRPR	62	M	Inhalable Drugs	-	29/07/2017	SFB + DCV + RBV	-	-	F3/F4	-
47	1164963	POS	3a(96%)	N307G, R333K, A334S	D17763	SAE	VOTUP	65	M	Parenteral	-	30/03/2017	SFB + DCV	-	-	F2/F3	-
48	5386569	POS	1a(98%)	R300Q	H77	SAE	VOTUP	53	M	Tattoo	-	Naive				F1	-
49	286014	POS	1a (97%)	V251M, R300Q, Q309R*	H77	SAE	VOTUP	46	M	Parenteral	HIV	Naive				-	-
50	30403	POS	1b (97%)	I262V	EU781827	AMHV	SJRPR	57	M	Sexual	HIV	15/06/2017	SFB + DCV + RBV	-	-	Cirrhosis	-
51	2234879	POS	1b (97%)	R254K, Q309R*, C316N*, S335N	EU781827	AMHV	SJRPR	36	M	Inject. Drugs	HIV	15/07/2017	SFB + DCV + RBV	-	-	F2/A1	-
52	115520	POS	1b (96%)	I262V, Q309R*, C316N*, S335N	EU781827	SAE	VOTUP	54	F	Parenteral	-	31/01/2017	SFB + DCV	-	-	-	-
53	1317761	POS	1b (96%)	I262V, Q309R*, C316N*, S335N	EU781827	SAE	VOTUP	34	F	Parenteral	HIV	14/08/2017	SFB + DCV	-	-	-	-
54	4111112	POS	1a (97%)	-	H77	SAE	VOTUP	46	F	Ignored	-	2005	PEGINF + RBV	-	-	-	-
55	75988	POS	1a (97%)	C335S	HQ850279	AMHV	SJRPR	60	M	Blood Transf.	-	ABD	-	-	-	F2/F3	-
56	829600	POS	1a (97%)	R270K	H77	AMHV	SJRPR	47	M	Glass Syringe	-	30/08/2008	PEGINF + RBV	28/10/2017	SFB + DCV + RBV	F3/F4	NR
57	2445009	POS	1b (97%)	A246AT, T329PT	EU781827	AMHV	SJRPR	65	F	Blood Transf.	HIV	21/08/2007	PEGINF + RBV	06/07/2017	SFB + DCV	F2	ABD
58	214835	POS	1a (96%)	R300Q	H77	AMHV	SJRPR	67	F	Surgery	-	24/06/2017	SFB + DCV	-	-	Cirrhosis	-
59	473299	POS	1a (96%)	C316Y*, Q327A, C335S	HQ850279	AMHV	SJRPR	62	F	Blood Transf.	-	11/07/2017	SFB + DCV + RBV	-	-	F3/F4	-
60	2378178	POS	1a (95%)	Q248QR, V251MV, R300Q, Q309R*	H77	SAE	VOTUP	33	F	Inject. Drugs	HIV	Naive				-	-
61	297529	POS	1b (97%)	-	EU781827	SAE	STSUL	70	F	Blood Transf.	-	01/10/2017	SFB + DCV	-	-	-	-
62	426580	POS	1a (97%)	R300Q, Q309R*	H77	AMHV	SJRPR	38	M	Inject. Drugs	HIV	15/07/2017	SFB + DCV + RBV	-	-	F2/A1	-

Sample	Viral Load (UI/mL)	PCR DBS	LANL / Blastn NCBI	RAS	NS5B Reference	Institution/City		Age at the time of sampling	Gender	Probable Infection Rote	Co-infection (HIV, HBV)	Beginning of Treatment	Medication	Retreatment	Second Medication	Biopsy[1]	Clinical Response
63	225511	POS	1a (97%)	R300K	H77	CADIP	FERPO	50	F	Ignored	-	-	-	-	-	-	-
64	76517	POS	1a (96%)	R300Q	H77	HEC	CATAN	53	F	Inject. Medication	-	20/05/2015	PEGINF + RBV	-	-	F2/A1	-
65	376563	POS	3a (98%)	N307G, A334T	D17763	HEC	CATAN	45	M	Inject. Drugs	-	-	-	-	-	-	-
66	870827	POS	1a (96%)	C335S	HQ850279	CADIP	FERPO	68	M	Ignored	-	ABD	-	-	-	-	-
67	128576	NEG				CADIP	FERPO	38	M	Sexual	-	Naive				-	-
68	673620	NEG				HEC	CATAN	28	M	Sexual	HIV	17/01/2018	SFB + DCV	-	-	F2/A1	-
69	39312	NEG				SAE	STSUL	33	M	Parenteral	-	Naive				-	-
70	571623	NEG				AMHV	SJRPR	28	M	Inject. Drugs	-	1995	-	-	ABD	F3/F4	-
71	143741	NEG				AMHV	SJRPR	48	M	Syringe	-	23/12/2016	SFB + DCV	-	-	F3/F4	-
72	266308	NEG				CADIP	FERPO	36	M	Ignored	HIV	-	-	-	-	-	-
73	213351	NEG				SAE	JALES	61	M	Blood Transf.	-	02/04/2018	SFB + DCV	-	-	-	-
74	99563	NEG				SAE	JALES	41	F	Ignored	-	Naive				-	-
75	889118	NEG				HEC	CATAN	55	M	Inject. Drugs	-	25/06/2018	INF + RBV	-	-	F2/F3	-
76	450892	NEG				SAE	STSUL	33	M	Inject. Drugs	HIV	18/07/2017	SFB + DCV	-	-	-	-
77	979684	NEG				SAE	JALES	69	F	Ignored	-	09/02/2018	SFB + DCV + RBV	-	-	-	-
78	125141	NEG				AMHV	SJRPR	38	F	Inject. Drugs	-	05/12/2017	SFB + SMV	-	-	F2/F3	-
79	126011	NEG				HEC	CATAN	50	F	Blood Transf.	-	29/11/2017	SFB + DCV + RBV	-	-	F3	-
80	1265962	NEG				AMHV	SJRPR	52	M	Acupuncture	-	22/11/2004	PEGINF + RBV	25/10/2016	SFB + DCV + RBV	Cirrhosis	NR

Table 3: NS5B Resistance associated substitution (RAS) analysis. NS5B sequences generated for 66 samples were analysed with Geno2Pheno software for the presence of described RAS by genotype to DAAs.

Genotype	Number of patients	Patients with RAS	Amino acid position associated with NS5B RAS								
			L 159 F	D 244 N	S 282 T/R	Q 309 R	D 310 N	C 316 H/N/Y	L 320 F	V 321 A/I	A 333 E
1a	35	16				15		1 - C 316 Y			
1b	17	5				3		3 - C 316 N		2 – V 321 I	
2a	1	0									
2c	1	0									
3a	12	0									

FIGURE 1

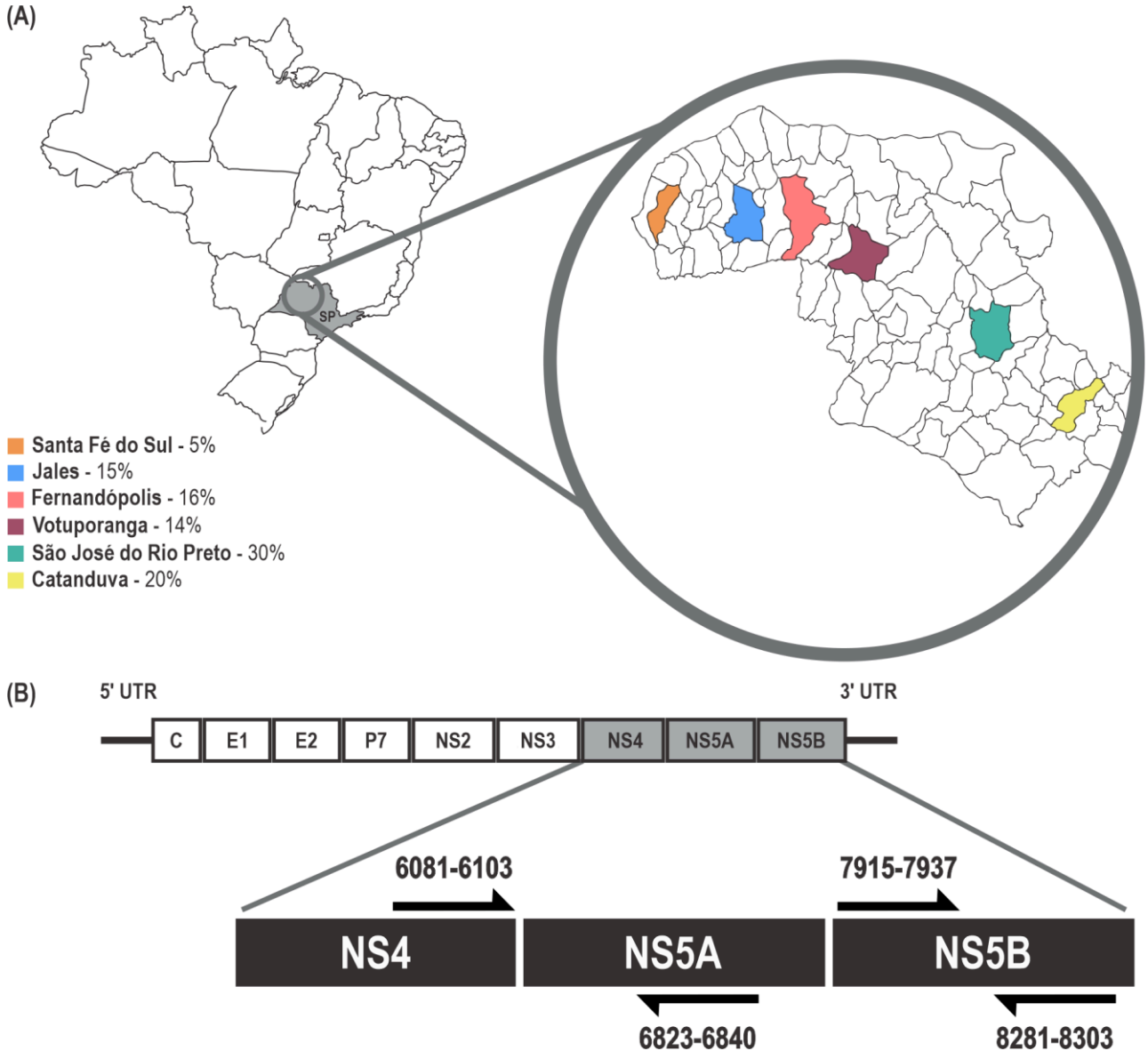

(A)
SP
Santa Fé do Sul - 5%
Jales - 15%
Fernandópolis - 16%
Votuporanga - 14%
São José do Rio Preto - 30%
Catanduva - 20%
(B)
5' UTR
3' UTR
C
E1
E2
P7
NS2
NS3
NS4
NS5A
NS5B
6081-6103
7915-7937
NS4
NS5A
NS5B
6823-6840
8281-8303

FIGURE 2

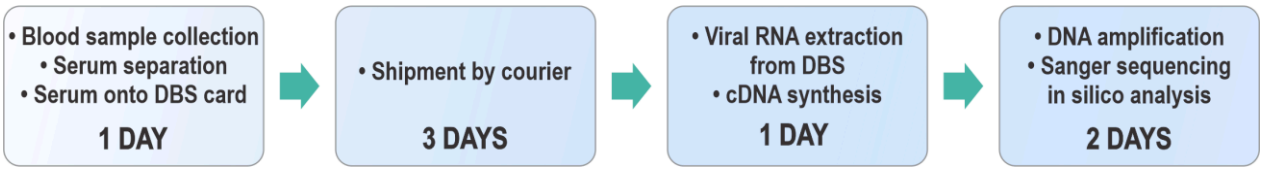
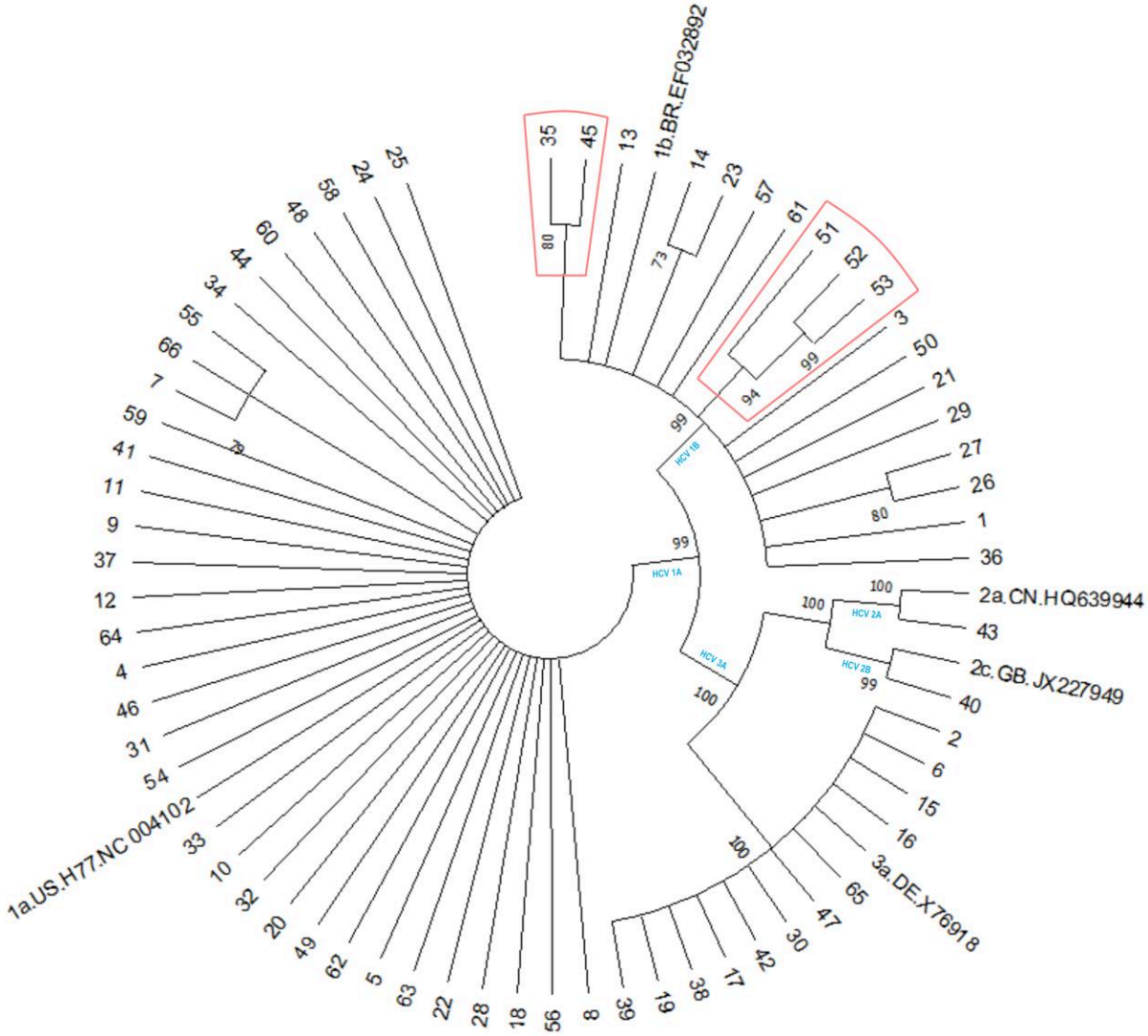

• Blood sample collection
• Serum separation
• Serum onto DBS card
1 DAY
• Shipment by courier
3 DAYS
• Viral RNA extraction from DBS
• cDNA synthesis
1 DAY
• DNA amplification
• Sanger sequencing in silico analysis
2 DAYS

FIGURE 3


1b.BR.EF032892
2a.CN.HQ639944
2c.GB.JX227949
3a.DE.X76918
1a.US.H77.NC 004102
HCV 1B
HCV 1A
HCV 3A
HCV 2A
HCV 2B