



INSTITUTO LEÔNIDAS E MARIA DEANE
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA DA INTERAÇÃO PATÓGENO
HOSPEDEIRO (PPGBIO-INTERAÇÃO)

CARACTERIZAÇÃO GENÉTICA DO SARS-COV-2 NO AMAZONAS, BRASIL

ÁGATHA KELLY ARAÚJO DA COSTA

Manaus – AM

2023



INSTITUTO LEÔNIDAS E MARIA DEANE
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA DA INTERAÇÃO PATÓGENO
HOSPEDEIRO (PPGBIO-INTERAÇÃO)

ÁGATHA KELLY ARAÚJO DA COSTA

CARACTERIZAÇÃO GENÉTICA DO SARS-COV-2 NO AMAZONAS

Dissertação apresentada ao Programa de Pós-Graduação em Biologia da Interação Patógeno Hospedeiro do Instituto Leônidas e Maria Deane – ILMD Fiocruz como requisito para obtenção do título de mestre.

Linha de pesquisa: Bioquímica, biologia celular e molecular de patógenos e seus vetores.

Orientador: Dr. Felipe Gomes Naveca

Co-orientadora: Dra. Lígia Fernandes Abdalla

**Manaus – AM
2023**

FICHA CATALOGRÁFICA

C837c

Costa, Ágatha Kelly Araújo da
Caracterização genética do SARS-COV-2 no Amazonas. / Ágatha
Kelly Araújo da Costa. – Manaus: Instituto Leônidas e Maria Deane, 2023.

61 f.

Dissertação (Mestrado em Biologia da Interação Patógeno-
Hospedeiro) – Instituto Leônidas e Maria Deane, 2023.

Orientador: Prof^o. Dr. Felipe Gomes Naveca.

Co-orientador: Prof^a. Dr^a. Lígia Fernandes Abdalla.

1. Covid-19 2. SARS-CoV-2 I. Título

CDU 616-022.6(813.3)(043.3)

CDD 616.2414

22. ed.

Elaborado por Ycaro Verçosa dos Santos - CRB-11/ 287

Dedico este trabalho ao meu tio Klinger de Oliveira Araújo (in memoriam) que sempre foi um de meus maiores incentivadores e com muito orgulho apoiou minha jornada na Ciência, mas infelizmente não pôde usufruir de seu avanço e foi uma vítima da COVID-19.

AGRADECIMENTOS

Agradeço primeiramente à Fundação Oswaldo Cruz – Instituto Leônidas e Maria Deane (FIOCRUZ-ILMD) por proporcionar a possibilidade de realização deste projeto de pesquisa, por disponibilizar laboratórios, equipamentos, corpo docente e demais servidores tornando essa experiência possível, enriquecedora e agradável.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte financeiro por meio da bolsa de estudos e pela flexibilidade em meio à pandemia.

Ao meu orientador Dr. Felipe Gomes Naveca por tantos ensinamentos, tanta compreensão, tanta paciência. Por ter me permitido viver coisas inesquecíveis, principalmente em meio a uma pandemia, e ser uma pessoa a qual sempre admirarei e reverenciarei.

Ao meu psicólogo Danilo de Matos Areosa Alves, sem o qual a conclusão deste trabalho não seria possível. Agradeço por me acolher em meus piores momentos, me ouvir quando preciso, me mostrar que existe um caminho além de desistir e acreditar em mim quando nem eu mesma consigo.

À minha psiquiatra Núbia Viana Santoro que pacientemente atendeu minhas necessidades e além do acompanhamento médico forneceu laudos necessários.

À Stefanie Costa Pinto Lopes, e demais membros da coordenação, que com muita bondade e compreensão entenderam minhas dificuldades e fizeram tudo que estive ao alcance para me ajudar.

À minha amiga de curso e companheira de laboratório Fernanda Oliveira do Nascimento que vivenciou todas as mesmas alegrias e dificuldades comigo. Obrigada por sempre me incentivar, me dar aquele empurrão necessário e me auxiliar em todas as etapas que precisei.

Ao grupo de pesquisa que tive o orgulho de participar: Matilde, Val, Maju, Débora, Karina, Victor, André, George, Luciana, Michele. Por contribuírem com este trabalho, por me auxiliarem, me fazerem crescer, por compartilharem tantos momentos únicos.

Às minhas amigas de curso Carol e Tallita por dividirem esta jornada, dos momentos mais difíceis aos mais alegres, e sempre me acolherem. Por todo o laço formado e por tornarem tudo mais leve.

À minha mãe Rejane, ao meu pai Habdul e à minha irmã Árina por sempre acreditarem em meu potencial desde o início da minha vida, por serem meu porto seguro em meu momento mais sombrio e pelo privilégio de uma família que respeita a saúde mental. Destaque para minha mãe que seguiu cada pequeno passo ao meu lado e às vezes até caminhou por mim quando não pude.

Aos meus gatos Miss Fame e Vida e meus cachorros Alejandro e Sayuri por serem a minha sanidade e minha fonte de puro amor incondicional.

Aos meus familiares por serem desde sempre meus maiores incentivadores.

Aos meus amigos por me encorajarem, por serem meu alívio necessário e por aplaudirem todas as minhas conquistas.

Enfim, a todos que contribuíram direta ou indiretamente para o início, a execução e a conclusão deste trabalho, muito obrigada.

RESUMO

Introdução: Os coronavírus são vírus pertencentes à família *Coronaviridae* e podem infectar mais de 200 espécies de hospedeiros vertebrados diferentes, causando principalmente manifestações clínicas. Dois coronavírus já foram responsáveis por ocasionar surtos no mundo, sendo eles o SARS-CoV (*Severe Acute Respiratory Syndrome Coronavirus*) e MERS-CoV (*Middle East Respiratory Syndrome Coronavirus*). Em dezembro de 2019 diversos casos de pneumonia sem etiologia conhecida foram notificados na cidade de Wuhan, na China. Análises genômicas de amostras desses pacientes identificaram a presença de um novo Coronavírus nomeado SARS-CoV-2 (*Severe acute respiratory syndrome-related coronavirus 2*). A doença causada pelo SARS-CoV-2 foi denominada COVID-19 (Coronavirus disease 2019) e pode causar sintomas como febre, tosse, fadiga, dor de garganta, cefaleia, dificuldade para respirar, diarreia, perda do olfato ou paladar. Em 11 de março de 2020, a OMS (Organização Mundial de Saúde) declarou a COVID-19 como pandemia. O SARS-CoV-2 é um vírus envelopado e esférico, seu genoma é RNA fita simples linear de polaridade positiva. Até o presente momento, existem mais de 770 milhões de casos confirmados e mais de 6 milhões de óbitos no mundo. No Brasil, o número de casos confirmados é de aproximadamente 37 milhões e mais de 705 mil óbitos. No Amazonas, foram confirmados mais de 638 mil casos e mais de 14 mil óbitos (dados obtidos em 10 de outubro de 2023).

Objetivo: Diante desse cenário, o objetivo deste trabalho foi descrever a introdução e a dinâmica de circulação das linhagens de SARS-CoV-2 no estado do Amazonas no período de março de 2020 a janeiro de 2021.

Metodologia: Para isso, o genoma completo viral foi recuperado através de um protocolo inhouse utilizando a plataforma Illumina. As linhagens de SARS-CoV-2 definidas, identificamos as mutações no genoma dos vírus detectados e caracterizamos a dispersão viral no estado.

Resultados: A análise genômica e epidemiológica de 250 genomas de SARS-CoV-2, verificou-se que o primeiro abupto dos casos se deu principalmente pela dispersão da linhagem B.1.195, que posteriormente foi substituída pela linhagem B.1.1.28. A ocorrência da segunda onda no Amazonas coincidiu com o surgimento da Variante de Preocupação (VOC) P.1 (Gamma).

Conclusão: Esses resultados possibilitam maior compreensão sobre a disseminação do SARS-CoV-2 e os mecanismos envolvidos nas ondas da COVID-19.

Palavras-chave: SARS-CoV-2; COVID-19; Epidemiologia; Amazonas; Brasil; P.1 (Gamma).

ABSTRACT

Introduction: Coronaviruses are viruses belonging to the Coronaviridae family and can infect more than 200 different species of vertebrate hosts, mainly causing clinical manifestations. Two coronaviruses have already been responsible for causing outbreaks around the world, namely SARS-CoV (Severe Acute Respiratory Syndrome Coronavirus) and MERS-CoV (Middle East Respiratory Syndrome Coronavirus). In December 2019, several cases of pneumonia with no known etiology were reported in the city of Wuhan, China. Genomic analyzes of samples from these patients identified the presence of a new Coronavirus named SARS-CoV-2 (Severe acute respiratory syndrome-related coronavirus 2). The disease caused by SARS-CoV-2 was called COVID-19 (Coronavirus disease 2019) and can cause symptoms such as fever, cough, fatigue, sore throat, headache, difficulty breathing, diarrhea, loss of smell or taste. On March 11, 2020, the WHO (World Health Organization) declared COVID-19 a pandemic. SARS-CoV-2 is an enveloped and spherical virus, its genome is linear single-stranded RNA of positive polarity. To date, there are more than 770 million confirmed cases and more than 6 million deaths worldwide. In Brazil, the number of confirmed cases is approximately 37 million and more than 705 thousand deaths. In Amazonas, more than 638 thousand cases and more than 14 thousand deaths were confirmed (data obtained on October 10, 2023).

Objective: Given this scenario, the objective of this work was to describe the introduction and circulation dynamics of SARS-CoV-2 strains in the state of Amazonas in the period from March 2020 to January 2021.

Methodology: For this, the complete viral genome was recovered through an in-house protocol using the Illumina platform. The SARS-CoV-2 lineages were defined, we identified the mutations in the genome of the detected viruses and characterized the viral dispersion in the state.

Results: The genomic and epidemiological analysis of 250 SARS-CoV-2 genomes revealed that the first abrupt occurrence of cases was mainly due to the dispersion of the B.1.195 lineage, which was later replaced by the B.1.1.28 lineage. The occurrence of the second wave in Amazonas coincided with the emergence of the Variant of Concern (VOC) P.1 (Gamma).

Conclusion: These results provide greater understanding of the spread of SARS-CoV-2 and the mechanisms involved in the waves of COVID-19.

Keywords: SARS-CoV-2; COVID-19; Epidemiology; Amazon; Brazil; P.1 (Gamma).

LISTA DE ABREVIACOES E SIGLAS

ACE2 – *Angiotensin-converting enzyme 2*

CDC – *Centers for Disease Control and Prevention*

cDNA – DNA complementar

CEP – Comitê de Ética em Pesquisa

COVID-19 – *Coronavirus disease 2019*

ELISA – *Enzyme-linked immunosorbent assay*

GC – Guanina-Citosina

IBV – *Infectious bronchitis virus*

IgG – Imunoglobulina G

IgM – Imunoglobulina M

kb – Kilobases

MERS – Middle East Respiratory Syndrome

MERS-CoV – Middle East Respiratory Syndrome Related Coronavirus

mRNA – RNA mensageiro

Nsp – Proteínas não estruturais

OMS – Organização Mundial de Saúde

ORF – *Open Reading Frame*

PCR – *Polymerase Chain Reaction*

PEG – Polietilenoglicol

RBD – *receptor-binding domain*

RNA – Ácido ribonucleico

rpm – Rotações por minuto

RT-PCR – *Reverse Transcription – Polymerase chain reaction*

SARS – Severe Acute Respiratory Syndrome

SARS-CoV – Severe Acute Respiratory Syndrome Related Coronavirus

SARS-CoV-2 – Severe Acute Respiratory Syndrome Related Coronavirus 2

SRAG – Síndrome Respiratória Aguda Grave

VOC – Variante de Preocupação

VOI – Variante de Interesse

LISTA DE FIGURAS

Figura 1 – Coronavírus.....	09
Figura 2 – Classificação taxonômica da família <i>Coronaviridae</i>	10
Figura 3 – Representação do genoma do SARS-CoV-2.....	13
Figura 4 – Linha do tempo do surgimento da pandemia de COVID-19.....	15

LISTA DE TABELAS

Tabela 1 – Função das proteínas não estruturais do SARS-CoV-2.....	12
Tabela 2 – Primers para amplificação do genoma completo.....	22

SUMÁRIO

1. REFERENCIAL TEÓRICO	9
1.1 Histórico	9
1.2 Síndromes Respiratórias causadas por Coronavírus em Humanos	11
1.3 SARS-CoV-2	11
1.3.1 Estrutura	11
1.3.2 Genoma	12
1.4 Replicação	13
1.5 Pandemia de COVID-19	14
1.6 Manifestações clínicas	15
1.7 Epidemiologia	16
1.8 Diagnóstico	16
1.9 Controle	17
2. JUSTIFICATIVA	18
3. OBJETIVOS	19
4. METODOLOGIA	20
4.1 Aspectos éticos	20
4.2 Amostras	20
4.3 Extração de RNA	20
4.4 Transcrição Reversa e reação em cadeia da Polimerase em tempo real (RT-PCR) para detecção do SARS-CoV-2	21
4.5 Produção de cDNA	21
4.6 PCR para amplificação do genoma completo	22
4.7 Purificação de amplicons por Polietilenoglicol (PEG)	22
4.8 Sequenciamento	23
4.9 Análise dos dados	23
5. RESULTADOS	24
6. CONCLUSÃO	25
7. REFERÊNCIAS	26
8. ANEXOS	28

1. REFERENCIAL TEÓRICO

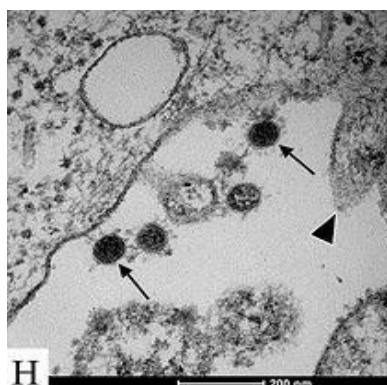
1.1 Histórico

Os coronavírus são vírus pertencentes à família *Coronaviridae* e à ordem *Nidovirales* e podem infectar mais de 200 hospedeiros vertebrados diferentes, entre eles mamíferos e pássaros, causando principalmente manifestações respiratórias, além de entéricas, neurológicas e hepáticas (Siddell et al. 1983, Lai & Cavanaght 2020, Masters 2020). A família *Coronaviridae* é dividida em duas subfamílias, a *Coronavirinae* e a *Toronavirinae*, compreendendo cinco gêneros, 26 subgêneros e 46 espécies (ICTV, 2019).

Análises filogenéticas apontam que o Coronavírus ancestral comum mais recente que infecta mamíferos surgiu 7 a 8 mil anos atrás, enquanto o que infecta aves surgiu há cerca de 10 mil anos (Vijgen et al. 2006). Porém estima-se que sua presença data de 50 a 100 mil anos atrás com a dispersão da população humana pelo planeta, visto que houve início de atividades rurais, como desmatamento de florestas para a criação de animais, o que ocasionou mudanças no ecossistema e na dinâmica do vírus por incluir novos hospedeiros de diferentes espécies (Chan et al. 2020).

Os vírus compreendidos nesta família são envelopados, esféricos, medem 120-220 nm e possuem glicoproteínas de superfície que se assemelham a uma coroa (**Figura 1**), por isso o nome *Coronaviridae* (coroa do latim *corona*) (Yassine Kasmi & Ennaji 2020). O seu genoma é o maior dentre os vírus de RNA, apresenta polaridade positiva e fita simples (Siddell et al. 1983).

Figura 1 – Coronavírus



Microscopia eletrônica do SARS-CoV-2 em 11 de fevereiro de 2020. Fonte: Vieira et al. 2021 (<https://memorias.ioc.fiocruz.br/article/10754/0443-morphology-and-morphogenesis-of-sars-cov-2-in-vero-e6-cells>). Acesso em 18 de janeiro de 2022.

O primeiro coronavírus foi identificado em 1932, o *Infectious bronchitis virus* (IBV), posteriormente nomeado Avian coronavirus, que infecta aves domésticas (*Gallus gallus*), é uma das principais causas de perdas econômicas na indústria avícola, afetando o desempenho tanto das aves de corte como das aves produtoras de ovos (Beaudette & RF, 1958; Cavanagh 2006). Análises na morfologia do IBV verificaram a presença de projeções na superfície viral que o tornavam diferente de outros vírus até então conhecidos, como o vírus Influenza ou os Myxovírus (Berry 1964). Nos anos seguintes outras doenças causadas por novos vírus foram descritas, como encefalomielite e hepatite em ratos, e em 1965 quadros de resfriado em seres humanos. Estudos realizados com esses novos vírus obtidos ao longo do tempo identificaram que todos apresentavam características morfológicas semelhantes às do IBV (Tyrrell et al. 1965). Assim, 30 anos após a detecção da primeira espécie, a família *Coronaviridae* foi descrita, fundamentada por diversos estudos confirmando que estes vírus eram de fato distintos de outros já identificados (Tyrrell et al, 1975).

Na atual classificação (maio de 2022), segundo o Comitê Internacional de Taxonomia de Vírus – ICTV (do inglês *International Committee on Taxonomy of Viruses*), a Família *Coronaviridae* possui três subfamílias, seis gêneros, 28 subgêneros e 54 espécies, como demonstrado na **Figura 2**.

Figura 2 – Classificação taxonômica da família *Coronaviridae*

Virus Taxonomy: 2021 Release
 EC 53, Online, July 2021
 Email ratification March 2022 (MSL #37)
Coronaviridae: 3 subfamilies, 6 genera, 28 subgenera, 54 species

– Family: <i>Coronaviridae</i>
+ Subfamily: <i>Letovirinae</i>
– Subfamily: <i>Orthocoronaviirinae</i>
+ Genus: <i>Alphacoronavirus</i>
– Genus: <i>Betacoronavirus</i>
+ Subgenus: <i>Embecovirus</i>
+ Subgenus: <i>Hibecovirus</i>
+ Subgenus: <i>Merbecovirus</i>
+ Subgenus: <i>Nobecovirus</i>
– Subgenus: <i>Sarbecovirus</i>
Species: <i>Severe acute respiratory syndrome-related coronavirus</i>
+ Genus: <i>Deltacoronavirus</i>
+ Genus: <i>Gammacoronavirus</i>
+ Subfamily: <i>Pitovirinae</i>

Esquema de classificação taxonômica dos Coronavírus, demonstrando família, subfamílias, gêneros, subgêneros e espécies. Fonte:

https://talk.ictvonline.org/ictv-reports/ictv_9th_report/positive-sense-rna-viruses-2011/w/posrna_viruses/222/coronaviridae. Acesso em 10 de abril de 2022.

1.2 Síndromes Respiratórias causadas por Coronavírus em humanos

Dois coronavírus já foram responsáveis por ocasionar surtos principalmente na Ásia e Oriente Médio, sendo eles o SARS-CoV e MERS-CoV.

Em março de 2003, a OMS emitiu um alerta para uma pneumonia de origem desconhecida que desde novembro de 2002 havia acometido mais de 300 pacientes na China, além de casos semelhantes em Hong Kong, Toronto, Singapura e Vietnã (Christian et al. 2004, Peiris et al. 2004). A doença foi denominada Síndrome Respiratória Aguda Grave (SARS, do inglês *Severe Acute Respiratory Syndrome*) e agente etiológico foi identificado como um novo coronavírus que até então não atingia humanos, nomeado SARS-CoV. O hospedeiro intermediário do SARS-CoV foi a civeta, um mamífero que após infectado por morcegos, era vendido em mercados (Drosten et al. 2003).

A SARS se manifesta com febre, tosse, dispneia, cefaleia e falência respiratória. O surto atingiu 29 países em cinco continentes, com 8096 casos confirmados e 774 mortes, em uma taxa de letalidade de aproximadamente 9,5% (Peiris et al. 2004, Gomes et al. 2020).

Uma pneumonia semelhante à SARS surgiu no Oriente Médio em 2012 e, até junho de 2013, 55 casos do surto denominado MERS (*Middle East Respiratory Syndrome*) foram confirmados, resultando em 31 mortes. O agente etiológico foi identificado como o coronavírus MERS-CoV, que também é originário dos morcegos, que infectam camelídeos, de forma que a transmissão do vírus se iniciou por meio de contato direto de seres humanos com os camelídeos e seus fluidos. O MERS-CoV continuou a causar surtos e desde abril de 2012 até janeiro de 2020 foram confirmados 2519 casos e 866 mortes, com uma taxa de letalidade de 34,3% (Zaki et al. 2012, Gomes et al. 2020).

1.3 SARS-CoV-2

O SARS-CoV-2. É um vírus envelopado e esférico que mede aproximadamente 120 a 140 nm de diâmetro. Seu genoma é de RNA fita simples linear de polaridade positiva com 29,9 kilobases (kb) e 11 genes codificam para 12 proteínas (Abu et al. 2020, Junga 2020).

1.3.1 Estrutura

O vírion do SARS-CoV-2 possui quatro tipos de proteínas estruturais: a proteína de espícula (S, do inglês *Spike*), a proteína de envelope (E), a proteína de membrana (M) e a de nucleocapsídeo (N) (Abu et al. 2020, Wang et al. 2020b).

A proteína S é responsável pela entrada do vírion nas células hospedeiras. Ela forma homotrímeros que se deslocam para fora da superfície viral. A proteína é composta por duas subunidades funcionais: S1 e S2. A função da subunidade S1 é ligar-se ao receptor na superfície da célula hospedeira e a função da subunidade S2 é fundir as membranas do vírus e da célula (Wang et al. 2020a).

A **Tabela 1** demonstra as proteínas não estruturais e suas funções:

Tabela 1 – Função das proteínas não estruturais do SARS-CoV-2

Nome	Função
Nsp1	participa do processamento do RNA e a replicação
Nsp2	modula a sinalização da célula hospedeira
Nsp3	supostamente separa as proteínas traduzidas
Nsp4	modifica a membrana do retículo endoplasmático
Nsp5	está envolvida no processamento da poliproteína durante a replicação
Nsp6	domínio transmembrana. Induz formação de autofagossomos e vesículas de dupla membrana
Nsp7	aumentam a interação de nsp12 e o molde de RNA
Nsp8	
Nsp9	liga-se ao RNA viral
Nsp10	é crítica para a metilação do mRNA viral
Nsp12	é a RNA polimerase RNA dependente.
Nsp13	participa do processo de replicação e transcrição
Nsp14	domínio exoribonuclease com atividade de metiltransferase
Nsp15	tem atividade dependente de endoribonuclease
Nsp16	Metiltransferase

Fonte: Adaptações de (Abu et al. 2020, Wang et al. 2020).

1.3.2 Genoma

O primeiro genoma obtido do SARS-CoV-2 foi denominado Wuhan-Hu-1 (NC_045512), sendo isolado e sequenciado na China em janeiro de 2020 (Junga 2020). O genoma apresenta 96% de semelhança com o do coronavírus de morcego BatCoV RaTG13, 80% de semelhança com o SARS-CoV e 50% com o MERS-CoV (Gralinski & Menachery 2020, Wu et al. 2020).

O genoma do SARS-CoV-2 é RNA linear de fita simples e senso positivo. Mede aproximadamente 29,9 kb, o maior entre os vírus de RNA, e contém um cap 5' e uma cauda poli-A 3' (**Figura 3**). Possui 14 fases de leitura abertas (ORFs, do inglês *Open Reading Frames*), 38% de proporção Guanina-Citosina (GC) e codifica 29 proteínas (Abu et al. 2020, Ortiz-prado et al. 2020).

Figura 3 – Representação do genoma do SARS-CoV-2



Esquemática do genoma do SARS-CoV-2, demonstrando as estruturas e ORFs. Fonte: adaptado de Abu et al. 2020.

A ORF1ab é o maior gene e codifica a poliproteína 1ab que é clivada em 15 proteínas não estruturais (nsp1 a nsp10 e nsp12 a nsp16). No terminal 3' estão as 4 proteínas estruturais S, E, M e N além de 8 proteínas acessórias (3a, 3b, 6, 7^a, 7b, 8b, 9b e ORF 14) (Elbe & Buckland-merrett 2017, Shu & Mccauley 2017).

Devido à variabilidade genética, o SARS-CoV-2 foi classificado em linhagens que facilitam a compreensão da epidemiologia e da evolução viral, contribuindo para o acompanhamento da pandemia. Através de análises filogenéticas, três sistemas de nomenclatura de linhagens foram estabelecidos: GISAID, Nextstrain e Pango. Para melhor compreensão da população em geral e para que não haja algum tipo de estigmatização, a OMS recomenda a utilização de letras do alfabeto grego, como Alpha, Beta, Gamma e Delta. No sistema Pango, adotado para a realização deste trabalho, inicialmente duas linhagens principais foram descritas, A e B, e a partir delas outras sublinhagens. De acordo com dados

atualizados até o dia 20 de maio de 2022, foram identificadas 1927 linhagens e sublinhagens virais. Além disso, por conta da contínua evolução viral, algumas linhagens podem ser designadas como Variante de Preocupação – VOC (*Variant of Concern*), Variante de Interesse – VOI (*Variant of Interest*) ou Variante em alerta para monitoramento – VUM (*Variant Under Monitoring*). As VOCs causam aumento na transmissibilidade ou na virulência ou diminuição da efetividade de medidas como diagnóstico, vacinas e tratamento. As VOIs possuem mutações que são conhecidas por alterar características como gravidade da doença, escape do sistema imune ou do tratamento e identificadas em relevantes aumentos de casos em uma região. As VUMs possuem alterações genéticas que são suspeitas de afetar características virais que possam oferecer risco, porém as evidências disso não são concretas, portanto é necessário monitoramento (Rambaut et al. 2020).

1.4 Replicação

A entrada do SARS-CoV-2 nas células é mediada pela proteína S. A subunidade S1 possui o domínio de ligação com o receptor (RBD, do inglês *receptor-binding domain*) que se liga diretamente ao domínio de peptídeo da enzima conversora da angiotensina 2 (ACE2, do inglês *angiotensin-converting enzyme 2*). A região RBD é a mais variável do genoma do SARS-CoV-2. A ACE2 é uma proteína de membrana que participa na maturação da angiotensina, um hormônio que controla a vasoconstrição e a pressão sanguínea. Essa proteína está amplamente expressa no trato respiratório, nas células dos alvéolos, da traqueia, dos brônquios e dos monócitos e macrófagos alveolares (Kuba et al. 2005, Andersen et al. 2020, Wang et al. 2020b).

Depois da ligação do RBD ao receptor ACE2, a proteína S é clivada pela proteína de membrana TMPRSS2, o que ativa a subunidade S2 para a fusão das membranas viral e celular. A entrada do vírus ocorre por endocitose e depois ocorre a liberação do capsídeo viral no citosol da célula (Wilde et al. 2018).

Ocorre então a tradução da ORF1a e ORF1b nas poliproteínas pp1a e pp1ab. As poliproteínas são então processadas por proteases virais o que gera 15 replicases que se unem em um complexo replicação-transcrição que sintetiza RNA de polaridade negativa. A replicação ocorre em uma rede de membranas modificadas do retículo endoplasmático (Kikkert et al. 2008).

A montagem de novos vírions ocorre com a junção de novos RNAs e componentes estruturais. Depois da montagem, os vírions são transportados em vesículas e liberados por exocitose (Ortiz-prado et al. 2020).

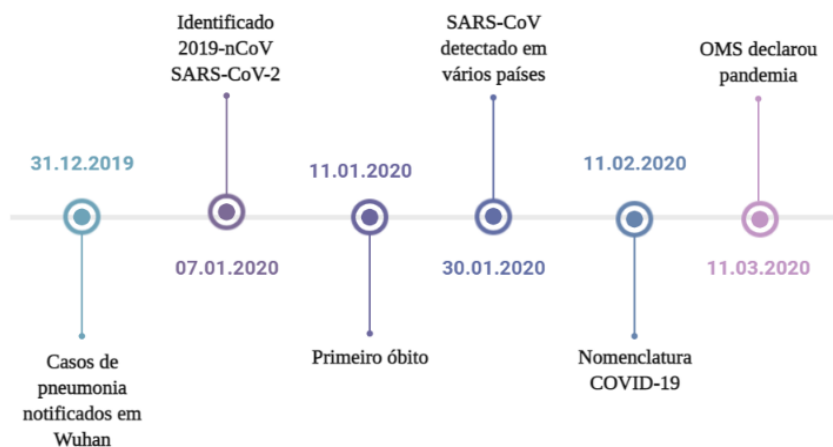
1.5 Pandemia de COVID-19

Em dezembro de 2019, diversos casos de pneumonia sem etiologia conhecida foram notificados na cidade de Wuhan, localizada na província de Hubei, na China (Sohrabi et al. 2020). Investigações apontaram o mercado de frutos do mar como possível fonte inicial da infecção, sendo então este local fechado dias após os primeiros casos serem identificados (Rothan & Byrareddy 2020).

Análises de swab de nasofaringe de pacientes identificaram a presença de um novo Coronavírus, nomeado inicialmente 2019-nCoV (Guan et al. 2019). Posteriormente, o vírus foi renomeado como SARS-CoV-2 (*Severe acute respiratory syndrome-related coronavirus 2*) (Gorbalenya et al. 2020). A doença causada pelo SARS-CoV-2 foi denominada COVID-19 (*Coronavirus disease 2019*) e pode causar sintomas como febre, tosse, fadiga, dor de garganta, cefaleia, dificuldade para respirar, diarreia, perda do olfato ou paladar (Huang et al, 2020).

De 18 a 29 de dezembro de 2019, 5 pacientes foram hospitalizados e um deles veio a óbito. Em 2 de janeiro de 2020, 41 pacientes hospitalizados tiveram o diagnóstico laboratorial de COVID-19 (Huang et al, 2020). Em 22 de janeiro, 571 casos foram reportados em 25 províncias da China e até esse momento havia 17 mortes. Em 25 de janeiro 1975 casos foram confirmados na China, com 56 mortes. Em 30 de janeiro, 7734 casos foram confirmados na China e 90 casos em outros países como Taiwan, Tailândia, Vietnam, Malásia, Nepal, Sri Lanka, Japão, Estados Unidos, Índia, Austrália, Canadá, França e Alemanha (Rothan & Byrareddy 2020). Em 30 de janeiro de 2020 a OMS (Organização Mundial de Saúde) declarou o surto de COVID-19 uma emergência de saúde pública de importância internacional (Sohrabi et al. 2020). Posteriormente, em 11 de março de 2020, a OMS decretou pandemia para a COVID-19 (**Figura 4**).

Figura 4 – Linha do tempo do surgimento da pandemia de COVID-19



Elaborado em Biorender.com. Fontes: Sohrbi et al. 2020; Rothan e Byareddy 2020; Huang et al. 2020.

Os primeiros genomas foram compartilhados na plataforma EpiCoV do GISAID (www.gisaid.org) em 15/12/2019, isso foi fundamental para desenvolvimento de testes de diagnósticos. Em 20/05/2022 contabilizamos 10.959.965 genomas depositados, isso evidencia um movimento inédito na comunidade científica, visto que nunca se sequenciou tantos genomas de um vírus em tão pouco tempo (GISAID, 2022).

1.6 Manifestações clínicas

A infecção pelo SARS-CoV-2 pode ocasionar desde quadros assintomáticos até casos fatais. Estudos apontam que pessoas com infecção assintomática possuem alta capacidade de transmitir o vírus, por isso são peças importantes para o surgimento de novos surtos. No entanto, apesar de diversas hipóteses levantadas, como um padrão diferente de expressão de ACE2, não existem evidências que respondam de forma definitiva o porquê de alguns indivíduos não apresentarem sintomas (Gao et al. 2020).

Dentre os sintomas mais comuns da infecção pelo SARS-CoV-2 estão febre, fadiga e tosse que podem ocorrer no período de uma semana após a infecção. Após 5 a 8 dias do início dos sintomas pode ocorrer dispneia. Outros sintomas comuns são cefaleia, náusea e diarreia (Chen et al. 2020).

A manifestação mais séria da COVID-19 é a pneumonia, que é caracterizada por tosse, dispneia e infiltrados em exames de imagem do tórax. Como complicação da pneumonia pode ocorrer a SRAG, a síndrome respiratória aguda grave. Pode se desenvolver em uma média de 8 dias de sintomas e em 12,3% dos casos é necessária implementação de respiração mecânica (Guan et al. 2019). Estudos indicam que a necessidade de suplementação de oxigênio por sonda nasal foi em 50% dos pacientes hospitalizados, de ventilação mecânica não invasiva em 30% dos pacientes e de ventilação mecânica invasiva em até 3% dos pacientes (Ortiz-prado et al. 2020).

Como principais fatores de risco para a forma mais grave da COVID-19 estão idade maior que 65 anos, obesidade, diabetes e hipertensão (Ortiz-prado et al. 2020).

1.7 Epidemiologia

Em 26 de fevereiro de 2020 foi reportado o primeiro caso confirmado de COVID-19 no Brasil, na cidade de São Paulo (REF). Em seguida foram confirmados casos na Bahia, em Brasília e no Rio Grande do Sul (REF). A região norte foi a última a registrar uma infecção por SARS-CoV-2 (REF). Em 13 de março, uma mulher que viajou para a Inglaterra e retornou para Manaus foi o primeiro caso confirmado no Amazonas (Nascimento et al. 2020, Amazonas 2020, Saúde 2020).

A incidência mundial do SARS-CoV-2 é de mais de 770 milhões de casos confirmados e mais de 6 milhões de óbitos. No Brasil, o número de casos confirmados é de aproximadamente 37 milhões e mais de 705 mil óbitos (World Health Organization, 2023). No Amazonas, foram confirmados mais de 638 mil casos e mais de 14 mil óbitos (Dados obtidos até 10/10/2023).

1.8 Diagnóstico

O padrão ouro para o diagnóstico de SARS-CoV-2 é a RT-PCR em tempo real, uma técnica desenvolvida para detecção de RNA. Ela baseia-se na conversão do RNA em cDNA através da enzima transcriptase reversa e na amplificação do número de cópias do cDNA pela PCR, utilizando primers específicos e sondas fluorescentes. Uma grande vantagem da RT-PCR em tempo real é que a amplificação e a análise dos resultados são realizadas

simultaneamente. A RT-PCR é considerada uma técnica confiável, rápida, sensível e específica (Yüce et al. 2020).

Em fevereiro de 2020, o CDC (do inglês, *Centers for Disease Control and Prevention*) validou um kit de RT-PCR em tempo real e em seguida outros laboratórios tiveram validação aprovada para seus kits (FDA, 2020). Para realização do teste é utilizada amostra de nasofaringe, orofaringe, swabs nasais, aspirado do trato respiratório inferior, lavado broncoalveolar e saliva.

Outra forma de diagnóstico do SARS-CoV-2 é através da detecção de anticorpos. Os testes detectam níveis de IgM ou IgG no sangue, soro ou plasma. A detecção pode ser feita através de testes rápidos ou de ELISA. Apesar de ser uma alternativa prática e rápida, a detecção de anticorpos pode apresentar reação cruzada e podem ocorrer falsos negativos devido ao fato de os anticorpos levarem dias para serem desenvolvidos, sendo detectados apenas dias após a infecção.

O teste de antígeno detecta componentes virais e revela a infecção ativa através de anticorpos. As amostras usadas são de swab de nasofaringe e urina. Essa técnica apresenta alta especificidade porém média sensibilidade (Yüce et al. 2020).

1.9 Controle

A pandemia de COVID-19, causada pelo coronavírus SARS-CoV-2, tem desafiado nações em todo o mundo desde o seu surgimento. Dois pilares cruciais na resposta a essa crise de saúde global são as medidas não farmacológicas (NPIs) e a vacinação. Essas estratégias desempenham papéis complementares na contenção da disseminação do vírus e na redução da gravidade da doença, juntamente com a esperança de restaurar a normalidade à sociedade.

As NPIs, como o distanciamento social, o uso de máscaras faciais, a higienização das mãos regular e as restrições de viagem, foram implementadas em várias fases da pandemia em todo o mundo. Essas medidas têm se mostrado eficazes na redução da transmissão do vírus, principalmente quando aplicadas de maneira coordenada e em resposta à evolução das taxas de infecção em uma determinada região. O distanciamento social, por exemplo, ajudou

a reduzir a propagação do vírus, evitando grandes aglomerações e interações próximas entre as pessoas. O uso de máscaras faciais se tornou uma medida de proteção comum, ajudando a capturar partículas virais e impedir que se espalhem pelo ar. A higienização das mãos e a etiqueta respiratória também desempenham um papel importante na prevenção da disseminação do vírus, especialmente em ambientes de alto risco, como instalações de saúde e locais públicos.

No entanto, as NPIs não são uma solução permanente e têm implicações socioeconômicas significativas, incluindo impactos na saúde mental, na economia e nas atividades cotidianas das pessoas. É por isso que a vacinação se tornou uma ferramenta crucial para controlar a pandemia. As vacinas COVID-19 têm se mostrado altamente eficazes na prevenção de casos graves da doença, hospitalizações e mortes. Além disso, a vacinação em massa é fundamental para alcançar a imunidade de rebanho, um ponto em que uma porcentagem significativa da população está protegida contra o vírus, tornando mais difícil a transmissão comunitária. Isso alivia a pressão sobre os sistemas de saúde, permitindo que eles funcionem de maneira mais eficaz e evitem sobrecargas.

A distribuição e o acesso equitativo às vacinas COVID-19 têm sido desafios importantes em muitas partes do mundo. Além disso, a evolução do vírus e o surgimento de variantes preocupantes destacam a necessidade contínua de monitoramento e adaptação das estratégias de vacinação. As variantes, como a Delta e a Ômicron, demonstraram maior transmissibilidade e, em alguns casos, resistência parcial à imunização. Isso enfatiza a importância de continuar aprimorando as vacinas e adaptando as estratégias de vacinação conforme necessário.

Em resumo, as medidas não farmacológicas e a vacinação são duas abordagens cruciais na resposta à pandemia de COVID-19. As NPIs desempenham um papel fundamental na redução da transmissão do vírus, enquanto a vacinação oferece esperança na prevenção de doenças graves e na restauração da normalidade. Ambas as estratégias devem ser implementadas em conjunto e adaptadas às circunstâncias em evolução para alcançar o controle eficaz da pandemia e proteger a saúde pública.

2. JUSTIFICATIVA

A pesquisa proposta tem como objetivo essencial descrever a introdução e a dinâmica de circulação das linhagens do SARS-CoV-2 no estado do Amazonas durante um período crítico da pandemia, de março de 2020 a janeiro de 2021. O contexto em que esta pesquisa se baseia é o cenário global de uma pandemia sem precedentes, com a COVID-19 causando um impacto significativo na saúde pública, na economia e na sociedade em geral.

A justificativa para esta dissertação é multifacetada. Primeiramente, a compreensão da introdução e dispersão do SARS-CoV-2 é de importância crítica para lidar com a pandemia e tomar decisões informadas sobre medidas de saúde pública. O Amazonas, como muitas outras regiões do mundo, foi duramente afetado pela COVID-19, e a investigação das linhagens do vírus em circulação é vital para entender como a doença se espalhou e evoluiu em uma área específica.

Além disso, o estudo se concentra nas mutações no genoma do vírus, incluindo a identificação de variantes de preocupação, como a VOC P.1 (Gamma), que surgiu durante o período de análise. As variantes do SARS-CoV-2 têm implicações críticas para a transmissibilidade do vírus, a eficácia das vacinas e as estratégias de mitigação da pandemia. Portanto, a pesquisa ajuda a contribuir para o conhecimento em evolução sobre a genética do vírus e seu impacto na saúde pública.

A pesquisa também assume relevância regional, já que se concentra especificamente no Amazonas, que enfrentou desafios únicos devido à sua geografia, densidade populacional e recursos de saúde limitados. Compreender como o vírus se comportou nesse contexto específico pode fornecer insights valiosos para o planejamento de resposta a pandemias em áreas semelhantes no futuro.

Em um contexto global, os resultados deste estudo podem contribuir para o conhecimento científico mais amplo sobre a pandemia de COVID-19 e ajudar a orientar as estratégias de saúde pública em todo o mundo. No geral, esta dissertação é justificada pela necessidade urgente de compreender a disseminação do SARS-CoV-2 e os mecanismos envolvidos nas ondas da COVID-19, fornecendo assim informações cruciais para enfrentar a pandemia e preparar-se para desafios semelhantes no futuro.

3. OBJETIVOS

Geral

Realizar a caracterização genética do SARS-CoV-2 que circularam no Amazonas no período de março de 2020 a janeiro de 2021.

Específicos

1. Descrever as linhagens do SARS-CoV-2 circulantes em Manaus no período de 2020 a 2021;
2. Identificar as mutações presentes no genoma do SARS-CoV-2 de amostras coletadas na região do estudo;
3. Caracterizar a dispersão viral na cidade de Manaus.

4. METODOLOGIA

4.1 Aspectos éticos

Este estudo está incluído no projeto “Vigilância genômica de vírus respiratórios circulantes nos anos de 2019 e 2020 em Manaus, Amazonas”, cuja realização foi aprovada pelo Comitê de Ética em Pesquisa (CEP) da Universidade do Estado do Amazonas (UEA) sob parecer consubstanciado CAAE: 25430719.6.0000.5016.

4.2 Amostras

As amostras utilizadas nesse estudo foram coletadas por de swab de faringe e nasofaringe de 644 residentes dos sexos feminino e masculino do Estado do Amazonas que foram testados em rotina de diagnóstico para SARS-CoV-2 através da metodologia de RT-PCR em tempo real para detecção do alvo XXXXX de SARS-CoV-2 no período de março de 2020 a janeiro de 2021. Dessas, 250 tiveram o genoma completo viral sequenciado pelo protocolo inhouse de SARS-CoV_2. Além disso, diante da emergência da VOC Gamma no estado foi desenvolvido um protocolo de inferência para triagem de vírus com deleção na NSP6 característico das P1/VOCs através de RT-PCR em tempo real e 394 foram avaliadas.

As amostras têm procedência do Instituto Leônidas e Maria Deane – ILMD; do Laboratório Central do Amazonas – LACEN-AM; do Hospital Militar de Área de Manaus – HMAM; e da Fundação Alfredo da Matta – FUAM.

4.3 Extração de RNA

O processo de extração de RNA das amostras foi realizado de modo semi-automatizado com o uso do Maxwell® RSC 48 Instrument (Promega). As amostras foram extraídas utilizando kit de extração Promega - Maxwell® RSC Viral Total Nucleic Acid, seguindo as recomendações do fabricante. Em síntese, é adicionado 20µL de Proteinase K, 200µL de Tampão de lise e 200µL de amostra. Essa mistura é brevemente homogeneizada em vórtex e segue para uma incubação de 10 minutos a 56°C. Em seguida, todo o volume é

transferido para o cartucho do kit acoplado ao robô. O RNA é eluído ao final do processo em 50 μ L de Água Livre de Nucleases e armazenado a -80°C até a sua utilização.

4.4 Transcrição Reversa e reação em cadeia da Polimerase em tempo real (RT-PCR) para detecção do SARS-CoV-2

A detecção do SARS-CoV-2 foi realizada através de três ensaios comerciais diferentes: SARS-CoV-2 (E/RP) – Biomanguinhos; Allplex 2019-nCoV Assay – Seegene; ou protocolo in-house seguindo as diretrizes do USA/CDC.

- SARS-CoV-2 (E/RP) Biomanguinhos: Adiciona-se a um tubo de 1,5mL previamente identificado 393,8 μ L de mistura para PCR, 7,9 μ L de ROX e 648,9 μ L de mix E/RP. Homogeneiza-se a mistura com uma pipeta. Adiciona-se 10 μ L da mistura em cada poço de uma placa óptica de RT-PCR. Em seguida utiliza-se o vórtex para homogeneizar as misturas e um adesivo óptico para selar a placa. Leva-se ao termociclador seguindo o protocolo de corrida recomendado pelo fabricante do kit.
- Allplex 2019-nCoV Assay Seegene: Adiciona-se a um tubo de 1,5mL previamente identificado os seguintes reagentes: 2019-n-CoV MOM, água livre de RNase, 5X real-time One-step buffer e a enzima de RT-PCR. Mistura-se e centrifuga-se brevemente o tubo. Adiciona-se 17 μ L da mistura de reagentes e 8 μ L de RNA. Inicia-se a PCR conforme a programação do fabricante.
- USA/CDC: Adiciona-se a um tubo de 1,5mL previamente identificado o mix de primers e sondas N1, N2 e RNP, água livre de nucleases e a enzima para RT-PCR. Mistura-se os reagentes com auxílio de uma pipeta e então leva-se à centrífuga. Na placa óptica, adiciona 15 μ L da mistura em cada poço e 5 μ L de RNA. Inicia-se a PCR de acordo com o protocolo do fabricante.

4.5 Produção de cDNA

Para a produção do cDNA utilizou-se a enzima SuperScript IV Reverse Transcriptase (Invitrogen) e primers randômicos. O termociclador utilizado foi o ProFlex (Applied

Biosystems). Adicionou-se 7 μ L de RNA e 1 μ L de random primer a 50nM em um microtubo e leva-se ao termobloco por 5 minutos a 65°C. Em seguida adiciona-se ao microtubo o restante do mix composto por água, buffer 5X, DTT a 0,1M, DNTPs a 10mM, Superase e a enzima. Levou-se ao termobloco na seguinte programação: 25°C por 15 minutos / 37°C por 15 minutos / 45°C por 15 minutos / 50°C por 15 minutos / 70°C por 15 minutos / 4°C

4.6 PCR para amplificação do genoma completo

Para a amplificação do genoma completo de SARS-CoV-2 foi utilizada a enzima Platinum Super Fi II Green e conjuntos de primers que compreendem 9 amplicons, conforme descrito na tabela 2. O termociclador foi o Veriti (Applied Biosystems).

Tabela 2 – Primers para amplificação do genoma completo

Amplicon	Forward	Reverse	Bp
1	24_F	3986_R	3962
2	3841_F	7966_R	4125
3	7836_F	11989_R	4153
4	11857_F	16060_R	4203
5	15911_F	20034_R	4123
6	19865_F	24144_R	4279
7	23997_F	26032_R	2036
8	25827_F	W_HU_1_28946R	3120
9	28155_F	29672_R	1515

A ciclagem utilizada foi: 1 ciclo a 98°C por 30 segundos, 35 ciclos de 98°C a 10 segundos; 60°C para os amplicons 1-8 e 55°C para o amplicon 9 por 10 segundos; 72°C por 2 minutos, 1 ciclo a 72°C por 5 minutos e 1 ciclo a 4°C ∞ .

4.7 Purificação de amplicons por Polietilenoglicol (PEG)

Os produtos de PCR foram purificados pelo método de precipitação com Polietilenoglicol (PEG 8000) para remoção de primers que não foram incorporados na PCR, de forma que não interfiram no sequenciamento. Cada amplicon foi purificado separadamente e ao final todos foram quantificados no Qubit.

O volume do produto de PCR é transferido para um tubo de 1,5mL e adiciona-se o mesmo volume de PEG 8000 a 20%. Agita-se o tubo suavemente em vórtex por 10 segundos e leva-se à incubação a 37°C por 15 minutos. Em seguida, o tubo é centrifugado a 6.000rpm por 15 minutos. Então, o sobrenadante é descartado e é adicionado 125µL de etanol a 80% gelado, com posterior centrifugação a 450 rpm por 2 minutos. Novamente o sobrenadante é descartado e o tubo é levado ao SpeedVAC ou banho seco com a tampa aberta para que os resíduos do etanol sequem. Certificando-se que não haja resíduos de etanol, adiciona-se o mesmo volume inicial do produto de PCR de água livre de nucleases. Agita-se o tubo em vórtex por 10 segundos e após deixar o tubo por aproximadamente 2 horas a 5°C para melhor eluição, armazena-se a amostra a -20°C para posterior reação de sequenciamento.

4.8 Sequenciamento

A etapa de sequenciamento foi realizada conforme descrito em Nascimento et al. 2020. Os *amplicons* de cada amostra foram normalizados e agrupados. As bibliotecas foram construídas usando o *Nextera DNA Flex Library Prep* e agrupadas sequenciadas com MiSeq Reagent Micro Kit v2 (500 ciclos), seguindo os protocolos do fabricante. O sequenciamento de nucleotídeos foi realizado na plataforma MiSeq (Illumina, San Diego, CA, Estados Unidos), instalada no ILMD, em uma execução pareada (2x150 ciclos).

4.9 Análise dos dados

As leituras no formato FASTQ foram geradas pelo pipeline da Illumina no BaseSpace (<https://basespace.illumina.com>). Todos os arquivos foram baixados e importados para o Geneious v10.2.6 para a montagem do *contig* de cada amostra, utilizando ferramentas

embutidas como BBDuk e BBMap (v37.25). Em síntese, as sequências foram aparadas, removendo as extremidades que apresentam baixa qualidade, e para evitar qualquer viés assim como regiões de ligação dos iniciadores foram removidas. As sequências foram montadas utilizando a sequências de referência disponível no GenBank (NC_045512.2).

5. RESULTADOS

Como resultado, foi publicado o artigo:

Naveca FG, Nascimento V, de Souza VC, Corado AL, Nascimento F, Silva G, Costa Á, Duarte D, Pessoa K, Mejía M, Brandão MJ, Jesus M, Gonçalves L, da Costa CF, Sampaio V, Barros D, Silva M, Mattos T, Pontes G, Abdalla L, Santos JH, Arantes I, Dezordi FZ, Siqueira MM, Wallau GL, Resende PC, Delatorre E, Gräf T, Bello G. COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence. *Nat Med.* 2021 Jul;27(7):1230-1238. doi: 10.1038/s41591-021-01378-7. Epub 2021 May 25. PMID: 34035535. (ANEXO)

Neste artigo abordamos um estudo epidemiológico realizado em 250 amostras de genomas do SARS-CoV-2 na região do Amazonas, que destaco os seguintes resultados:

O estudo em questão se concentra na análise da diversidade genética das variantes do SARS-CoV-2 que circularam no estado do Amazonas, Brasil, durante a pandemia de COVID-19, no período de março de 2020 a janeiro de 2021. Para isso, foram geradas 250 sequências de genoma completo de alta qualidade do SARS-CoV-2 a partir de amostras de indivíduos em 25 municípios do Amazonas.

Os resultados do estudo revelaram que a primeira fase de crescimento exponencial da COVID-19 no Amazonas foi impulsionada principalmente pela disseminação da linhagem B.1.195. No entanto, essa linhagem foi gradualmente substituída pela linhagem B.1.1.28 entre maio e junho de 2020. A segunda onda da pandemia coincidiu com o surgimento da Variante de Preocupação (VOC) P.1, que evoluiu a partir de uma clado local B.1.1.28 no final de novembro de 2020 e substituiu a linhagem parental em menos de dois meses.

Esses resultados sugerem que as substituições sucessivas de linhagens no Amazonas foram impulsionadas por uma combinação complexa de níveis variáveis de medidas de distanciamento social e pelo surgimento de uma variante P.1 mais transmissível do vírus. Essas descobertas têm implicações importantes para entender os mecanismos subjacentes às ondas epidêmicas de COVID-19 e o risco de disseminação da variante P.1 do SARS-CoV-2 não apenas no Brasil, mas também em escala global.

O estudo também fornece informações valiosas sobre a dinâmica da pandemia no Amazonas e destaca a importância da vigilância genômica contínua do vírus para orientar as

medidas de saúde pública e o desenvolvimento de estratégias eficazes de combate à COVID-19. Além disso, os resultados sugerem que as características genéticas das variantes, como a VOC P.1, podem influenciar a transmissibilidade e a gravidade da doença, enfatizando a necessidade de pesquisas adicionais nessa área.

Sendo um trabalho importante destaco minha contribuição principalmente no diagnóstico e produção dos genomas, através de coletas, extração de material genético e auxílio nas etapas de PCR, purificação, sequenciamento e análise de dados.

6. CONCLUSÃO

Com base nos resultados apresentados no estudo epidemiológico genômico realizado no estado do Amazonas, fica claro que a região foi profundamente afetada pela epidemia de COVID-19, experimentando duas ondas de crescimento exponencial no início e no final de 2020. Durante a primeira fase de crescimento exponencial, a disseminação predominante foi atribuída à linhagem B.1.195, que gradualmente foi substituída pela linhagem B.1.1.28 entre maio e junho de 2020. A segunda onda coincidiu com o surgimento da variante de preocupação (VOC) P.1, que evoluiu a partir de um clado local da linhagem B.1.1.28 no final de novembro de 2020 e substituiu a linhagem parental em menos de dois meses.


Esses achados apoiam a conclusão de que as sucessivas substituições de linhagem no Amazonas foram impulsionadas por uma combinação complexa de medidas variáveis de distanciamento social e pelo surgimento de um vírus VOC P.1 mais transmissível. Esses dados fornecem informações essenciais para a compreensão dos mecanismos subjacentes às ondas epidêmicas de COVID-19 e ao risco de disseminação do VOC P.1 do SARS-CoV-2 no Brasil e, potencialmente, em todo o mundo.

Em resumo, o estudo destaca a importância de uma resposta pública eficaz e adaptável às mudanças nas características do vírus, como as variantes, e a necessidade contínua de medidas de saúde pública para controlar a disseminação da COVID-19, especialmente em regiões altamente afetadas, como o Amazonas, a fim de proteger a saúde da população.

7. REFERÊNCIAS

- Abu A, Naqvi T, Fatima K, Mohammad T, Fatima U, Singh IK 2020. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach.
- Amazonas F de V em S do 2020. Situação Epidemiológica de COVID-19 e da Síndrome Respiratória Aguda Grave no Estado do Amazonas, 2020. 2: 1–11.
- Beaudette CH, RF 1958. INFECTION OF THE CLOACA WITH THE VIRUS OF INFECTIOUS BRONCHITIS. 76: 1958.
- Berry DM 1964. The Structure of Infectious Bronchitis Virus. 403407.
- Chan JF, To KK, Tse H, Jin D, Yuen K 2020. Interspecies transmission and emergence of novel viruses : lessons from bats and birds.
- Christian MD, Poutanen SM, Loutfy MR, Muller MP, Low DE 2004. Severe Acute Respiratory Syndrome. 38.
- D. A. J. Tyrrell, J. D. Almeida, C. H. Cunningham, W. R. Dowdle, M.S. Hofstad, K. McIntosh, M. Tajima, L. Ya. Zakstelskaya, B.C. Easterday AK and RWB, An 1975. Coronaviridae. 82: 76–82.
- Drosten C, Günther S, Preiser W, Werf S van der, Brodt H-R, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RAM, Berger A, Burguière A-M, Cinatl J, Eickmann M, Escriou N, Grywna K, Kramme S, Manuguerra J-C, Müller S, Rickerts V, Stürmer M, Vieth S, Klenk H-D, Osterhaus ADME, Schmitz H, Doerr HW 2003. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *N. Engl. J. Med.* 348: 1967–1976.
- Elbe S, Buckland-merrett G 2017. Data , disease and diplomacy : GISAID ' s innovative contribution to global health. : 33–46.
- Gomes V, Lázaro M, Marielena M, Saivish V 2020. The emergence of SARS , MERS and novel SARS - 2 coronaviruses in the 21st century. *Arch. Virol.*
- Gorbalenya A e., Baker SC, Baric RS, Groot RJ de, Drosten C, Gulyaeva AA, Haagmans bart l., , Chris lauber2 AM leontovich3, Neuman9 benjamin W, Penzar3 D, Perlman10 S, , leo l. M. Poon11, Dmitry V. Samborskiy3 Igor A. Sidorov2 IS and JZ 2020. The species Severe acute respiratory syndrome- related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. 5.
- Gralinski LE, Menachery VD 2020. Return of the coronavirus: 2019-nCoV. *Viruses* 12: 1–8.
- Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, Liu L, Shan H, Lei C, Hui DSC, Du B, Li L, Zeng G, Yuen K, Chen R, Tang C, Wang T, Chen P, Xiang J, Li S, Wang J, Liang Z, Peng Y, Wei L, Liu Y, Hu Y, Peng P, Wang J, Liu J, Chen Z, Li G, Zheng Z, Qiu S, Luo J, Ye C, Zhu S 2019. Clinical Characteristics of Coronavirus Disease 2019 in China. : 1–13.
- Junga BCCHYWXLLRZJZYHLZGFJXXGZCTYJXYWWWXXWYHLMMLYXHGLG 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China.
- Kikkert M, Worm SHE Van Den, Zevenhoven-dobbe JC 2008. SARS-Coronavirus

Replication Is Supported by a Reticulovesicular Network of Modified Endoplasmic Reticulum. 6.

Kristian G. Andersen^{1 2} , 3, Andrew Rambaut⁴ WI, Garry ECH⁵ and RF 2020. The proximal origin of SARS-CoV-2. : 2–4.

Kuba K, Imai Y, Rao S, Gao H, Guo F, Guan B, Huan Y, Yang P, Zhang Y, Deng W, Bao L, Zhang B, Liu G, Wang Z, Chappell M, Liu Y, Zheng D, Leibbrandt A, Wada T, Slutsky AS, Liu D, Qin C, Jiang C, Penninger JM 2005. A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus – induced lung injury. 11: 875–879.

Lai MMC, Cavanaght D 2020. THE MOLECULAR BIOLOGY OF CORONAVIRUSES.

Masters PS 2020. THE MOLECULAR BIOLOGY OF CORONAVIRUSES.

Nanshan Chen*, Min Zhou*, Xuan Dong*, Jieming Qu*, Fengyun Gong, Yang Han, Yang Qiu, Jingli Wang, Ying Liu, Yuan Wei, Jia'an Xia, Ting Yu, Xinxin Zhang, Li Zhang Nanshan Chen*, Min Zhou*, Xuan Dong*, Jieming Qu*, Fengyun Gong, Yang Han, Yang Qiu, Jingli Wa LZ 2020. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study.

Nascimento VA, Corado ADLG, Nascimento FO do, Costa ÁKA da, Duarte DCG, Luz SLB, Gonçalves LMF, Jesus MS De, Fernandes C, Delatorre E, Naveca FG 2020. Genomic and phylogenetic characterisation of an imported case of SARS-CoV-2 in Amazonas State , Brazil. 115: 1–6.

Ortiz-prado E, Simbaña-rivera K, Barreno LG-, Rubio-neira M, Cevallos-robalino D, Sanches-sanmiguel H, Unigarro L, Zalakeviciute R 2020. Clinical, molecular, and epidemiological characterization of the SARS-CoV-2 virus and the Coronavirus Disease 2019 (COVID-19), a comprehensive literature review. *Elsevier*.

Peiris JSM, Guan Y, Yuen KY 2004. Severe acute respiratory syndrome. 10: 88–97.

Rambaut A, Holmes EC, Toole ÁO, Hill V, Mccrone JT, Ruis C, Plessis L, Pybus OG 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology.

Rothan HA, Byrareddy SN 2020. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *Elsevier*.

Saúde M da 2020. Boletim Epidemiológico Especial COE COVID-19. : 1–48.

Shu Y, Mccauley J 2017. GISAID : Global initiative on sharing all influenza data – from vision to reality. : 2–4.

Siddell SG, Anderson R, Cavanagh D, Fujiwara K, Klenk HD, Macnaughton MR, Pensaert M, Stohman SA, Sturman L 1983. Coronaviridae. 189: 181–189.

Sohrabi C, Alsafi Z, Neill NO, Khan M, Kerwan A 2020. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19).

Tyrrell DAJ, Bynoe ML, Obst DRCOG 1965. Cultivation of Novel Type of Common-cold Virus in Organ Cultures.

Vijgen L, Keyaerts E, Lemey P, Maes P, Reeth K Van, Nauwynck H, Pensaert M, Ranst M Van 2006. Evolutionary History of the Closely Related Group 2 Coronaviruses : Porcine Hemagglutinating Encephalomyelitis Virus , Bovine. 80: 7270–7274.

Wang L, Wang Y, Ye D, Liu Q 2020a. Review of the 2019 novel coronavirus (SARS-CoV-2) based on current evidence. *Elsevier*.







- Wang M, Zhao R, Gao L, Gao X, Wang D, Gallagher T 2020b. SARS-CoV-2 : Structure , Biology , and Structure-Based Therapeutics Development. 10: 1–17.
- Wilde AH De, Snijder EJ, Kikkert M 2018. *Host Factors in Coronavirus Replication*. Springer International Publishing.
- Wu D, Wu T, Liu Q, Yang Z 2020. The SARS-CoV-2 outbreak: What we know.
- Yassine Kasmi KK, Ennaji AS and MM 2020. Coronaviridae: 100,000 Years of Emergence and Reemergence.
- Yüce M, Filiztekin E, Özkaya KG 2020. COVID-19 diagnosis —A review of current methods.
- Zaki AM, Boheemen S van, Bestebroer TM, Osterhaus ADME, Fouchier RAM 2012. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *N. Engl. J. Med.* 367: 1814–1820.

8. ANEXO

ARTICLES

nature
medicine

COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence

Felipe Gomes Naveca¹  , Valdinete Nascimento¹, Victor Costa de Souza¹, André de Lima Corado¹, Fernanda Nascimento¹, George Silva¹, Ágatha Costa¹, Débora Duarte¹, Karina Pessoa¹, Matilde Mejía¹, Maria Júlia Brandão¹, Michele Jesus², Luciana Gonçalves³, Cristiano Fernandes da Costa³, Vanderson Sampaio³, Daniel Barros³, Marineide Silva⁴, Tirza Mattos⁴, Gemilson Pontes⁵, Ligia Abdalla⁶, João Hugo Santos⁷, Ighor Arantes⁸, Filipe Zimmer Dezordi⁹, Marilda Mendonça Siqueira¹⁰, Gabriel Luz Wallau⁹, Paola Cristina Resende , Edson Delatorre , , ⁸ , ¹¹, Tiago Gräf¹² and Gonzalo Bello

The northern state of Amazonas is among the regions in Brazil most heavily affected by the COVID-19 epidemic and has experienced two exponentially growing waves, in early and late 2020. Through a genomic epidemiology study based on 250 severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genomes from different Amazonas municipalities sampled between March 2020 and January 2021, we reveal that the first exponential growth phase was driven mostly by the dissemination of lineage B.1.195, which was gradually replaced by lineage B.1.1.28 between May and June 2020. The second wave coincides with the emergence of the variant of concern (VOC) P.1, which evolved from a local B.1.1.28 clade in late November 2020 and replaced

the parental lineage in <2 months. Our findings support the conclusion that successive lineage replacements in Amazonas were driven by a complex combination of variable levels of social distancing measures and the emergence of a more transmissible VOC P.1 virus. These data provide insights to understanding the mechanisms underlying the COVID-19 epidemic waves and the risk of dissemination of SARS-CoV-2 VOC P.1 in Brazil and, potentially, worldwide.

Amazonas state reported the first confirmed SARS-CoV-2 case in Manaus, the state capital, in March 2020 in a traveler returning from Europe¹. By late February 2021,

>306,000 laboratory-confirmed cases and more than 10,400 deaths in Amazonas had been reported². The COVID-19 epidemic in Amazonas is, at the time of

writing, characterized by two exponentially growing curves of cases (Fig. 1a). Epidemiological data from surveillance of severe acute respiratory illness (SARI) and burials indicate that the first wave of the epidemic started in March 2020 and peaked around early May 2020, when the number of cases dropped and then remained roughly stable from June to November 2020. However, in mid-December the

number of cases started to grow exponentially, establishing the second wave of the epidemic.

A new SARS-CoV-2 VOC, designated P.1 and also known as N501Y.V3, recently emerged in Manaus. Lineage P.1 was first detected in four travelers returning to Japan from Amazonas state on 2 January 2021 (ref. ³) and was soon recognized as an emergent lineage in Manaus⁴. The VOC P.1 harbors 21 lineage-defining mutations, including ten in the Spike protein (L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, H655Y and T1027I). The emergence

of P.1 was touted as one of the putative causes of the second wave of COVID-19 in Manaus⁵. However, the precise relationship between circulating SARS-CoV-2 variants and epidemic dynamics in Amazonas remains unclear due to the paucity of viral sequences sampled in this Brazilian state before December 2020.

Results

Evidence of successive SARS-CoV-2 lineage replacements in Amazonas. To acquire a more in-depth understanding of the genetic diversity of SARS-CoV-2 variants circulating in Amazonas state since the early epidemic, we generated 250 SARS-CoV-2 high-quality, whole-genome sequences from individuals living in 25 municipalities, between 16 March 2020 and 13 January 2021 (Fig. 1a,b). Viral sequences were generated at FIOCRUZ Amazônia, which is part of both the Amazonas state health genomics network (REGESAM) and the consortium FIOCRUZ COVID-19 Genomics Surveillance Network of the Brazilian Ministry of Health (<http://www.genomahcov.fiocruz.br/>). Our genomic survey revealed that most sequences were classified into five lineages: B.1.1.28 ($n = 93$, 37%), P.1 ($n = 60$, 24%), B.1.195 ($n = 49$, 20%), B.1.1.33 ($n = 29$,

¹Laboratório de Ecologia de Doenças Transmissíveis na Amazônia, Instituto Leônidas e Maria Deane, Fiocruz, Manaus, Brazil. ²Laboratório de Diversidade Microbiana da Amazônia com Importância para a Saúde, Instituto Leônidas e Maria Deane, Fiocruz, Manaus, Brazil. ³Fundação de Vigilância em Saúde do Amazonas, Manaus, Brazil. ⁴Laboratório Central de Saúde Pública do Amazonas, Manaus, Brazil. ⁵Instituto Nacional de Pesquisas da Amazônia, Manaus, Brazil.

⁶Universidade do Estado do Amazonas, Manaus, Brazil. ⁷Hospital Adventista de Manaus, Manaus, Brazil. ⁸Laboratório de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, Brazil. ⁹Instituto Aggeu Magalhães, Departamento de Entomologia e Núcleo de Bioinformática, Fiocruz, Recife, Brazil. ¹⁰Laboratório de Vírus Respiratórios e Sarampo, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, Brazil. ¹¹Departamento de Biologia, Centro de Ciências Exatas, Naturais e da Saúde, Universidade Federal do Espírito Santo, Alegre, Brazil. ¹²Instituto Gonçalo Moniz, Fiocruz, Salvador, Brazil.

✉ e-mail: felipe.naveca@fiocruz.br; gbellobr@gmail.com

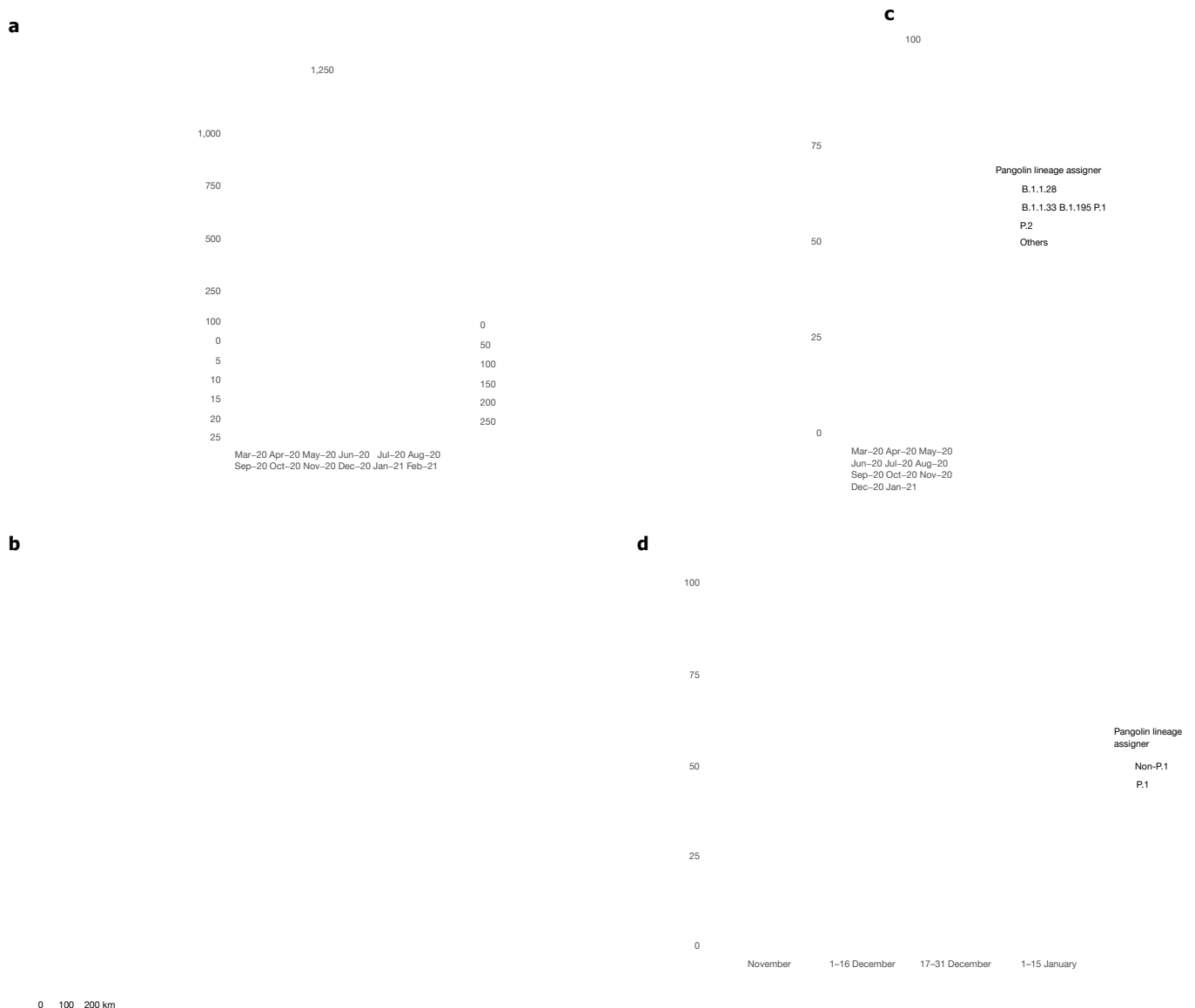


Fig. 1 | Geographic and temporal distribution of SARS-CoV-2-positive samples and viral lineages in Amazonas state. **a**, Graph depicting the temporal evolution of SARI cases and SARI deaths (based on date of symptom onset (source, <http://info.gripe.fiocruz.br>) and burials (source, SEMULSP-Manaus) as a proxy for the COVID-19 epidemic curve in Amazonas state, along with the number of SARS-CoV-2 whole-genome sequences generated in this study and samples genotyped by the P.1-specific real-time RT-PCR protocol. **b**, Map of Amazonas state showing municipalities covered by the SARS-Cov-2 sequences generated in this study. Numbers represent genomes generated for each municipality. **c**, Prevalence of major SARS-CoV-2 lineages between March 2020 and January 2021, estimated from whole-genome sequencing. **d**, Prevalence of SARS-CoV-2 lineages P.1 and non-P.1 between November 2020 and January 2021, estimated from whole-genome sequencing combined with RT-PCR screening.

12%) and P.2 ($n = 9$, 4%) (Supplementary Table 1), with a changing temporal prevalence over time (Fig. 1c). The lineage B.1.195 was the most prevalent during the first exponential growth phase. However, its prevalence gradually decreased after the first epidemic peak in early May and was surpassed by lineage B.1.1.28. This lineage persisted as the most prevalent from May to December 2020, when the second lineage replacement took place coinciding with the second phase of exponential growth. The VOC P.1 was first detected on 4 December 2020 in Manaus, and displayed an extremely rapid increase in prevalence up to January 2021.

To better estimate the temporal trajectory of the emergence of P.1 in Amazonas state in late 2020 and early 2021, we designed a real-time PCR with reverse transcription (RT-PCR) screening assay to detect the deletion at orf1b (NSP6: S106del, G107del, F108del), which is a genetic signature of VOCs P.1, B.1.1.7 and B.1.351. Thus, we evaluated a further 1,232 SARS-CoV-2-positive samples available at the central laboratory of Amazonas state (LACEN-AM) between 1 November 2020 and 31 January 2021. None of the SARS-CoV-2-positive samples genotyped by real-time RT-PCR

before 16 December was positive for the NSP6 deletion, supporting our sequencing results that indicate a very low prevalence of VOC

P.1 before mid-December 2020 in Amazonas. However, positive samples for the NSP6 deletion were frequently detected in the second half of December 2020 and in January 2021. Combining genome sequencing and real-time RT-PCR testing, we found the P.1 lineage in 0% of samples collected in November 2020 ($n = 0/355$), in 8.6% of samples from 1 to 15 December 2020 ($n = 33/384$), in 50.8% of samples from 16 to 31 December 2020 ($n = 177/348$) and in 73.8% of samples from 1 to 15 January 2021 ($n = 487/660$) (Fig. 1d). Similar findings

were observed when analyses were restricted to the city of Manaus: P.1 lineage in 0% of samples collected in November 2020 ($n = 0/211$), in 10.5% of samples from 1 to 15 December 2020 ($n = 32/305$), in 45.1% of samples from 16 to 31 December 2020 ($n = 106/235$) and in 66.0% of samples from 1 to 15 January 2021 ($n = 163/247$). These findings clearly show a sharp increase of lineage P.1 prevalence in Manaus, and in the countryside of Amazonas state, during the second epidemic wave.

Major SARS-CoV-2 local clades drove the COVID-19 epidemic in Amazonas state. To estimate the minimum number of viral

introductions into Amazonas state, we performed maximum likelihood phylogeographic analyses for the large datasets of B.1.1.28 ($n = 674$) and B.1.1.33 ($n = 602$), and Bayesian analysis for the medium-sized B.1.195 dataset ($n = 160$). Time-scaled maximum likelihood trees support at least 28 importation events of wide-spread Brazilian lineages B.1.1.28 ($n = 13$), P.2 ($n = 8$) and B.1.1.33 ($n = 7$) into Amazonas state; mostly from the southeastern (75%) and southern (18%) Brazilian regions (Fig. 2a,b and Extended Data Fig. 1). Despite multiple seeding events, most B.1.1.28 and B.1.1.33 viruses from Amazonas state branched into three highly supported (approximate likelihood-ratio test >80%) local clades (Fig. 2a,b). The clades 28-AM-I and 28-AM-II comprised 49% ($n = 46$) and 26% ($n = 24$) of all B.1.1.28 sequences from Amazonas, respectively, and clade 28-AM-II further gave origin to lineage P.1. The clade 33-AM-I comprises 48% ($n = 14$) of all B.1.1.33 sequences from Amazonas detected here. A Bayesian phylogeographic tree of B.1.195 sequences, sampled worldwide, supports a single importation event of this lineage into Amazonas state from either other Brazilian states or abroad (posterior probability = 0.61 and 0.39, respectively) (Fig. 2c). The clustering of B.1.195 Amazonian sequences in a single, highly supported (posterior probability = 0.92) monophyletic clade (195-AM) was resolved after incorporation of a nine-nucleotide deletion at nsp1 (delta 640–648: K141, S142, F143) as an informative trait in phylogenetic reconstruction. This deletion was the only lineage-defining molecular signature of clade 195-AM, and was also detected in two sequences from the Colombian Amazonian region and one from a Japanese traveler returning from Amazonas to Japan, which were nested within clade 195-AM.

Analysis of the identified Amazonian clades revealed single-lineage-defining mutations in clades 28-AM-I (C29284T), 28-AM-II (A6613G) and 33-AM-I (A28108C, ORF8:Q72P) relative

to other B.1.1.28 and B.1.1.33 Brazilian sequences. Furthermore, only one signature mutation at the Spike (V1176F) distinguishes Amazonian lineages B.1.195 and B.1.1.28, which contrasts with the high number of mutations accumulated by the lineage P.1. A closer inspection of the genetic diversity within the Amazonian clade 28-AM-II further revealed a P.1-like sequence sampled in Manaus on 23 December 2020 that branched basal to the P.1 lineage and also accumulated an unusually high number of genetic changes relative to other B.1.1.28 Brazilian sequences (Fig. 3a). The P.1-like sequence harbors 6/10 P.1 lineage-defining mutations in the Spike protein, including those at the receptor-binding domains K417T, E484K and N501Y. Those mutations raise concern because they may enhance ACE2 affinity and contribute to antibody evasion (Supplementary Table 2)⁶. To investigate the events underlying the emergence of P.1 and P.1-like lineages, we analyzed intrahost SARS-CoV-2 variability in our samples. We found a total of 484 well-supported minor variants (MVs) with a predominance of nonsynonymous ($n = 368$) over synonymous ($n = 105$) substitutions in 126 out of 250 genomes analyzed in this study (Fig. 3b). The frequency of MVs observed in samples taken during the early (March–September) and late

(October–January) epidemic phases was comparable (Extended Data Fig. 2), and no MVs corresponding to the Spike-lineage-defining mutations of clades P.1 and P.1-like were identified (Supplementary Table 3).

Differences in the epidemic trajectory of major SARS-CoV-2 Amazonian clades. Bayesian reconstructions of the spatiotemporal dissemination dynamic, using different molecular clocks (strict and local) and phylogeographic (continuous and discrete) models, support very similar scenarios (Supplementary Tables 4 and 5). These analyses support that the early prevalent local clade 195-AM probably emerged in mid-March 2020 in the city of Manaus and quickly spread from there to several municipalities of the metropolitan region and inner Amazonas state (Fig. 4a,b). The origin of clades 28-AM-I and 28-AM-II was estimated to be in late March

1232

2020 in the area of the Solimões river comprising the municipalities Manacapuru, Iranduba and Manaquiri, and from this area these clades were disseminated to Manaus and other inner municipalities (Fig. 4c–f). The clade 33-AM-I probably arose in mid-April 2020 and remained restricted to Manaus. These analyses traced the most recent common ancestor of lineages P.1 and P.1-like to Manaus in late August 2020, and further indicate that the VOC P.1 probably arose in Manaus around late November 2020 and rapidly spread to other municipalities of the metropolitan region, and also to municipalities located up to 1,100 km distant from Manaus at the borders with Peru, Colombia and Venezuela (Fig. 4e,f).

We next applied the birth–death skyline (BDSKY) model to estimate the effective reproductive number (R_e) of the Amazonian clades with >40 genomes. The estimated R_e trajectories matched the relative prevalence of lineages and social distancing metrics (Fig. 5) very closely. The R_e of clade 195-AM was high (2.6, 95% highest posterior density (HPD): 1.6–3.8) in March, but displayed a steep decrease to 1.0 (95% HPD): 0.8–1.2) in April, coinciding with an increase by >50% of social distancing in Manaus. Clade 28-AM-I, which was estimated to have emerged in rural Amazonas municipalities, also presented a high R_e (2.1, 95% HPD: 1.2–3.4) in its initial spread, reducing to 0.9 (95% HPD: 0.7–1.2) in May when the social distancing index reached 50% in the interior of Amazonas state. The increasing relative prevalence of clade 28-AM-I over 195-AM from April to June agrees with the estimated R_e differences during April. From June to August 2020, the R_e of clades 195-AM and 28-AM-I remained roughly stable at around 1.0. When the social distancing index decreased to <40% in September 2020, 195-AM became apparently extinct while the R_e of 28-AM-I increased to 1.2 (95% HPD: 0.9–1.6) and then remained roughly stable >1.0 up to the end of 2020, leading to an increasing prevalence of 28-AM-I between September and November 2020. The lineage P.1 arose in late November 2020 and displayed a high R_e of 2.6 (95% HPD: 1.5–4.5) during December 2020,

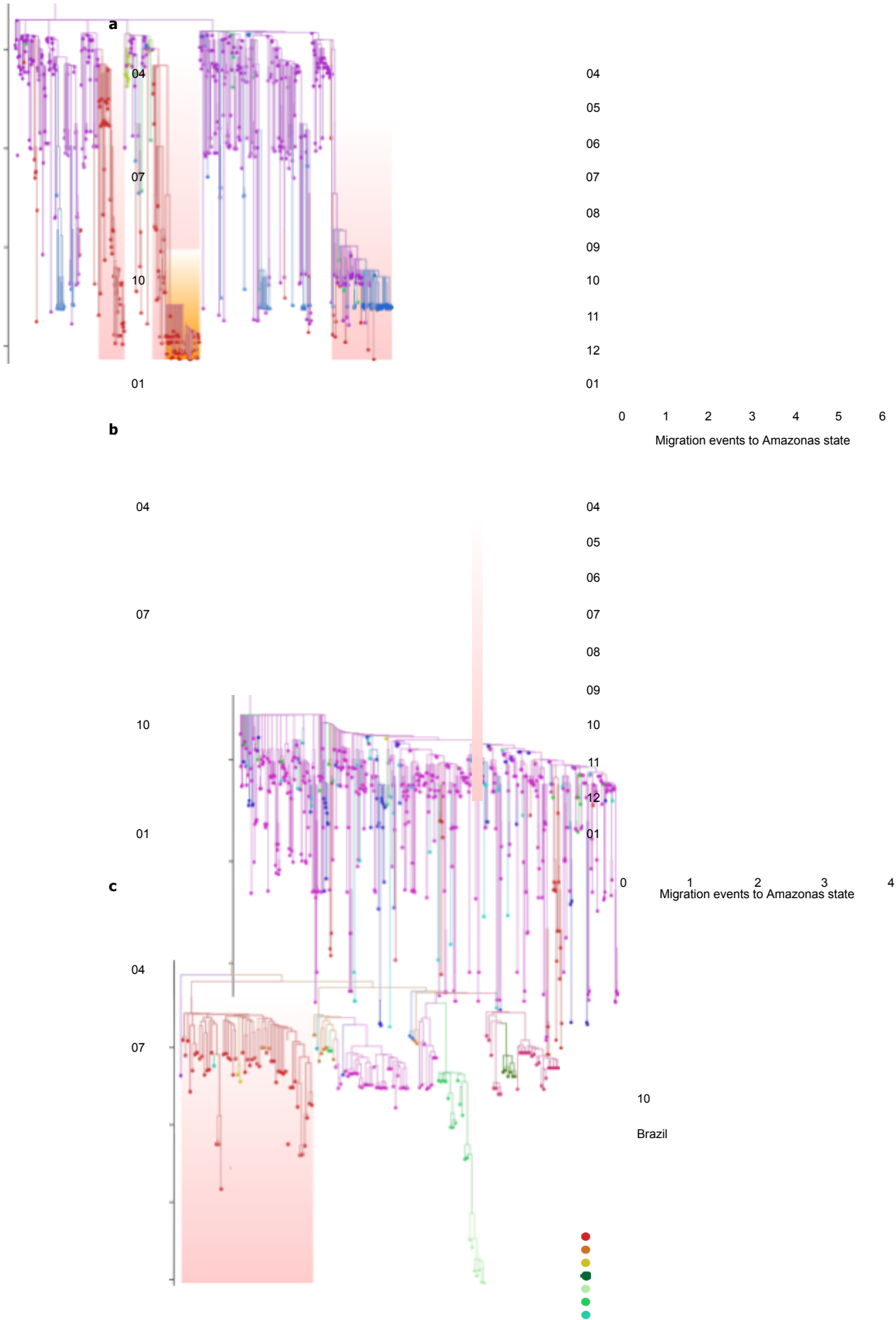
becoming the predominant lineage. With the consequently increased social distance imposed after the health system collapse, this VOC's R_e was estimated to have decreased to 1.2 (95% HPD: 0.9–1.6) in late December and January.

SARS-CoV-2 P.1 infections are associated with higher viral loads. The estimated R_e trajectories of SARS-CoV-2 Amazonian lineages support that VOC P.1 could be more transmissible than the earlier prevalent viral lineages. To test whether this estimated R_e difference might reflect a distinct virological phenotype, we used real-time RT-PCR cycle threshold (Ct) scores as a proxy of the viral load in the upper respiratory tract (URT) of P.1-positive and -negative samples collected at a similar time point of symptom onset (Fig. 6 and Supplementary Table 6). Our analysis revealed a significantly (Mann-Whitney two-sided test $P < 0.0001$) lower median score in URT samples from P.1-infected (Ct = 19.8) compared with non-P.1-infected (Ct = 23.0) subjects, indicating that viral load was

~tenfold higher in P.1 infections than in non-P.1. The median Ct scores for P.1 were significantly lower than for non-P.1 in adult (18– 59 years) men ($P = 0.0005$), adult women ($P < 0.0001$) and elderly (>59 years) women ($P = 0.0149$), but not significantly different in elderly men ($P = 0.4624$). Non-P.1-infected elderly men displayed higher viral load (Ct = 20.8) than younger male adults (Ct = 22.7) and females of all ages (Ct = 23.8). Older P.1-infected males, by contrast, displayed viral load (Ct = 19.8) similar to younger adult males (Ct = 19.4) and females of all ages (Ct = 20.1), suggesting that P.1-infected individuals >18 years old could be equally infectious, irrespective of sex and age.

Discussion

The present study is a comprehensive SARS-CoV-2 genomic investigation performed in Amazonas, one of the Brazilian states most heavily impacted by COVID-19. Our analyses revealed that



Netherlands Oman South Korea Spain
United Arab Emirates Uruguay

01

Japan

Fig. 2 | Origin of major SARS-CoV-2 lineages circulating in Amazonas state. **a,b**, Maximum likelihood phylogeographic analysis of lineages B.1.1.28/ P.1/P.2 (**a**) and B.1.1.33 (**b**) in Brazil. Graphs depict the estimated numbers of introductions into Amazonas state, colored by region of origin. **c**, Bayesian phylogeographic analysis of lineage B.1.195 sequences sampled worldwide. Tips and branches are colored according to sampling location and the most probable location state of their descendent nodes, respectively, as indicated in the legend. Shaded boxes highlight the major SARS-CoV-2 Amazonian clades. The trees were automatically rooted under the assumption of a strict molecular clock, and all horizontal branch lengths are drawn to a scale of years.

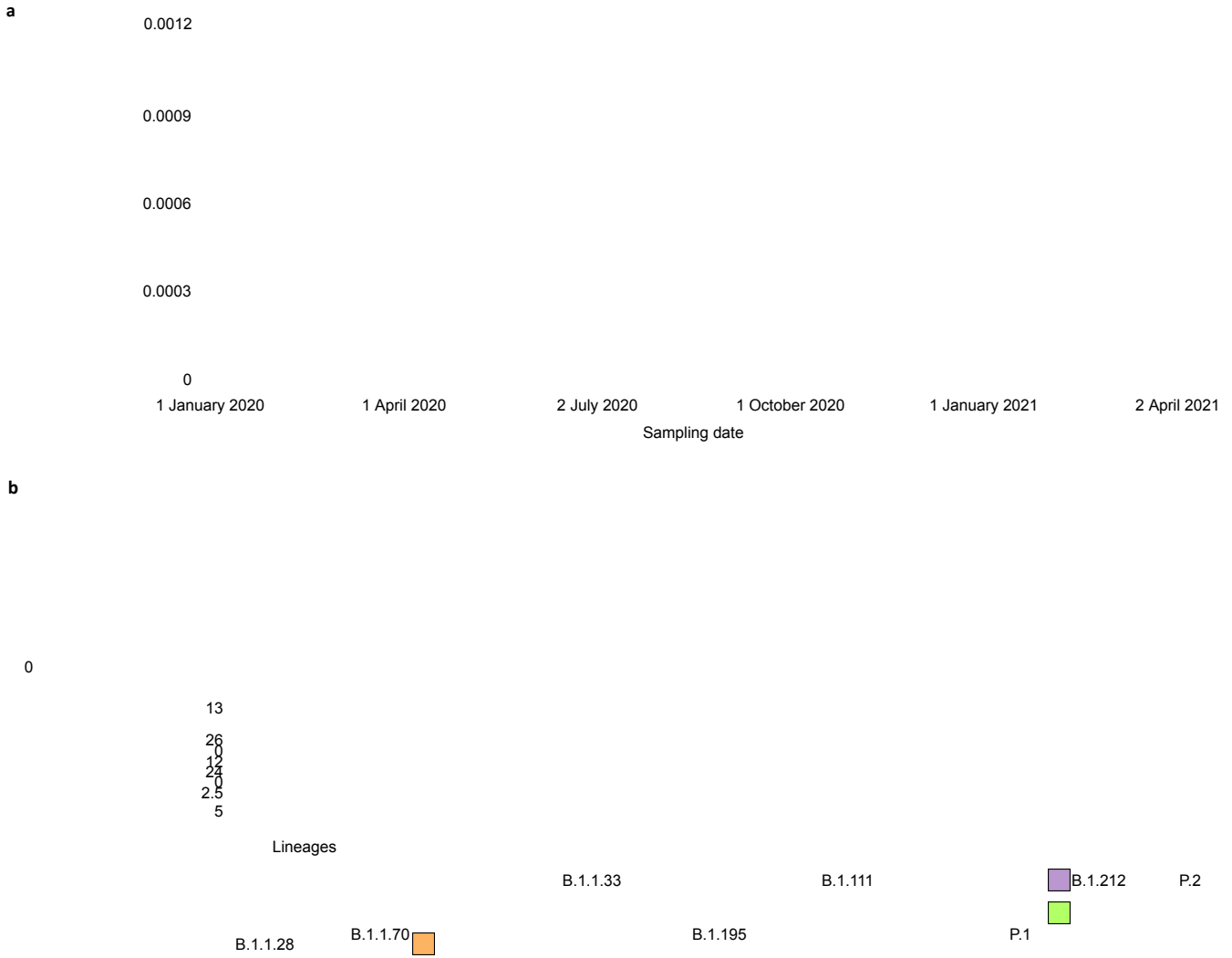


Fig. 3 | Genomic variability of SARS-CoV-2 in Amazonas state. **a**, Correlation between the sampling dates of B.1.1.28, P.1 and P.2 Brazilian sequences and their genetic distance from the root of the maximum likelihood phylogenetic tree. The tree is rooted using the oldest sequence. Sequences from Amazonian clades and VOCs are colored according to the legend at left. **b**, Maximum likelihood phylogenetic tree of the 250 genomes from Amazonas, with lineage assignment and total (black), nonsynonymous (red) and synonymous (blue) nucleotide changes estimated when incorporating MVs into the consensus genome.

most cases in Amazonas were driven by successful dissemination of a few local viral clades that, together, comprise 77% of the 250 SARS-CoV-2 Amazonian genomes sampled between March 2020 and January 2021 rather than multiple importation events. Early major SARS-CoV-2 Amazonian clades arose in either Manaus or the metropolitan region between mid-March and late April 2020 and were widely disseminated within Amazonas state, reaching the most isolated inner localities. We found almost no evidence of the spread of early local SARS-CoV-2 Amazonian lineages outside the state, supporting that Amazonas was not a major hub of viral dissemination within Brazil during 2020. Increasing travel during Christmas and New Year celebrations, combined with the emergence of a potentially more transmissible VOC, however, have changed this scenario, and lineage P.1 has rapidly spread across several Brazilian states up to March

2021 (<http://www.genomahcov.fiocruz.br>).

Two SARS-CoV-2 lineage replacements characterized the COVID-19 epidemic in Amazonas during early and late 2020. The first of these started after the primary epidemic peak and was a gradual process over nearly 5 months during which lineage B.1.1.28 progressively substituted lineage B.1.195, which was the dominant variant during the first epidemic wave. Our analyses suggest that

1234

epidemiological rather than virological factors probably drove this first lineage replacement. The SARS-CoV-2 Amazonian clades 28-AM-I and 28-AM-II, which became the dominant variants in the phase between peaks, displayed only single-lineage-defining synonymous mutations with one difference at the Spike protein when compared to clade 195-AM, and these evolved at a relatively constant rate

between April and November 2020. The most notable difference was that clade 195-AM arose in the city of Manaus, and the R_e value was considerably reduced around mid-April when social distancing in Manaus increased to >50%. Clades 28-AM-I and 28-AM-II, by contrast, arose outside the city of Manaus and the R_e of clade 28-AM-I remained >1.0 until mid-May 2020, when social distancing outside the capital city increased to >50%. When mitigation measures were relaxed and the social distancing index fell by <40% in September 2020, the R_e of clade 28-AM-I returned to >1.0 while clade B.1.195 became extinct, completing the lineage replacement process. Thus, the lower social distancing observed in Amazonas state inner municipalities compared to Manaus was the probable driver of the first lineage replacement.

A modeling study with data from blood donors, conducted in Manaus, estimated that the first wave of SARS-CoV-2 infected 76% (95% CI: 67–98) of the city's population by October 2020, suggesting

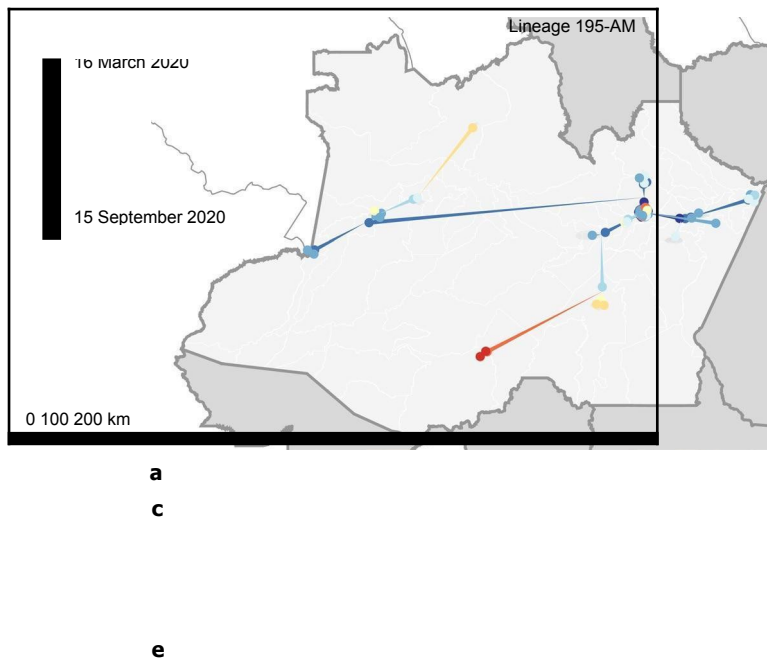


Fig. 4 | Spatial dissemination of the major SARS-CoV-2 Amazonian clades. **a–f**, Phylogeny branches were arranged in space according to internal node locations inferred by the continuous phylogeographic model for lineages 195-AM (**a,b**), 28-AM-I (**c,d**) and 28-AM-II (**e,f**). Branches are colored according to time, as indicated by the legend. Blue shaded areas depict the phylogeographic estimation uncertainty as the 80% highest posterior density interval of the nodes. **b,d,f**, Close-up views of migration events in the Manaus metropolitan area for each lineage. Stars represent the estimated location of the most recent common ancestor of each lineage; squares represent nodes comprising the P.1 lineage.

that the theoretical threshold for herd immunity was reached by mid-2020 and that a second COVID-19 wave would not be expected so soon⁷. Several hypotheses were proposed to explain the unexpected second wave that resulted in the collapse of the health system in Manaus between December 2020 and January 2021 (ref. ⁵). One hypothesis is that lineage P.1 might evade immunity generated in response to a previous infection and has the potential

to reinfect convalescent individuals. While some cases of reinfection with lineage P.1 were described in Manaus⁸, the extent to which reinfections effectively contribute to both onward transmission of SARS-CoV-2 and the surge of cases in the second wave in Amazonas remains controversial^{4,9}. Another hypothesis is that the SARS-CoV-2 attack rate in Manaus was overestimated. Our study supports that a drastic reduction in median R_e (from 2.1–2.6 to

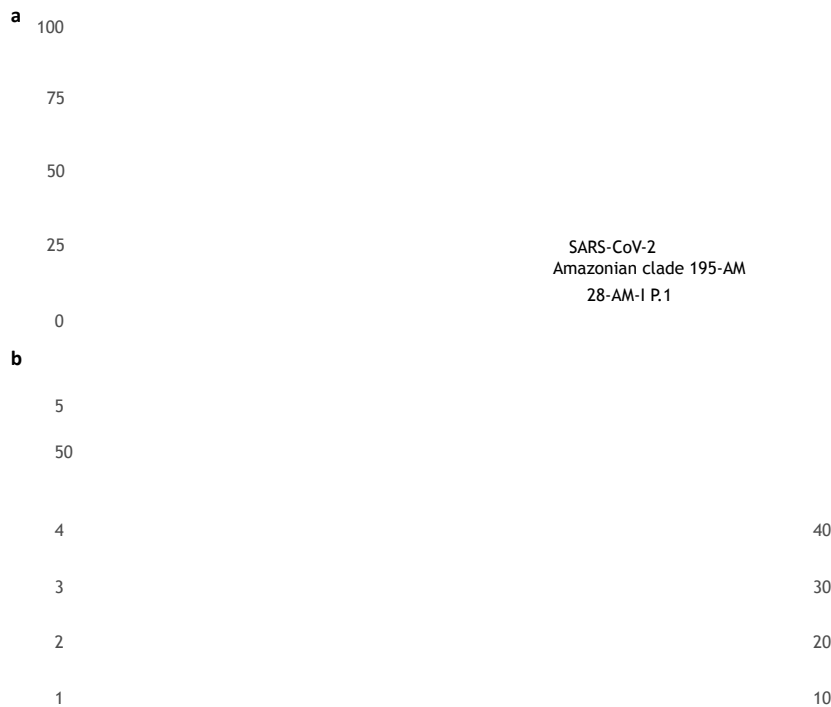


Fig. 5 | Epidemic temporal trajectories of major SARS-CoV-2 Amazonian clades. **a**, Relative prevalence of viral Amazonian clades between March 2020 and January 2021 estimated from whole-genome sequencing. **b**, Graph depicting the temporal variation in R_e of viral Amazonian clades, estimated using the BDSKY approach along with social distancing data trends, reported as percentage change measured against baseline in Manaus (dashed line) and outside Manaus (solid line). Lines represent the weekly average (source, <http://inloco.com.br>).

0.9–1.0) for Amazonian lineages B.1.195 and B.1.1.28 occurred around April–May 2020, consistent with independent epidemiological modeling studies^{10,11}, coinciding with the timing of implementation of nonpharmaceutical interventions (NPI) that effectively increased social distancing. Although those NPIs were not sufficiently stringent to consistently reduce R_e to <1.0 , they maintained a stationary state of low endemic viral community transmission from May to September 2020 (refs. ^{10,11}). These findings support that the NPIs implemented in Amazonas brought the first epidemic wave under relative control before herd immunity became established, and were sufficiently effective to provide population ‘herd protection’ until December 2020 (ref. ¹¹).

Mitigation measures were relaxed from September 2020 onwards, and the R_e of clade 28-AM-I returned to >1.0 . Nevertheless, the second epidemic wave started only in December 2020, coinciding with the emergence of the VOC P.1 and the second lineage replacement event. Several complementary evidences support that these events were probably driven by the emergence of a more transmissible variant in a context of relaxed social distancing. First, the second lineage replacement event was an abrupt process because the VOC P.1 emerged around late November 2020 and it required <2 months to become the dominant variant in Amazonas. This epidemiologic trajectory was later repro-

duced in other Brazilian states (<http://www.genomahcov.fiocruz.br/>). Second, the estimated median R_e of the VOC P.1 during December 2020 was 2.2-fold higher than that estimated for clade 28-AM-I in the same period, indicating that P.1 could have been nearly twofold more transmissible than the co-circulating B.1.1.28 parental lineage, consistent with two recent independent studies^{4,9}. Third, the level of SARS-CoV-2 RNA (estimated from the median Ct) in the URT samples from P.1 infections, particularly from adults (18–59 years), was ~tenfold higher than the level

1236

detected in non-P.1 infections, suggesting that P.1-infected adult individuals are more infectious than those harboring non-P.1 viruses^{4,9}. Fourth, recent experimental evidence supports that VOC P.1 displayed both higher affinity for the human receptor ACE2 and increased resistance to antibody neutralization^{6,12–14}, which might provide a substantial selective advantage for transmission of P.1 over other lineages.

Understanding the factors that drive the emergence and expansion of VOCs harboring multiple key mutations in the receptor-binding domain of the Spike protein is of crucial importance. One hypothesis is that the emergence of VOCs

resulted from a major change in the selective environment, probably imposed by partial herd immunity in heavily affected regions within which SARS-CoV-2 was evolving¹⁵. Our study of the Amazonian clades that locally evolved between April and December 2020, however, revealed no unusual patterns of intra- or interhost viral variability, showing that the local emergence of VOC is an evolutionary event that is challenging to anticipate from the analysis of parental lineages. Another hypothesis is that VOCs result from selective pressures and adaptation of the virus during prolonged individual infections¹⁶. However, the identification—in our study and during subsequent surveillance conducted in Manaus between December 2020 and January 2021 (ref. ¹⁷)—of P.1-like viruses that harbor several of the P.1 lineage-defining mutations indicates that P.1 mutations did not accumulate in a single, long-term infection but were acquired in sequential steps as observed in the VOC B.1.351 (ref. ¹⁸). The finding that P.1 and P.1-like viruses probably share a most recent common ancestor in late August 2020 further supports that SARS-CoV-2 variants carrying mutations of concern had circulated in Manaus for some time before the emergence of lineage P.1. Although only the lineage P.1 seems to have displayed a rapid dissemination to date, our findings alert for the potential spread of other P.1-related VOC in Amazonas state, or in other Brazilian states.

Fig. 6 | Estimation by RT-PCR of viral load in the URT of patients infected with P.1 versus non-P.1 viruses. **a–f**, Evaluation of sample collection time from the onset of symptoms between groups for those samples with available information: $n = 96$ (40 P.1, 56 non-P.1) (**a**), $n = 464$ (157 P.1, 307 non-P.1) (**b**), $n = 161$ (44 P.1, 117 non-P.1) (**c**), $n = 184$ (67 P.1, 117 non-P.1) (**d**), $n = 45$ (19 P.1, 26 non-P.1) (**e**) and $n = 46$ (20 P.1, 26 non-P.1) (**f**). Comparison of Ct values among different groups by sex and age range. Horizontal bars represent Ct medians and IQR. Two-sided P values for the nonparametric Mann–Whitney test are shown for each group. Exact P values are shown above, apart from those <0.0001 ; two-sided P values <0.05 were considered statistically significant.

It is important to stress that our study has some limitations. First, biased sampling across Brazilian states might influence phylogeographic reconstructions of within-country B.1.128 and B.1.1.33 lineage migrations, so the inferred number of importations into Amazonas should be interpreted as lower-bound estimates. Second, within-country spread of lineages B.1.1.28 and B.1.1.33 was inferred using a machine learning

phylogeographic approach that does not account for uncertainty in phylogenetic reconstruction, and thus the routes of viral migration described here are plausible hypotheses among alternatives not fully explored in our analyses. Third, estimates of R_e obtained here may be influenced by local epidemiological dynamics,

making it challenging to extrapolate the difference in viral transmissibility between P.1 and non-P.1 variants observed in Amazonas to other geographic regions. Fourth, although we removed potential confounders (for example, comparison of only PCR results using the same RNA extraction/real-time RT-PCR protocols and sampled at similar timepoints from symptom onset), Ct comparisons have weaknesses and must be analyzed with caution. VOC P.1 may cause more prolonged infections with a similar peak in viral concentration than non-P.1 lineages, as was recently described for B.1.1.7 (ref. ¹⁹). Moreover, we have no data regarding disease severity for group comparison. Therefore, the difference observed here should be confirmed in other geographic settings, including analysis of Ct dynamics in longitudinal sampled patients with different disease outcomes.

1238

In summary, our findings support that lineage replacements were a recurrent phenomenon in the local evolution of SARS-CoV-2 in Amazonas state, driven by ecological and virological factors. Our findings also indicate that NPIs deployed in Amazonas state in April 2020 were sufficiently effective to reduce the R_e of early prevalent local SARS-CoV-2 clades but were insufficient to keep the epidemic under control, allowing the establishment and local persistence of several endemic viral lineages and subsequent emergence of the VOC P.1 in late November/early December 2020. The lack of efficient social distancing and other mitigation measures probably allowed a sudden and accelerated transmission of VOC P.1. At the same time, the higher transmissibility of this VOC further fueled the rapid upsurge in SARS-CoV-2 cases and hospitalizations observed in Manaus following its emergence. Importantly, phylodynamic modeling indicates that NPIs implemented in Manaus since early January 2021 (Supplementary Note) effectively reduced the median R_e of the VOC P.1 by approximately 50%. Therefore, our results suggest that weak adoption of NPIs represents a risk for the continuous emergence of new variants. Implementation of efficient mitigation measures, combined with widespread vaccination, will be crucial to controlling the spread of SARS-CoV-2 VOCs in Brazil.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-021-01378-7>.

Received: 25 February 2021; Accepted: 29 April 2021;
Published online: 25 May 2021

References

1. Nascimento, V. A. D. et al. Genomic and phylogenetic characterisation of an imported case of SARS-CoV-2 in Amazonas State, Brazil. *Mem. Inst. Oswaldo Cruz* **115**, e200310 (2020).
2. Fundação em Vigilância e Saúde do Amazonas. *Boletim diários dos casos de COVID-19*. https://www.fvs.am.gov.br/media/publicacao/21_02_21_BOLETIM_DIARIO_DE_CASOS_COVID-19.pdf (2021).
3. Fujino, T. et al. Novel SARS-CoV-2 variant identified in travelers from Brazil to Japan. *Emerg. Infect. Dis.* <https://doi.org/10.3201/eid2704.210138> (2021).
4. Faria, N. R. et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* <https://doi.org/10.1126/science.abh2644> (2021).
5. Sabino, E. C. et al. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet* **397**, 452–455 (2021).
6. Dejnirattisai W. et al. Antibody evasion by the P.1 strain of SARS-CoV-2. *Cell* <https://doi.org/10.1016/j.cell.2021.03.055> (2021).
7. Buss, L. F. et al. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science* **371**, 288–292 (2021).
8. Naveca, F. et al. Three SARS-CoV-2 reinfection cases by the new Variant of Concern (VOC) P.1/501Y.V3. Preprint at *Res. Sq.* <https://doi.org/10.21203/rs.3.rs-318392/v1> (2021).
9. Coutinho, R. M. et al. Model-based estimation of transmissibility and reinfection of SARS-CoV-2 P.1 variant. Preprint at *medRxiv* <https://doi.org/10.1101/2021.03.03.21252706> (2021).
10. Mellan, T. A. et al. Subnational analysis of the COVID-19 epidemic in Brazil. Preprint at *medRxiv* <https://doi.org/10.1101/2020.05.09.20096701> (2020).
11. He, D., Artzy-Randrup, Y., Musa, S. S. & Stone, L. The unexpected dynamics of COVID-19 in Manaus, Brazil: was herd immunity achieved? Preprint at *medRxiv* <https://doi.org/10.1101/2021.02.18.21251809> (2021).
12. Hoffmann, M. et al. SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies. *Cell* <https://doi.org/10.1016/j.cell.2021.03.036> (2021).
13. Garcia-Beltran, W. F. et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* <https://doi.org/10.1016/j.cell.2021.03.013> (2021).
14. Wang, P. et al. Increased resistance of SARS-CoV-2 variant P.1 to antibody neutralization. *Cell Host Microbe* <https://doi.org/10.1016/j.chom.2021.04.007> (2021).
15. Martin, D. P. et al. The emergence and ongoing convergent evolution of the N501Y lineages coincides with a major global shift in the SARS-CoV-2 selective landscape. Preprint at *medRxiv* <https://doi.org/10.1101/2021.02.23.21252268> (2021).
16. McCormick, K. D., Jacobs, J. L. & Mellors, J. W. The emerging plasticity of SARS-CoV-2. *Science* **371**, 1306–1308 (2021).
17. Resende, P. C. et al. The ongoing evolution of variants of concern and interest of SARS-CoV-2 in Brazil revealed by convergent indels in the amino (N)-terminal domain of the Spike protein. Preprint at *medRxiv* <https://doi.org/10.1101/2021.03.19.21253946> (2021).
18. Tegally, H. et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* <https://doi.org/10.1038/s41586-021-03402-9> (2021).
19. Kissler, S. M. et al. Densely sampled viral trajectories suggest longer duration of acute infection with B.1.1.7 variant relative to non-B.1.1.7 SARS-CoV-2. Preprint at *medRxiv* <https://doi.org/10.1101/2021.02.16.21251535> (2021).

NATURE MEDICINE | VOL 27 | JULY 2021 | 1230–1238 | www.nature.com/naturemedicine

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

Methods

SARS-CoV-2 samples and ethical aspects. We collected nasopharyngeal and pharyngeal swabs from 644 residents in Amazonas state with available demographic data (320 male, median age 44 years (interquartile range (IQR) = 31.0–57.7); 324 female, median age 43 years (IQR = 30.2–56.0)) that were positively tested by real-time RT–PCR as a routine diagnostic for COVID-19 using any of the following commercial assays: SARS-CoV2 (E/RP) (Biomanguinhos), Allplex 2019-nCoV Assay (Seegene) or an in-house protocol following US Centers for Disease Control and Prevention (CDC) guidelines (<https://www.fda.gov/media/134922/download>). Among those 644 nasopharyngeal and pharyngeal swab samples, 250 were submitted to nucleotide sequencing and 394 were evaluated only for P.1/VOCs by the real-time RT–PCR developed in this study. Fiocruz/ ILMID is one of the official laboratories designated for SARS-CoV-2 testing under the auspices of a network coordinated by the Amazonas State Health Foundation (FVS-AM) and the Brazilian Ministry of Health. This study was conducted at the request of the SARS-CoV-2 surveillance program of FVS-AM. It was approved by the Ethics Committee of Amazonas State University (no. 25430719.6.0000.5016), which waived signed informed consent.

Detection of SARS-CoV-2 P.1/VOCs by RT–PCR. A total of 1,626 SARS-CoV-2- positive samples collected between 1 November 2020 and 31 January 2021 (including those 394 with demographic data) were submitted to a real-time RT–PCR screening test designed for the detection of VOCs that use a forward primer (P.1/VOCs-FNF 5'-GGGTGATGCGTATTATGACATGGTTGG), a reverse primer (P.1/VOCs-FNR 5'-CTAGCACCATCATCATAACACAGTTCTTGC) and a probe (P.1/VOCs-FNP 5' FAM (ZEN)- TGGTTGATACTAGTTTGAAGCTAAAA), to detect the ORF1b deletion (NSP6: S106del, G107del, F108del) found in the three VOCs (P.1, B.1.1.7 and B.1.351). Both primers were used at 300 nM and the probe at 150 nM (final concentration), with TaqMan one-step Fast Virus master Mix (ThermoFisher Scientific, no. 4444434). All real-time RT–PCR data collected in this experiment were acquired using the QuantStudio 5 Real-Time PCR System and QuantStudio design & analysis software v.1.4.1 (ThermoFisher Scientific). We previously validated this assay against 185 high-quality, full SARS-CoV-2 genomes, 59 non-P.1 and 126 P.1 (Supplementary Table 7). All oligos used in this study were manufactured by IDT DNA. Because we have not detected the circulation of VOCs B.1.1.7 and B.1.351 in Amazonas state, we use the frequency of NSP6 deletion among real-time RT–PCR positives as a reliable proxy for frequency of the VOC P.1.

SARS-CoV-2 amplification and sequencing. A total of 250 SARS-CoV-2-positive samples (122 male, 128 female; median age 43 years, IQR = 32–46) collected from residents of 25 out of 62 municipalities in Amazonas state, including the capital Manaus, between 16 March and 13 January were subjected to amplification and next-generation sequencing as previously described¹, now with a reduced number of amplicons (nine rather than 15) of mean average size ~3,500 bp (ref. ⁹). Briefly, RNAs were extracted with the Maxwell RSC Viral Total Nucleic Acid Purification Kit (Promega, no. AS1330) and then converted to complementary DNA with Superscript IV reverse transcriptase (ThermoFisher Scientific, no. 18090200). Amplicons were amplified with SuperFi II Green PCR master mix (a proofreading DNA polymerase with >300× Taq fidelity from ThermoFisher Scientific (no. 12369010)), precipitated with PEG 8000 (Promega, no. V3011) and quantified using a fluorimeter. Normalized pooled amplicons of each sample were used to prepare next-generation sequencing libraries with Nextera XT (no. FC-131-1096) and clustered with 500 cycles of MiSeq Reagent Kit v.2 (no. MS-102-2003) on 2× 250 cycles or 2× 150 cycles (no. MS-103-1002) of paired-end runs. All sequencing data were collected using the MiSeq sequencing platform and Miseq Control software v.2.6.2.1 (Illumina).

SARS-CoV-2 whole-genome consensus sequences and genotyping. FASTQ reads were generated by the Illumina pipeline at BaseSpace (<https://basespace.illumina.com>). All files were downloaded and imported into Geneious v.10.2.6 for trimming and assembly using a customized workflow employing BBDuk and BBDMap tools (v.37.25) and the NC_045512.2 RefSeq as a template. Using a threshold of at least 50% to call a base, we generated consensus sequences with mean depth coverage

of 2,689× (95% CI of mean 2,376–3,002), with only eight genomes having <1,000× depth coverage. The mean number of mapped reads was 512,967 (95% CI of

mean 441,779–584,154), covering at least 98.9% of the RefSeq genome. The final consensus sequences had at least 94% bases with Q score = 30, zero ambiguities and were carefully inspected when a disagreement with RefSeq was observed.

Coverage and pairwise identity percentages, as well as the total number of mapped reads (without duplicates) to RefSeq and the percentage of high-quality bases in consensus, were calculated for all consensus files and are shown in Supplementary Table 1. Consensus sequences were initially assigned to viral lineages according to the nomenclature proposed by Rambaut et al.²⁰, using the Pangolin web application (<https://pangolin.cog-uk.io>) and later confirmed by phylogenetic analyses.

Intrahost SARS-CoV-2 genomic variability. Raw sequencing reads and primer sequences were removed with Trimmomatic v.0.26 (ref. ²¹) using default parameters. Reads that passed quality filtering were then mapped against the

reference genome (NC_045512.2) using the Bowtie2 software v.2.3.5.1 (ref. ²²). A bed file was generated with bedtools v.2.15.0 (ref. ²³), SAMtools v.1.10 (ref. ²⁴) and vcftools v.0.1.13 (ref. ²⁵) using the internal parameter vcf-annotated (parameters-filter Qual = 20/MinDP = 100/SnpGap = 20), meaning that only those nucleotide variants supported by reads with mapping quality >20 and at least 100 sequencing coverage depth would be retained in the intermediate variant call file. To characterize the viral intrahost population, we identified all MVs found in the samples—that is, nucleotides highly supported by 10–49% of the reads in a given position and that were not included in the final majority consensus genome. We then replaced the nucleotides supported by the majority of reads by MVs in the consensus genome to evaluate the impact of synonymous and nonsynonymous nucleotide variation between major and minor variants. We performed the synonymous and nonsynonymous analysis using an R pipeline developed for SARS-CoV-2 (ref. ²⁶) with R v.4.0.3 and RStudio v.1.4.1103.

Discrete maximum likelihood and Bayesian phylogeography. All high-quality (<1% N, or non-identified nucleotide) complete (>29 kb) SARS-CoV-2 genomes of lineages B.1.1.28 ($n = 512$) and B.1.1.33 ($n = 595$) sampled in Brazil, and of lineage B.1.195 sampled worldwide ($n = 110$), that were available on GISAID (<https://www.gisaid.org/>) as of 13 January 2021, were downloaded. SARS-CoV-2 complete genome sequences were aligned using MAFFT v.7.475 (ref. ²⁷). The B.1.1.28 and B.1.1.33 datasets were subjected to maximum likelihood phylogenetic analysis using IQ-TREE v.2.1.2 (ref. ²⁸) under a general time-reversible (GTR) model of nucleotide substitution with a gamma-distributed rate variation among sites and four rate categories (G4), a proportion of invariable sites (+I) and empirical base frequencies (+F), as selected by the ModelFinder application^{29,30}. Branch support was assessed by the approximate likelihood-ratio test based on the Shimodaira–Hasegawa-like procedure with 1,000 replicates. Time-scaled phylogeographic maximum likelihood phylogenetic trees of Brazilian B.1.1.28 and B.1.1.33 datasets were reconstructed using Treetime v.0.8.1 (ref. ³¹), with a fixed substitution rate (8×10^{-4} substitutions per site per year) coupled with an ancestral character reconstruction of epidemic locations using PASTML v.1.9.15 (ref. ³²) with marginal posterior probabilities approximation and an F81-like model. A time-scaled Bayesian phylogeographic analysis was performed for B.1.195 sampled worldwide using the Bayesian Markov chain Monte Carlo (MCMC) approach, implemented in BEAST v.1.10.4 (ref. ³³) with BEAGLE library v.3 (ref. ³⁴), to improve computational time. The Bayesian tree was reconstructed using the GTR + F + I + G4 nucleotide

substitution model, the nonparametric Bayesian skyline model as the coalescent tree prior³⁵, a strict molecular clock model with a uniform substitution rate prior ($8-10 \times 10^{-4}$ substitutions per site per year) and a reversible discrete phylogeographic model³⁶ with a continuous-time Markov chain (CTMC) rate reference prior³⁷. Additionally, the nine-nucleotide deletion at nsp1 (delta 640–648: K141, S142, F143) characteristic of the 195-AM clade was incorporated as an informative trait in phylogenetic reconstruction, and transitions were modeled with a symmetric CTMC rate prior. Three MCMC chains were run for 100 million generations and then combined to ensure stationarity and good mixing. Convergence (effective sample size >200) in parameter estimates was assessed using TRACER v.1.7 (ref. ³⁸). The maximum clade credibility (MCC) tree was summarized with TreeAnnotator v.1.10. Maximum likelihood and MCC trees were visualized using FigTree v.1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Continuous Bayesian phylogeography. The phylogenetic diffusion of SARS-CoV-2 clades from Amazonas state identified by the maximum likelihood analysis (195-AM, 28-AM-I and 28-AM-II) was estimated with the heterogeneous relaxed random walk model and a Cauchy distribution³⁹, previously applied to SARS-CoV-2 in Brazil⁴⁰, using BEAST v.1.10.4 (ref. ³³) as explained above. We used strict and local molecular clock models^{41,42} with a uniform substitution rate prior ($8-10 \times 10^{-4}$ substitutions per site per year) to estimate evolutionary rates. Viral spatial-temporal diffusion was analyzed and visualized in SPREAD v.1.0.7 (ref. ⁴³), and further projected in maps generated with QGIS v.3.10.2 software (<http://qgis.org>) using public access data downloaded from the GADM v.3.6 database (<https://gadm.org>). For each lineage and molecular clock model, one MCMC chain was run for 150 million generations and stationarity and mixing were checked as explained above.

Data for social distancing trends. The social distancing trends were obtained from a commercial company (<http://inloco.com.br>). Inloco's isolation index analyzes people's movements for different levels, states, cities and microregions inferred from proprietary technology. Thus, the higher the index the greater the degree of isolation estimated from the site. This index has been used by several Brazilian states' decision-making authorities since the beginning of the pandemic.

Estimation of R_e . To estimate the R_e of the Amazonian SARS-CoV-2 clades over time we used the BDSKY model⁴⁴ implemented within BEAST 2 v.2.6.2 (ref. ⁴⁵). The sampling rate (d) was set to zero for the period before the oldest sample and then estimated from the data. The BDSKY prior settings were as follows: become uninfected rate (exponential, mean = 36); reproductive number (log normal, mean = 0.8, s.d. = 0.5); sampling proportion (beta, alpha = 1, beta = 100). Origin parameter was conditioned to root height, and R_e was estimated in a piecewise

manner over six time intervals (monthly) to the 195-AM clade, five time intervals (bimonthly) to the 28-AM-I clade and two equal time intervals to the P.1 clade. Time intervals were defined from the date of the most recent sample up to the root of the tree. The molecular clock and substitution model were as in the phylogeographic analysis. One MCMC chain was run for 20 million generations and then checked for stationarity and mixing, as explained above.

Statistical analysis. Descriptive statistics, testing for normal distribution (D'Agostino and Pearson and Anderson–Darling) and the nonparametric Mann–Whitney test were used to compare the Ct of SARS-CoV-2 RT–PCR-positive samples from the URT of patients infected with P.1 versus non-P.1 viruses. To avoid bias, only Ct values from samples analyzed by the same RNA extraction method (Promega Maxwell) and the same real-time RT–PCR diagnostic assay (for example, the CDC assay) were compared. The threshold for statistical significance was set to $P < 0.05$ using two-sided tests. Graphics and statistical analyses were performed using GraphPad v.9.01 and v.9.02 (Prism Software).

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

Data availability

All SARS-CoV-2 genomes generated and analyzed in this study are available at the EpiCoV database in GISAID (<https://www.gisaid.org>) under IDs EPI_ISL_792560, EPI_ISL_801386–801403, EPI_ISL_811148, EPI_ISL_811149, EPI_ISL_833131–833140, EPI_ISL_1034304–1034306, EPI_ISL_1068078–1068292 and EPI_ISL_1661250–1661252. Figures 1a and 5b were created with data provided by <http://info.gripe.fiocruz.br>, SEMULSP-Manaus and FVS-AM. Administrative areas presented in Fig. 4 were provided by the GADM v.3.6 database (<http://gadm.org>). Detailed results of MV detection are available at GitHub (https://github.com/dezordi/mFinder/tree/naveca_et_al_2021/supplementary_data). Source data are provided with this paper.

Code availability

We have made available all scripts used for MV detection (<https://github.com/dezordi/mFinder>) and BEAST XML files (https://github.com/dezordi/mFinder/tree/naveca_et_al_2021/phylo_files) in GitHub folders.

References

20. Rambaut, A. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* <https://doi.org/10.1038/s41564-020-0770-5> (2020).
21. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
22. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
23. Quinlan, A. R. BEDTools: the Swiss-Army tool for genome feature analysis. *Curr. Protoc. Bioinformatics* **47**, 11.12.1–34 (2014).
24. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
25. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
26. Mercatelli, D. & Giorgi, F. M. Geographic and genomic distribution of SARS-CoV-2 mutations. *Front. Microbiol.* **11**, 1800 (2020).
27. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
28. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating

- maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
29. Tavaré, S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences* https://www.damtp.cam.ac.uk/user/st321/CV_&_Publications_files/STpapers-pdf/T86.pdf (1986).
30. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermini, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
31. Sagulenko, P., Puller, V. & Neher, R. A. TreeTime: maximum-likelihood phylodynamic analysis. *Virus Evol.* <https://doi.org/10.1093/ve/vex042> (2018).
32. Ishikawa, S. A., Zhukova, A., Iwasaki, W. & Gascuel, O. A fast likelihood method to reconstruct and visualize ancestral scenarios. *Mol. Biol. Evol.* **36**, 2069–2085 (2019).
33. Suchard, M. A. et al. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, vey016 (2018).
34. Suchard, M. A. & Rambaut, A. Many-core algorithms for statistical phylogenetics. *Bioinformatics* **25**, 1370–1376 (2009).
35. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192 (2005).
36. Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* **5**, e1000520 (2009).
37. Ferreira, M. A. R. & Suchard, M. A. Bayesian analysis of elapsed times in continuous-time Markov chains. *Can. J. Stat.* **36**, 355–368 (2008).
38. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* <https://doi.org/10.1093/sysbio/syy032> (2018).
39. Lemey, P., Rambaut, A., Welch, J. J. & Suchard, M. A. Phylogeography takes a relaxed random walk in continuous space and time. *Mol. Biol. Evol.* **27**, 1877–1885 (2010).
40. Candido, D. S. et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science* **369**, 1255–1260 (2020).
41. Drummond, A. J. & Suchard, M. A. Bayesian random local clocks, or one rate to rule them all. *BMC Biol.* **8**, 114 (2010).
42. Yoder, A. D. & Yang, Z. Estimation of primate speciation dates using local molecular clocks. *Mol. Biol. Evol.* **17**, 1081–1090 (2000).
43. Bielejec, F., Rambaut, A., Suchard, M. A. & Lemey, P. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics* **27**, 2910–2912 (2011).
44. Stadler, T., Kühnert, D., Bonhoeffer, S. & Drummond, A. J. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Natl Acad. Sci. USA* **110**, 228–233 (2013).
45. Bouckaert, R. et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).

Acknowledgements

We thank all the health care workers and scientists who have worked hard to deal with this pandemic threat, the GISAID team and all the EpiCoV database submitters—in particular, the Japanese National Institute of Infectious Diseases members T. Sekizuka, K. Itokawa, R. Tanaka and M. Hashino—for publication of genomes. A GISAID acknowledgment table containing sequences used in this study is shown in Supplementary Table 8. We also thank N. Faria for sharing unpublished findings regarding the SARS-CoV-2 B.1.1.28 lineage. We also appreciate the support of Genomic Coronavirus Fiocruz Network members and the Respiratory Viruses Genomic Surveillance Network of the General Laboratory Coordination of the Brazilian Ministry of Health, Brazilian Central Laboratory States and the Amazonas surveillance teams for their partnership in viral surveillance in Brazil. Funding support is acknowledged from Fundação de Amparo à Pesquisa do Estado do Amazonas – FAPEAM (PCTI-EmergeSaude/AM call 005/2020 and Rede Genômica de Vigilância em Saúde), Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (grant no. 403276/2020-9) and Inova Fiocruz/Fundação Oswaldo Cruz (grant no.

VPPCB-007-FIO-18-2-30—Geração de conhecimento), received by F.G.N., and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro – FAPERJ (grant no. E-26/202.896/2018), received by G.B. F.G.N. and G.B. are also supported by CNPq through their productivity research fellowships (306146/2017-7 and 302317/2017-1, respectively). In loving memory of F. R. Naveca, C. R. dos Santos, V. L. Costa de Souza and all the relatives and colleagues we have lost to COVID-19.

Author contributions

F.G.N. contributed to writing of the report, data analysis, laboratory management and obtaining financial support. V.N., V.C.d.S., A.d.L.C., F.N., G.S., A.C., D.D., K.P., M.M., M.J.B., M.J. and L.G. contributed to diagnostics and sequencing analysis. C.F.d.C., V.S., D.B., M.S., T.M., G.P., L.A. and J.H.S. contributed to patient and public health surveillance data. I.A. and F.Z.D. contributed to formal data analysis of sequence diversity. M.M.S., G.L.W., P.C.R., E.D., T.G. and G.B. contributed to formal data analysis and writing and editing of the report.

Competing interests

The authors declare no competing interests.

Additional information

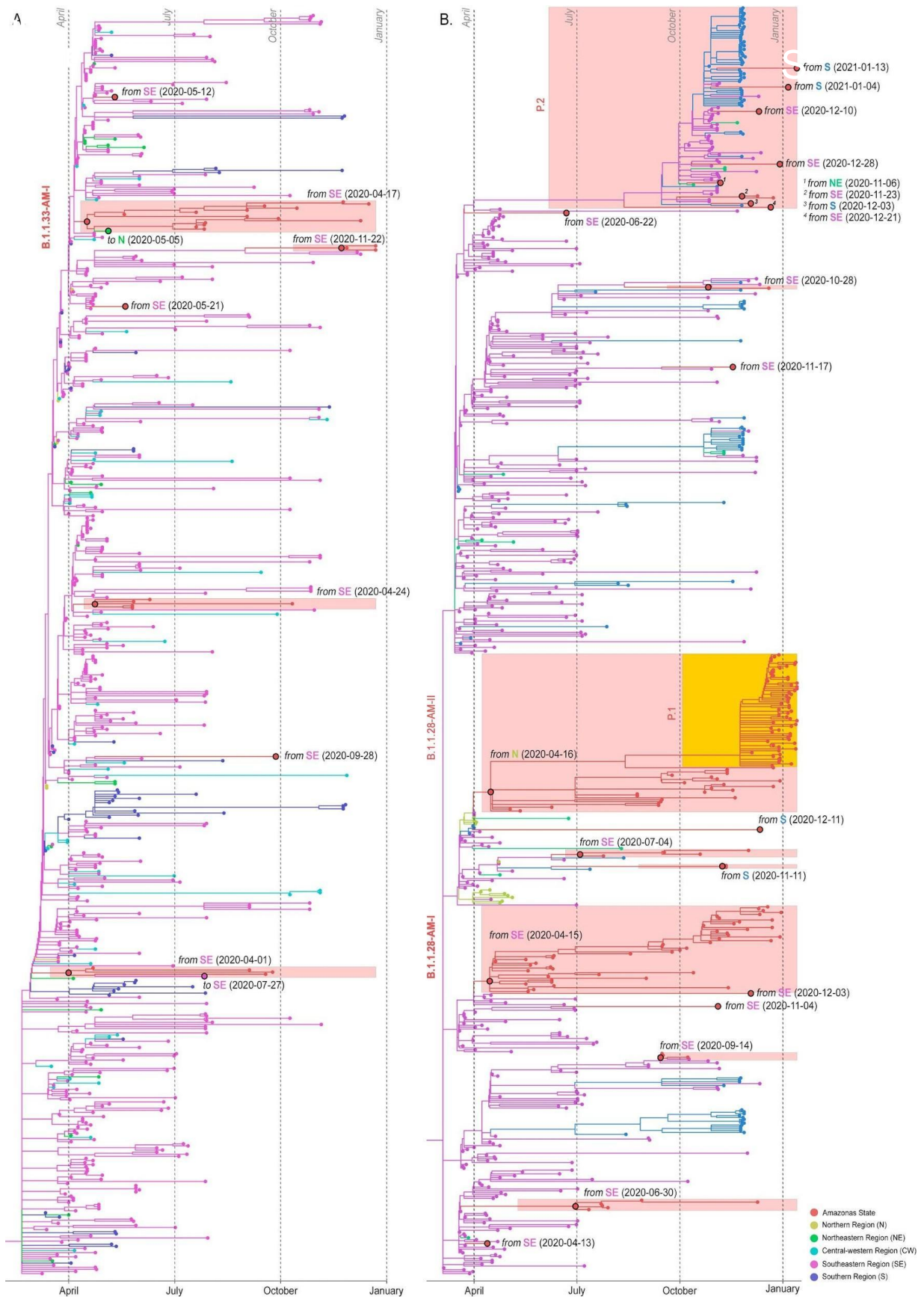
Extended data is available for this paper at <https://doi.org/10.1038/s41591-021-01378-7>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41591-021-01378-7>.

Correspondence and requests for materials should be addressed to F.G.N. or G.B.

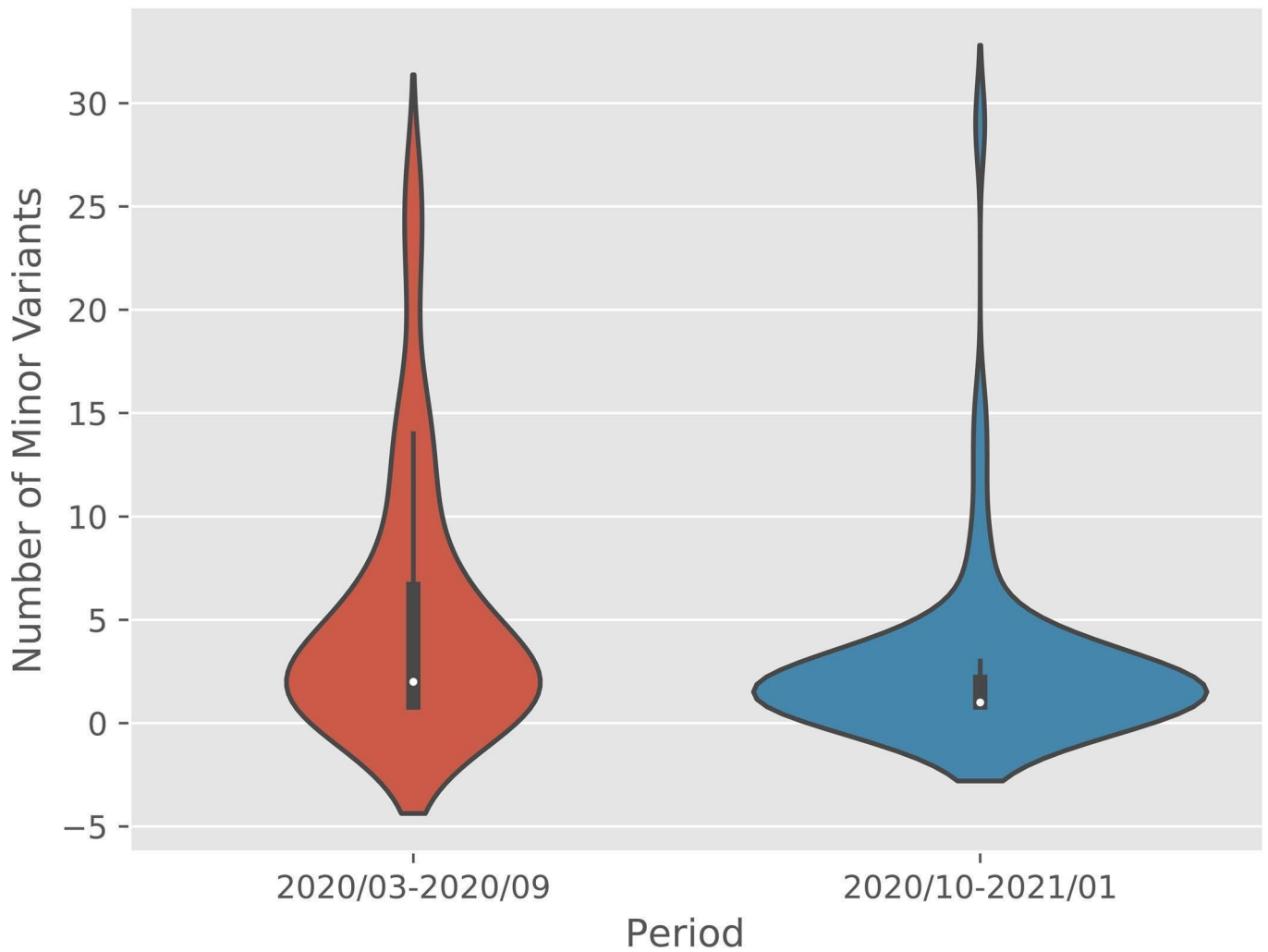
Peer review information *Nature Medicine* thanks Richard Neher, Xiang Ji and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Alison Farrell was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Maximum Likelihood (ML) phylogeographic analysis. ML phylogeographic analysis of lineages B.1.1.28/P.2 ($n = 674$) (**a**) and B.1.1.33 ($n = 602$) (**b**) in Brazil. Ancestral character state reconstruction was done in PastML with time-scaled trees. Importation and exportation events deduced from location state changes toward ($n = 28$) and from ($n = 2$) the Amazon state are detailed. The singleton Amazonian sequences and the MRCA of Amazonian clusters are indicated by black outlined circular shapes. Shaded boxes indicate the Amazonian clusters and the sub-clades that define lineages P.1 and P.2. The origin or destination of the event is indicated alongside the estimated date (Amazonian clusters) or sampling date (singletons). All locations are colored according to the legend in the bottom right.



Extended Data Fig. 2 | Minor variants density representation. Violin plot showing the density of samples with varying degrees of minor variants in two different sampling periods (reddish March-September 2020 - n = 67 biologically independent samples and blueish October 2020 to January 2021 - n = 60 biologically independent samples). Data are presented as mean values (white dots into violin plots) +/- standard deviation (SD based on values between first and third quartile, black boxplot) and adjacent values (Upper and Lower) present with vertical black lines.