ELSEVIER

Genomic monitoring of the SARS-CoV-2 B1.1.7 (WHO VOC Alpha) in the Sao Paulo state, Brazil

Svetoslav Nanev Slavov [a,b,*], Rafael dos Santos Bezerra [a], Evandra Strazza Rodrigues [a], Elaine Vieira Santos [a], Joseane Serrano Borges [a], Débora Glenda Lima de la Roque [a], José Salvatore Leister Patané [b], Alex Ranieri Jerônimo Lima [b], Gabriela Ribeiro [b], Vincent Louis Viala [b], Loyze Paola Oliveira de Lima [b], Antonio Jorge Martins [b], Claudia Renata dos Santos Barros [b], Elaine Cristina Marqueze [b], Jardelina de Souza Todão Bernardino [b], Debora Botequio Moretti [b], Ricardo Augusto Brassaloti [c], Raquel de Lello Rocha Campos Cassano [c], Pilar Drummond Sampaio Corrêa Mariani [d], João Paulo Kitajima [e], Bibiana Santos [e], Patricia Akemi Assato [f], Felipe Allan da Silva da Costa [f], Mirele Daiana Poleti [g], Jessika Cristina Chagas Lesbon [g], Elisangela Chicaroni Mattos [g], Cecilia Artico Banho [h], Lívia Sacchetto [h], Marília Mazzi Moraes [h], Rejane Maria Tommasini Grotto [f,i], Jayme A. Souza-Neto [f], Marta Giovanetti [j,k], Luiz Carlos Junior de Alcantara [j,k], Maurício Lacerda Nogueira [h], Heidge Fukumasu [g], Luiz Lehmann Coutinho [d], Rodrigo Tocantins Calado [a,l], Raul Machado Neto [b], Dimas Tadeu Covas [a,b,l], Sandra Coccuzzo Sampaio [b], Maria Carolina Elias [b], Simone Kashima [a]

[a] University of São Paulo, Ribeirão Preto Medical School, Blood Center of Ribeirão Preto, Ribeirão Preto, SP, Brazil
[b] Butantan Institute, São Paulo, Brazil
[c] University of São Paulo, Centro de Genômica Funcional da ESALQ, Piracicaba, SP, Brazil
[d] NGS Soluções Genômicas, Piracicaba, SP, Brazil
[e] Mendelics Análise Genômica SA, Sao Paulo, Brazil
[f] São Paulo State University (UNESP), School of Agricultural Sciences, Department of Bioprocesses and Biotechnology, Botucatu, Brazil
[g] Department of Veterinary Medicine, School of Animal Science and Food Engineering, University of Sao Paulo, Pirassununga, Sao Paulo, Brazil
[h] Medicine School of São José do Rio Preto (FAMERP), São José do Rio Preto, São Paulo 15090-000, Brazil
[i] Molecular Biology Laboratory, Applied Biotechnology Laboratory, Clinical Hospital of the Botucatu Medical School, Brazil
[j] Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
[k] Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil
[l] University of São Paulo, Ribeirão Preto Medical School, Ribeirão Preto, SP, Brazil

ARTICLE INFO

ABSTRACT

The SARS-CoV-2 alpha VOC (also known as lineage B.1.1.7) initially described in the autumn, 2020 in UK, rapidly became the dominant lineage across much of Europe. Despite multiple studies reporting molecular evidence suggestive of its circulation in Brazil, much is still unknown about its genomic diversity in the state of São Paulo, the main Brazilian economic and transportation hub. To get more insight regarding its transmission dynamics into the State we performed phylogenetic analysis on all alpha VOC strains obtained between February and August 2021 from the *Sao Paulo state Network for Pandemic Alert of Emerging SARS-CoV-2 variants*. The performed phylogenetic analysis showed that most of the alpha VOC genomes were interspersed with viral strains sampled from different Brazilian states and other countries suggesting that multiple independent Alpha VOC introductions from Brazil and overseas have occurred in the São Paulo State over time. Nevertheless, large monophyletic clusters were also observed especially from the Central-West part of the São Paulo State (the city of

Bauru) and the metropolitan region of the São Paulo city. Our results highlight the Alpha VOC molecular epidemiology in the São Paulo state and reinforce the need for continued genomic surveillance strategies for the real-time monitoring of potential emerging SARS-CoV-2 variants during the ever-growing vaccination process.

1. Introduction

The SARS-CoV-2 pandemic has been a source for the rapid emergence and spread of variants of concern (VOC) with major genomic changes (Davies et al., 2021). Those VOCs appear to have the ability to substitute the circulating ancestral strains, exerting thus enormous pressure on the healthcare systems as reported worldwide (Faria et al., 2021; Choudhary et al., 2021). SARS-CoV-2 variants are typically characterized by constellations of specific mutations which can enhance their transmissibility, immunological escape and phenotypic characteristics (Volz et al., 2021). The SARS-CoV-2 Alpha VOC (B.1.1.7/GR/501Y.V1/20I) which was initially reported in late 2020 in the United Kingdom, rapidly spread across the whole country between November 2020 and January 2021 (Volz et al., 2021) and later in continental Europe (Gaymard et al., 2021; Loconsole et al., 2021) and the USA (Washington et al., 2021). This is a result of the increased transmissibility of the Alpha VOC which is up to 90% more transmissible compared to the predecessor lineages (Davies et al., 2021).

Brazil is currently a world leader in the number of confirmed SARS-CoV-2 cases (21,849,137 confirmed cases by November, 5th, 2021, https://covid.saude.gov.br/) and represents a complex mixture of circulating variants. The first Alpha VOC cases in Brazil were accidentally identified in December 2020 due to an uncommon PCR amplification pattern observed during routine SARS-CoV-2 diagnosis in the city of São Paulo (Claro et al., 2021). The Alpha VOC molecular epidemiology in Brazil is still little examined as the only study which traces the molecular Alpha VOC profile analyzes only 25 complete genomes which were obtained from different Brazilian Federal States (Moreira et al., 2021). Currently, São Paulo State in Brazil demonstrates the highest number of confirmed SARS-CoV-2 cases (4409,404 by November, 5th, 2021; https://www.saopaulo.sp.gov.br/coronavirus/). Due to the high number of SARS-CoV-2, *Sao Paulo State Network for Pandemic Alert of Emerging SARS-CoV-2 Variants* was implemented with the objective to monitor the VOCs and VOIs in the State. This raised important questions regarding the dissemination and molecular epidemiology of the Alpha VOC in this Brazilian region. From a total number of 21,700 SARS-CoV-2 genomes sequenced by the *São Paulo Network for Pandemic Alert of Emerging SARS-CoV-2 Variants* and which passed quality control, 1.9% has been classified as alpha VOC but no systematic research regarding this VOC has been performed in the country so far. In this report, we analyzed the Alpha VOC phylogenetics using 366 strains obtained from different Health Districts of the São Paulo State during the genomic surveillance performed by the *Sao Paulo State Network for Pandemic Alert of Emerging SARS-CoV-2 Variants*.

2. Materials and methods

2.1. Clinical samples

From February to August 2021 the *Network for Pandemic Alert of*

Fig. 1. Division of the São Paulo State in Regional Health Departments. The São Paulo State is divided in 17 Regional Health Departments which are identified by roman numbers.

Emerging SARS-CoV-2 Variants generated 366 full genome sequences which were classified as the Alpha VOC using the PANGO lineages (https://cov-lineages.org/lineages.html). These sequences were obtained from all 17 Regional Health Divisions of the São Paulo State (Fig. 1) as follows: Araraquara (13 sequences), Baixada Santista (6 sequences), Barretos (5 sequences), Bauru (79 sequences), Campinas (64 sequences), Franca (16 sequences), Metropolitan São Paulo (60 sequences), Marília (17 sequences), Piracicaba (21 sequences), Presidente Prudente (11 sequence), São Jose do Rio Preto (9 sequences), São Joao da Boa Vista (11 sequences), Sorocaba (10 sequences), Taubaté (22 sequences), Ribeirão Preto (11 sequences) and Araçatuba (11 sequences). The sample representativity from the total number of positive cases per each Regional Health Division varied between 0.2% (10th epidemiological week) to 6.4% (26th epidemiological week). From the beginning of the 20th epidemiological week, the sampling significance was superior of 82% (G*Power, Statistical Power Analysis software). SARS-CoV-2 genomic libraries were generated using the COVIDSeq kit (Illumina, San Diego, CA) following the manufacturer's specifications. The normalized sample libraries were sequenced using Illumina MiSeq platform using the MiSeq Reagent v.2 Kit (2 × 300 cycles) (Illumina, San Diego, CA, USA).

2.2. Bioinformatic pipeline

The raw sequence data obtained were submitted to quality control analysis using the FastaQC (Andrews, 2010) software version 0.11.8. Trimming was performed using Trimmomatic (Bolger et al., 2014) version 0.3.9 in order to select the sequences with best quality. Sequences with quality scores> 30 were only used. We mapped the trimmed sequences against the SARS-CoV-2 reference (Genbank refseq NC_045512.2) using BWA (Li, 2013) software and samtools (Li et al., 2013) for read indexing. The mapped files were submitted to refinement with the software Pilon (Walker et al., 2014) in order to obtain the indels and insertions in the most correct way possible, afterwards the trimmed sequences were subjected to a remap against the genome refined by Pilon. Finally, we use bcftools (Narasimhan et al., 2016) for variant calling, and seqtk (Shen et al., 2016) for creation of consensus genomes. Positions covered by fewer than 10 reads (DP<10) and bases of quality lower than 30 were considered a gap in coverage and converted to Ns. Coverage values for each genome were calculated using samtools v1.12 (Li et al., 2009). We assessed genome consensus sequence quality using Nextclade v0.8.1 (https://clades.nextstrain.org).

2.3. Phylogenetic analysis

The sequences which were generated in this study were initially classified as belonging to the Alpha VOC using the Pangolin 3.1.11 and pangoLEARN 1.2.56 software (O'Toole et al., 2021). A set of representative Alpha VOC genomes (and associated metadata) across countries, dates, and lineages (curated weekly) was downloaded from GISAID (Shu&McCauley, 2017), including all Brazilian samples for a set of 1103 genomes (247 genomes obtained from Brazil, 490 world genomes and 366 from our study). Only genomes >29,000 bp and <1% of ambiguities were retrieved, low-quality genomes (> 10% of ambiguous positions) were excluded. Sequence alignment was performed using MAFFT v7.453 (Katoh &Standley, 2013) with subsequent realignment of misaligned stretches done in Aliview (Larsson, 2014). The beginning and end of the alignment were removed to avoid these regions of lower coverage (Larsson, 2014). Maximum Likelihood (ML) trees were estimated using IQ-TREE v1.6.12 (Nguyen et al., 2015) applying the maximum likelihood algorithm with statistical support of ultrafast bootstrap with 1000 replicates. We inferred time-scaled trees and rooted these with least-squares criteria and the evolutionary rate of $>1.1 \times 10^{-3}$ substitutions/site/year estimated by S. Duchene et al., 2020 using the TreeTime v.1 software (Sagulenko et al., 2018).

3. Results

The inferred phylogenetic history of the Alpha VOC circulating in the state of São Paulo using a number of representative reference strains obtained from Brazil and worldwide showed that our isolates were randomly interspersed with other Alpha VOCs and grouped predominantly with Alpha VOC strains circulating in Brazil and Europe (Fig. 1). Additionally, we observed several large monophyletic clusters composed of Alpha VOC strains obtained from different regions of the state including the metropolitan region of São Paulo and especially Bauru. The largest observed cluster belonged to sequences obtained from the Central-West part of the state of São Paulo and principally from the city of Bauru. Interestingly from this location we obtained the largest number of strains (total number of 79 Alpha VOC, 21.6%) which may be related to sustained Alpha VOC transmission and dissemination in this region. Another 60 genomes (16.4%), which were sequenced from the metropolitan region of the state of São Paulo formed also several smaller monophyletic clusters which might be also related to a number of independent introductions of this VOC in this location which is additionally justified by the importance of the city of São Paulo as a national and international travel, economic and demographic hub (Fig. 2).

4. Discussion

The performed study uses considerable number of complete Alpha VOC genomes obtained from the state of São Paulo in order to give detailed information on the dissemination and molecular epidemiology of this important VOC in Southeast Brazil. All of the examined sequences were obtained between February-August 2021 and this roughly corresponded to the second SARS-CoV-2 wave which hit the São Paulo State between December and February 2021. The second SARS-CoV-2 pandemic wave was further related with implementation of restriction measures and lockdowns along the State municipalities and coincided with the wide dissemination of the gamma VOC (P.1) nationwide.

Although in the reconstructed phylogenetic tree the sequences analyzed in this study were randomly interspersed with other Alpha VOC strains, several large monophyletic clusters composed of sequences obtained from different geographic regions of the state were observed (observe boxes Fig. 2). Such a local monophyletic clustering might be related to multiple introductions of the Alpha VOC in specific areas of the São Paulo State. In addition to the multiple introductions, we found phylogenetic evidence for a local sustained transmission of this VOC in several geographic regions of the State and especially the Central-West part (city of Bauru) and metropolitan areas of the city of São Paulo. Interestingly the largest cluster containing the largest number of sequences belonged to the city of Bauru, which is located ~ 340 km from the city of São Paulo. We suggest that the introduction of Alpha VOC in this region led to sustained transmission and wide dissemination of this VOC, the detection of which was only possible due to the massive sequencing provided by the *Butantan Network for Pandemic Alert of SARS-CoV-2 variants*. Higher number of Alpha VOC sequences were also obtained from the metropolitan area of the city of São Paulo, but these were interspersed in several smaller clusters along the phylogenetic tree. Such a clustering might be related to multiple Alpha VOC introductions in the city of São Paulo which is justified with the national importance of this metropolis. São Paulo city is served by the largest international airport of the country and is the most important economical, port (port of Santos) and transport center of South America. Given the intense economic, transportation and communication characteristics it is possible that different Alpha VOC introductions have occurred along time which led to the sustained Alpha VOC transmission in São Paulo city. This finding is similar to other studies which show the contribution of the large urban and densely populated areas as principal hubs for introduction, diversification and dissemination of SARS-CoV-2 (Alteri et al., 2021; Kraemer et al., 2021).

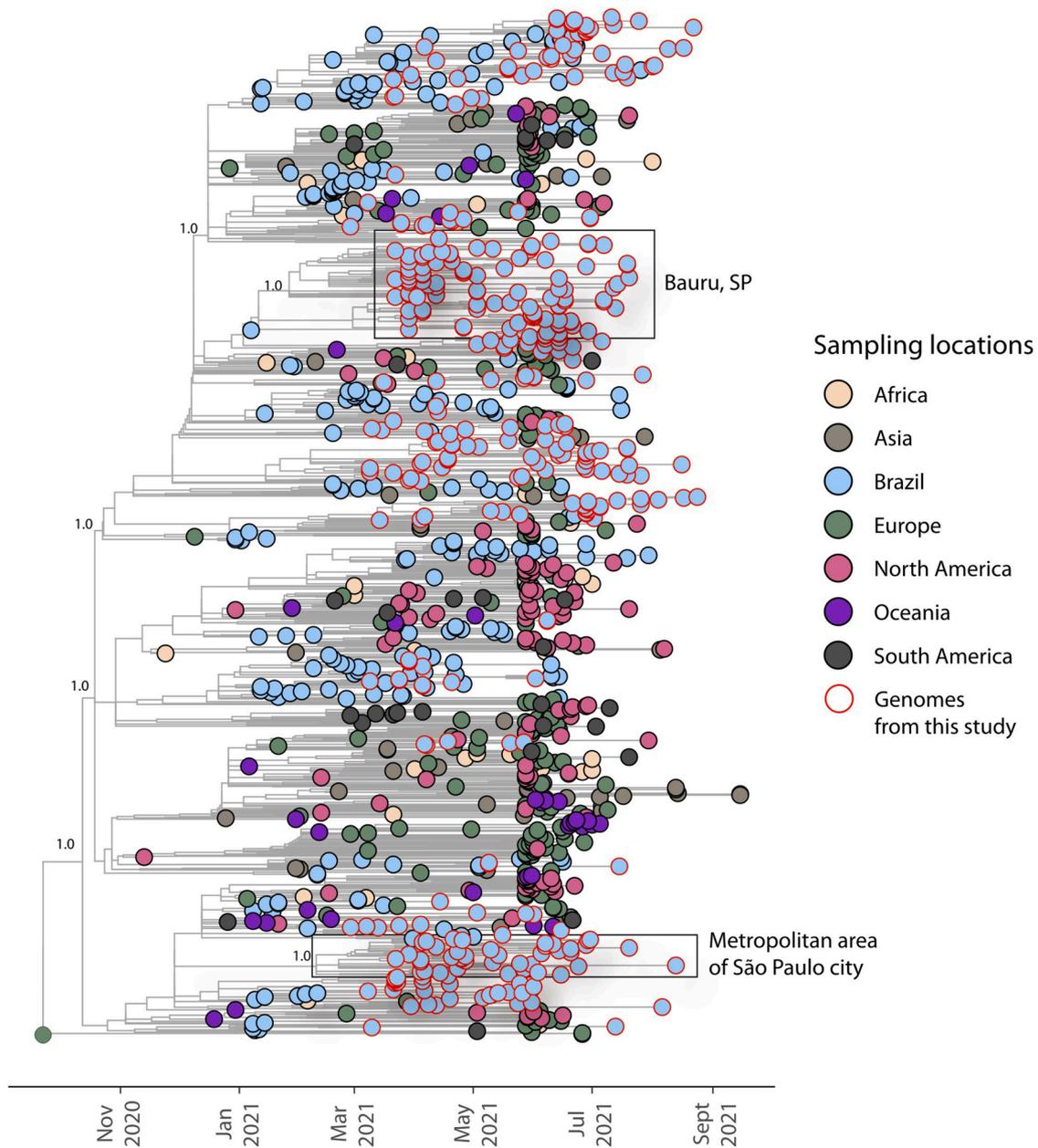Despite that the alpha VOC strains in this study were obtained from

Fig. 2. Genomic characterization of Alpha SARS-CoV-2 variants of concern (VOC) circulating in São Paulo state, Brazil. Approximate maximum likelihood (ML) phylogenetic time tree, including the 366 newly sequenced Alpha VOC SARS-CoV-2 isolates obtained in this study as well as all 247 Alpha VOC strains obtained in Brazil and 490 representative global Alpha VOC strains obtained from GISAID (https://www.gisaid.org) until October 2021. The sequences obtained in this study are presented by light blue dots with red outlining. The different clusters which were formed by the sequences obtained in our study is suggestive of multiple introductions of this VOC in the São Paulo State. In large boxes on the phylogenetic tree we also represented the largest monophyletic clusters suggesting sustained transmission of the Alpha VOC in the cities of Bauru and São Paulo.

districts that encompass the whole State territory, the proportion of the alpha VOC lineages was not prominent especially compared to Europe (Davis et al., 2021). We could not explain why the Alpha VOC in Brazil did not reach epidemic proportions. We believe that probably, the most important factor was related to a background of a scenario of a fully dominant P.1 VOC in Brazil. By the time when Alpha VOC was identified for the first time in São Paulo, Brazil has been already experiencing the second gamma VOC wave and a significant part of the inhabitants were already SARS-CoV-2 immunologically competent which may have not created a favorable environment for Alpha VOC dissemination. Additionally, the substitution of a given SARS-CoV-2 lineage is a complex process which depends on a series of factors including high incidence of other VOC infections, previous SARS-CoV-2 seroprevalence, vaccination

rates and human mobility/regional transportation characteristics (Kraemer et al., 2021).

The dissemination profile of the alpha VOC in the São Paulo State was obtained through extensive SARS-CoV-2 molecular surveillance program, the Network for Pandemic Alert of Emerging SARS-CoV-2 Variants. This effort was fundamental for a more profound understanding of VOC and VOI circulation in the State and monitoring the emergence and distribution of novel or underrepresented SARS-CoV-2 variants which also included the alpha VOC. Therefore, SARS-CoV-2 genomic surveillance remains the most important tool which predicts the spread of SARS-CoV-2 VOCs and is a fundamental approach for pandemic response. Moreover, the unification of efforts towards SARS-CoV-2 molecular surveillance will be imperative to understand the

presence and selection of variants in an ever-growing scenario of vaccine application and coverage.

Nucleotide sequence accession numbers

All of the generated sequences were deposited in GISAID (www.gisaid.org). All of the identification numbers are given in a Supplementary File 1.

Ethics approval

Credit statement

ERS, EVS, ME, LPOL, RAB, PDSCM, RLRCC, JPK, DS, PAA, FASC, MDP, JCCL, ECM, CAB, and LS designed and performed the experiments. RBS, JSLP, VLV, VF, SNS, MCE, SCS, SK, ARJL, GR analyzed the data and wrote the article. MB, PMMG, NNF, RTC, and DTC evaluated the clinical/epidemiological data and reviewed the article. GRM, AJM, CRSB and ECM evaluated the epidemiological survey. MG, LCJA, LLC, RMTG, JASN, HF, MLN, MCE, and SNS edited and reviewed the article. MCE, SCS, SK, and DTC supervised this work. All the authors agreed on the submission of the final manuscript.

Funding

Supplementary material

Supplementary table 3

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejps.2020.105216.

References

Alteri, C., Cento, V., Piralla, A., Costabile, V., Tallarita, M., Colagrossi, L., Renica, S., Giardina, F., Novazzi, F., Gaiarsa, S., Matarazzo, E., Antonello, M., Vismara, C., Fumagalli, R., Epis, O.M., Puoti, M., Perno, C.F., Baldanti, F., 2021. Genomic epidemiology of SARS-CoV-2 reveals multiple lineages and early spread of SARS-CoV-2 infections in Lombardy, Italy. Nat Commun. 12 (1), 434.

Andrews, S., Bielejec, F., Baele, G., Vrancken, B., Suchard, M.A., Rambaut, A., Lemey, P., 2010. FastQC: a quality control tool for high throughput sequence data. SpreaD3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes Mol. Biol. Evol. 33 (8), 2167–2169.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30 (15), 2114–2120.

I. Choudhary, O.P., Priyanka, Singh, Rodriguez-Morales, A.J., 2021. Second wave of COVID-19 in India: dissection of the causes and lessons learnt Travel Med. Infect. Dis., 102126

Claro, I.M., da Silva Sales, F.C., Ramundo, M.S., Candido, D.S., Silva, C.A.M., de Jesus, J. G., Manuli, E.R., de Oliveira, C.M., Scarpelli, L., Campana, G., Pybus, O.G., Sabino, E. C., Faria, N.R., Levi, J.E., 2021. Local Transmission of SARS-CoV-2 Lineage B.1.1.7, Brazil. December 2020 Emerg. Infect. Dis. 27 (3), 970–972.

CMMID COVID-19 Working Group, COVID-19 Genomics UK (COG-UK) Consortium Davies, N.G., Abbott, S., Barnard, R.C., Jarvis, C.I., Kucharski, A.J., Munday, J.D., Pearson, C.A.B., Russell, T.W., Tully, D.C., Washburne, A.D., Wenseleers, T., Gimma, A., Waites, W., Wong, K.L.M., van Zandvoort, K., Silverman, J.D., Diaz-Ordaz, K., Keogh, R., Eggo, R.M., Funk, S., Jit, M., Atkins, K.E., Edmunds, W.J., 2021. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science 372 (6538), eabg3055.

Duchene, S., Featherstone, L., Haritopoulou-Sinanidou, M., Rambaut, A., Lemey, P., Baele, G., 2020. Temporal signal and the phylodynamic threshold of SARS-CoV-2. Virus Evol 6 (2), veaa061.

Faria, N.R., Mellan, T.A., Whittaker, C., Claro, I.M., Candido, D.D.S., Mishra, S., Crispim, M.A.E., Sales, F.C.S., Hawryluk, I., McCrone, J.T., Hulswit, R.J.G., Franco, L.A.M., Ramundo, M.S., de Jesus, J.G., Andrade, P.S., Coletti, T.M., Ferreira, G.M., Silva, C.A.M., Manuli, E.R., Pereira, R.H.M., Peixoto, P.S., Kraemer, M.U.G., Gaburo Jr., N., Camilo, C.D.C., Hoeltgebaum, H., Souza, W.M., Rocha, E.C., de Souza, L.M., de Pinho, M.C., Araujo, L.J.T., Malta, F.S.V., de Lima, A. B., Silva, J.D.P., Zauli, D.A.G., Ferreira, A.C.S., Schnekenberg, R.P., Laydon, D.J., Walker, P.G.T., Schlüter, H.M., Dos Santos, A.L.P., Vidal, M.S., Del Caro, V.S., Filho, R.M.F., Dos Santos, H.M., Aguiar, R.S., Proença-Modena, J.L., Nelson, B., Hay, J.A., Monod, M., Miscouridou, X., Coupland, H., Sonabend, R., Vollmer, M., Gandy, A., Prete Jr., C.A., Nascimento, V.H., Suchard, M.A., Bowden, T.A., Pond, S.L. K., Wu, C.H., Ratmann, O., Ferguson, N.M., Dye, C., Loman, N.J., Lemey, P., Rambaut, A., Fraiji, N.A., Carvalho, M.D.P.S.S., Pybus, O.G., Flaxman, S., Bhatt, S., Sabino, E.C., 2021. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science 372 (6544), 815–821.

ANRS MIE AC43 COVID-19, French viro COVID group Gaymard, A., Bosetti, P., Feri, A., Destras, G., Enouf, V., Andronico, A., Burrel, S., Behillil, S., Sauvage, C., Bal, A., Morfin, F., Van Der Werf, S., Josset, L., Blanquart, F., Coignard, B., Cauchemez, S., Lina, B., 2021. Early assessment of diffusion and possible expansion of SARS-CoV-2., 2021. Lineage 20I/501Y.V1 (B.1.1.7, variant of concern 202012/01) in France, January to March 2021. Euro Surveill 26 (9), 2100133.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30 (4), 772–780.

COVID-19 Genomics UK (CoG-UK) consortium Kraemer, M.U.G., Hill, V., Ruis, C., Dellicour, S., Bajaj, S., McCrone, J.T., Baele, G., Parag, K.V., Battle, A.L., Gutierrez, B., Jackson, B., Colquhoun, R., O'Toole, Á., Klein, B., Vespignani, A., Volz, E., Faria, N.R., Aanensen, D., Loman, N.J., du Plessis, L., Cauchemez, S., Rambaut, A., Scarpino, S.V., Pybus, O.G., 2021. Spatiotemporal invasion dynamics of SARS-CoV-2 lineage B.1.1.7 emergence. Science. eabj0113.

Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30 (22), 3276–3278.

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv, 1303.3997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools Bioinformatics 25 (16), 2078–2079.

Loconsole, D., Centrone, F., Morcavallo, C., Campanella, S., Sallustio, A., Accogli, M., Fortunato, F., Parisi, A., Chironna, M., 2021. Rapid Spread of the SARS-CoV-2 Variant of Concern 202012/01 in Southern Italy (December 2020-March 2021). Int. J. Environ. Res. Public. Health 18 (9), 4766.

Moreira, F.R.R., Bonfim, D.M., Zauli, D.A.G., Silva, J.P., Lima, A.B., Malta, F.S.V., Ferreira, A.C.S., Pardini, V.C., Magalhães, W.C.S., Queiroz, D.C., Souza, R.M., Geddes, V.E.V., Costa, W.C., Moreira, R.G., Faria, N.R., Voloch, C.M., Souza, R.P., Aguiar, R.S., 2021. Epidemic Spread of SARS-CoV-2 Lineage B.1.1.7 in Brazil. Viruses 13 (6), 984.

Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., Durbin, R., 2016. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics 32 (11), 1749–1751.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015 Jan;32(1):268-74. doi: 10.1093/molbev/msu300. Epub 2014 Nov 3. PMID: 25371430; PMCID: PMC4271533.

O'Toole, Á., Scher, E., Underwood, A., Jackson, B., Hill, V., McCrone, J.T., Colquhoun, R., Ruis, C., Abu-Dahab, K., Taylor, B., Yeats, C., du Plessis, L., Maloney, D., Medd, N., Attwood, S.W., Aanensen, D.M., Holmes, E.C., Pybus, O.G., Rambaut, A., 2021. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol 7 (2), veab064.

Sagulenko, P., Puller, V., Neher, R.A., 2018. TreeTime: maximum-likelihood phylodynamic analysis. Virus Evol 4 (1), vex042.

Shen, W., Le, S., Li, Y., Hu, F., 2016. SeqKit: a Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS ONE 11 (10), e0163962.

Shu, Y., McCauley, J., 2017. GISAID: global initiative on sharing all influenza data–from vision to reality. Eurosurveillance 22 (13), 30494.

COVID-19 Genomics UK (COG-UK) consortium Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg, L., Hinsley, W.R., Laydon, D.J., Dabrera, G., O'Toole, Á., Amato, R., Ragonnet-Cronin, M., Harrison, I., Jackson, B., Ariani, C.V., Boyd, O., Loman, N.J., McCrone, J.T., Gonçalves, S., Jorgensen, D., Myers, R., Hill, V.,

Jackson, D.K., Gaythorpe, K., Groves, N., Sillitoe, J., Kwiatkowski, D.P., Flaxman, S., Ratmann, O., Bhatt, S., Hopkins, S., Gandy, A., Rambaut, A., Ferguson, N.M., 2021. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. Nature 593 (7858), 266–269.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE 9 (11), e112963.

Washington, N.L., Gangavarapu, K., Zeller, M., Bolze, A., Cirulli, E.T., Schiabor Barrett, K.M., Larsen, B.B., Anderson, C., White, S., Cassens, T., Jacobs, S., Levan, G., Nguyen, J., Ramirez 3rd., J.M., Rivera-Garcia, C., Sandoval, E., Wang, X., Wong, D., Spencer, E., Robles-Sikisaka, R., Kurzban, E., Hughes, L.D., Deng, X., Wang, C., Servellita, V., Valentine, H., De Hoff, P., Seaver, P., Sathe, S., Gietzen, K., Sickler, B., Antico, J., Hoon, K., Liu, J., Harding, A., Bakhtar, O., Basler, T., Austin, B., MacCannell, D., Isaksson, M., Febbo, P.G., Becker, D., Laurent, M., McDonald, E., Yeo, G.W., Knight, R., Laurent, L.C., de Feo, E., Worobey, M., Chiu, C.Y., Suchard, M. A., Lu, J.T., Lee, W., Andersen, K.G., 2021. Emergence and rapid transmission of SARS-CoV-2 B1.1.7 in the United States. Cell 184 (10), 2587–2594.e7.