

¹Universidade de São Paulo, Instituto de Medicina Tropical de São Paulo, São Paulo, São Paulo, Brazil

²Instituto Adolfo Lutz, Laboratório Estratégico, São Paulo, São Paulo, Brazil

³University of Oxford, Department of Zoology, Oxford, United Kingdom

⁴Instituto Adolfo Lutz, Centro de Virologia, São Paulo, São Paulo, Brazil

⁵Universidade Federal do Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, Brazil

⁶University of Birmingham, Birmingham, United Kingdom

⁷Ministério da Saúde, Secretaria de Vigilância em Saúde, Coordenação Geral de Laboratórios de Saúde Pública, Brasília, DF, Brazil

⁸Universidade Federal da Grande Dourados, Laboratório de Pesquisa em Ciências da Saúde, Dourados, Mato Grosso do Sul, Brazil

⁹Fundação Osvaldo Cruz Campo Grande, Mato Grosso do Sul, Brazil

¹⁰University of Edinburgh, Institute of Evolutionary Biology, Edinburgh, United Kingdom

¹¹Imperial College, School of Public Health, Department of Infectious Disease Epidemiology, London, United Kingdom

*This authors contributed equally

Correspondence to: Nuno Rodrigues Faria
University of Oxford, Department of Zoology, South Parks Rd, Oxford OX1 3SY, United Kingdom

E-mail: nuno.faria@zoo.ox.ac.uk

Received: 21 April 2020

Accepted: 22 April 2020

Importation and early local transmission of COVID-19 in Brazil, 2020

Jaqueline Goes de Jesus^{1*}, Claudio Sacchi^{2*}, Darlan da Silva Candido^{3*}, Ingra Morales Claro^{1*}, Flávia Cristina Silva Sales^{1*}, Erika Regina Manuli^{1*}, Daniela Bernardes Borges da Silva⁴, Terezinha Maria de Paiva⁴, Margarete Aparecida Benega Pinho⁴, Katia Correa de Oliveira Santos⁴, Sarah Catherine Hill³, Renato Santana Aguiar⁵, Filipe Romero⁵, Fabiana Cristina Pereira dos Santos⁴, Claudia Regina Gonçalves², Maria do Carmo Timenetsky⁴, Joshua Quick⁶, Julio Henrique Rosa Croda^{7,8,9}, Wanderson de Oliveira⁷, Andrew Rambaut¹⁰, Oliver G. Pybus³, Nicholas J. Loman⁶, Ester Cerdeira Sabino^{1*}, Nuno Rodrigues Faria^{1,3,11*}

ABSTRACT

We conducted the genome sequencing and analysis of the first confirmed COVID-19 infections in Brazil. Rapid sequencing coupled with phylogenetic analyses in the context of travel history corroborate multiple independent importations from Italy and local spread during the initial stage of COVID-19 transmission in Brazil.

KEYWORDS: Public health surveillance. COVID-19. SARS-CoV-2.

INTRODUCTION

The severe acute respiratory syndrome 2 (SARS-CoV-2) was first identified in Wuhan, Central China, in early December 2019, and reported to the World Health Organization (WHO) country office in China on December 31, 2019¹. SARS-CoV-2 infection causes coronavirus-associated acute respiratory disease in humans, a disease named corona virus disease 19 (COVID-19)². COVID-19 is the third documented spill over of a coronavirus from an animal reservoir to humans in the last two decades to have caused a serious public health threat³.

On January 30, 2020, COVID-19 was declared a Public Health Emergency of International Concern. On March 16, 2020, WHO reported 153,517 confirmed cases across the globe, and 5,735 confirmed deaths in 143 countries, territories or areas. Within Latin America, Brazil is the country with the largest number of confirmed cases. The country has reported 234 cases across 15 federal States; Sao Paulo, Rio de Janeiro and Bahia States have confirmed local transmission⁴.

During the early stages of an epidemic disease, molecular surveillance can inform on the tracking and control of the virus spread across the global and at local scales. Moreover, viral genomes can help to design effective molecular diagnostics, improve vaccine design and complement the contact tracking^{5,6}. However, the resolution of the transmission networks reconstructed from genetic data will depend on the rate at which genetic changes accumulate across viral genomes. Within outbreaks, short timescales mean that not all the observed changes will become fixed at the population level⁷. To investigate the early transmission dynamics of imported and local cases in

Brazil, we set up a genomic observatory in Sao Paulo where we sequenced and analysed two complete SARS-CoV-2 genomes in less than 48 h after the cases confirmation. Here we investigate the transmission patterns from phylogenetic analysis of the earliest six SARS-CoV-2 cases in Brazil.

MATERIALS AND METHODS

Samples from suspected SARS-CoV-2 cases underwent confirmatory diagnostic real-time RT-PCR testing⁸ at the Instituto Adolfo Lutz (IAL), the regional reference laboratory for SARS-CoV-2 detection in Sao Paulo State, Southeast Brazil. Samples obtained from the Reference Centre for Arbovirus of Sao Paulo, Adolfo Lutz Institute (IAL) have been processed in agreement with routine surveillance activities from the Brazilian Ministry of Health.

We used the open COVID-19 sequencing available and the bioinformatics protocols developed by the ARTIC network. Sequencing protocols, multiplex PCR primers, and bioinformatic pipelines are described in detail at <https://artic.network/ncov-2019>. In brief, cDNA synthesis was conducted in duplicate for each sample and the concentration of PCR products was measured using a Qubit dsDNA High Sensitivity kit on a Qubit 3.0 fluorometer (Thermo Fisher Scientific, Waltham, USA). Library preparation was conducted without a barcoding step and libraries were sequenced on an R9.4.1 flow cell using MinKNOW version 19.10.1 (Oxford Nanopore Technologies, Oxford, UK) for over 12 h. The open-source software RAMPART version 10.5 was used to assign and map reads in real-time. Raw files were base-called with Guppy, demultiplexed and trimmed with Porechop (<https://github.com/rwick/Porechop>) and mapped against reference sequence Wuhan-Hu-1 (GenBank Accession Number MN908947). Variants were called using nanopolish 0.11.3. Low coverage regions were masked with N characters. Coverage for the SPBR1 and SPBR2 was 96.9 and 99.6%, with 552730 and 3461754 mapped reads, respectively (Table S1). Raw read data for SPBR1 is available for inspection from <https://cadde.s3.climb.ac.uk/covid-19/BR1.sorted.bam>.

We investigated the transmission dynamics of the early COVID-19 cases in Brazil (SPBR1 to SPBR6) by analysing genetic changes among the early genomes from Brazil belonging to imported and local cases, and by estimating a maximum likelihood phylogenetic tree together with a set of global reference sequences. We added the SBPBR1 and SPBR2 consensus sequences from Sao Paulo to a curated dataset of complete genomes available from GISAID that included four additional sequences from Brazil (available on GISAID of 15th March 2020). A multiple sequence alignment comprising 347 complete genomes from several

countries was generated using MAFFT⁹ and manually edited. A maximum likelihood (ML) phylogenetic tree was estimated using PhyML version 3.0¹⁰ using a Hasegawa-Kishino-Yano nucleotide substitution model with a gamma-distributed rate variation across sites.

RESULTS

Four of the six patients self-reported travelling from European countries to Sao Paulo city (SPBR1 to SPBR4) (Table S1). Two patients (SPBR5 and SPBR6) reported direct contact with SPBR1 and no travel outside Brazil. Patient SPBR1 (60-65 year old male) self-reported arriving from Italy on the February 21, 2020 he started symptoms on the February 24 and tested positive two days later. Patients SPBR5 and SPBR6 were in direct contact with patient SPBR1 on February 22, and tested positive on February 29, 2020. Figure 1 shows the clade containing Brazilian sequences along with location of infection (squares) and reporting (circles). Our analyses show that the SPBR1 genome is identical to the SPBR5 and SPBR6 contacts (illustrated by zero branch lengths in Figure 1; detailed tree with annotated tips and travel history information for the clade containing Brazilian sequences can be found in the Figure S1).

We found that the SPBR1, SPBR5 and SPBR6 are identical to several other genomes circulating in Italy and elsewhere collected between February 20 and March 2, 2020 (Figure 1). The lack of changes among SARS-CoV-2 genomes collected during this period is not surprising given the evolutionary rate of the virus that results in an average of 1 to 2 mutations per month¹¹. These data highlight the critical importance of contextualizing phylogenetic information with travel history when investigating early transmission dynamics of SARS-CoV-2. As no epidemiological information was available for SPBR5 and SPBR6, one could not exclude an alternative scenario based on sequencing data alone that would suggest additional independent introductions from Italy or elsewhere.

Patients SPBR2, SPBR3 and SPBR4 all reported travelling to Italy, where the of incidence of COVID-19 has been the highest outside Wuhan in China¹². Consistent with the travel history, sequences from these patients are found interspersed in the tree in agreement with the multiple independent introductions of SARS-CoV-2 to Sao Paulo from Italy. This finding highlights the key role of human mobility in the early stages of the pandemic and is in line with a recent analysis on the risk of importation of COVID-19 based on the history of air traveling data and the incidence data¹³. Given that the air traveling to Brazil from Italy has reduced, it is possible that the proportion

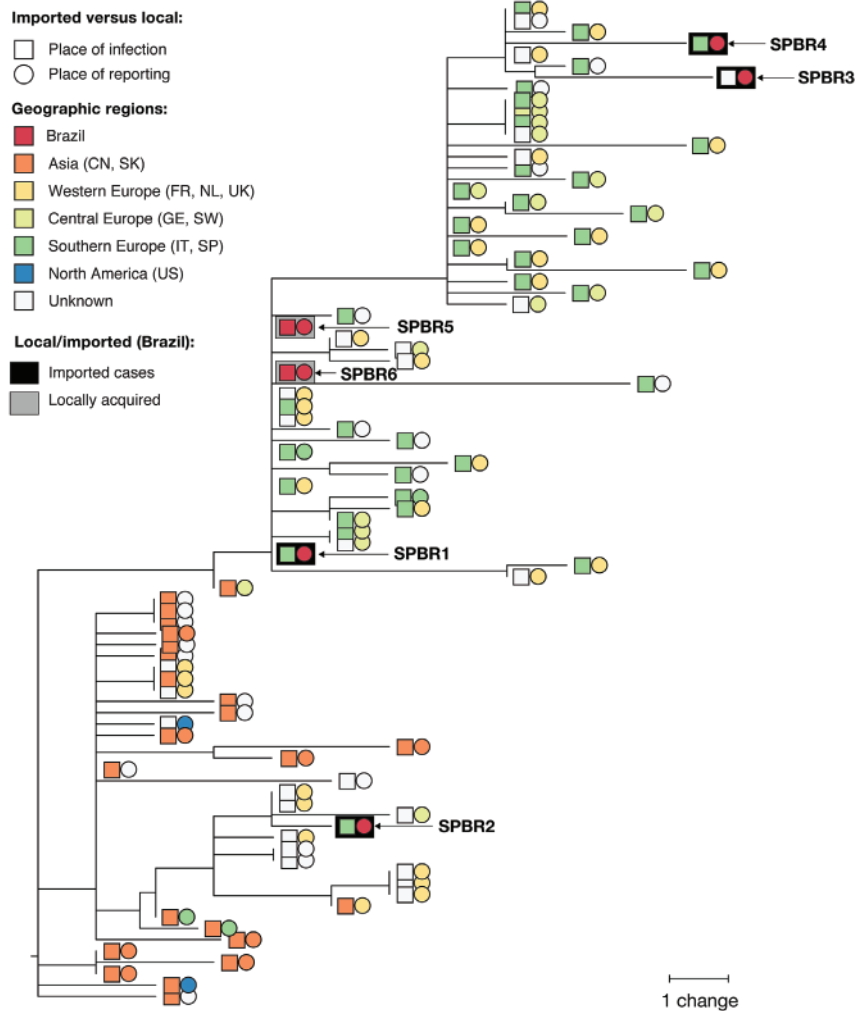


Figure 1 - Maximum likelihood phylogeny (n=88) including Brazilian SARS-CoV-2 genomes from the first confirmed cases in Brazil. Squares and circles are coloured according to the place of infection and the place of reporting, respectively. Local cases are highlighted with a grey background, imported cases are highlighted with a black background. A full tree (n=347) can be found in the [Supplementary Material \(Figure S1\)](#).

of SARS-CoV-2 imported cases from other countries, particularly the USA, may increase¹³.

DISCUSSION

Our study provides a snapshot of the early establishment of the COVID-19 pandemic in Brazil, characterized by multiple independent introductions from Italy, followed by local transmission of the virus in Sao Paulo. Phylogenetic analyses are broadly consistent with the patients' self-reported traveling histories. We show that the two genomes associated with local transmission are linked to a patient infected in Italy and are identical to other Italian genomes collected in the same time window. Given the within-outbreak rate of evolutionary change estimated for SARS-CoV-2¹¹, we caution against inferring directionality of transmission based on genetic data alone. Such inferences

can further be overshadowed by incomplete sampling due to delays, reflecting the lack of equitable access to diagnosis and genomic sequencing.

CONCLUSION

Given the findings of the present study, we conclude that phylogenetic data from the pandemic needs to be contextualized with appropriate metadata, including basic demographics, symptoms onset date, the sample collection date, the country of reporting and the self-reported travel history. Joint epidemiological and genomic surveillance of COVID-19 cases will be critical to rapidly identify possible clusters of local transmission in Brazil and in other countries, and to better understand and help mitigating the transmission in the community.

ACKNOWLEDGMENTS

We thank GISAID database for supporting rapid and open access real-time SARS-CoV-2 genomic data sharing. We would like to thank the authors who generated and shared data with the research community and all the patients involved, the staff and research groups who assisted with patient care, sample collection, genome sequencing and data sharing. We thank Josh Quick, Lucy Matkin and Julien Thezé for support.

AUTHORS' CONTRIBUTIONS

Conceptualization, Jaqueline de Jesus, Ester Sabino and Nuno Rodrigues Faria; data curation, Darlan da Silva Candido, Oliver G. Pybus and Nuno Rodrigues Faria; formal analysis, Darlan da Silva Candido, Sarah Catherine Hill, Filipe Romero, Andrew Rambaut, Nicholas James Loman and Nuno Rodrigues Faria; funding acquisition, Andrew Rambaut, Nicholas James Loman, Ester Cerdeira Sabino, Nuno Rodrigues Faria; investigation, Jaqueline de Jesus, Claudio Sacchi, Ingra Morales Claro, Flávia Cristina Silva Sales, Erika Regina Manuli, Daniela Bernardes Borges da Silva, Terezinha Maria de Paiva, Margarete Aparecida Benega, Katia Correa de Oliveira Santos, Renato Santana Aguiar, Fabiana Cristina Pereira Santos and Claudia Regina Gonçalves; methodology, Jaqueline de Jesus, Claudio Sacchi, Ingra Morales Claro, Fabiana Cristina Pereira dos Santos, Joshua Quick and Nicholas James Loman; project administration, Andrew Rambaut, Oliver G. Pybus, Nicholas James Loman, Ester Sabino and Nuno Rodrigues Faria; resources, Julio Henrique Rosa Croda and Wanderson de Oliveira, Oliver G Pybus, Andrew Rambaut, Ester Sabino and Nuno Rodrigues Faria; software, Filipe Romero, Andrew Rambaut, Nicholas J Loman; supervision, Maria do Carmo Timenetsky, Oliver G. Pybus, Ester Sabino and Nuno Rodrigues Faria; validation, Joshua Quick, Andrew Rambaut; writing – original draft, Jaqueline de Jesus, Darlan da Silva Candido and Nuno Rodrigues Faria; writing – review & editing, Andrew Rambaut, Oliver G. Pybus, Nicholas James Loman and Ester Sabino.

FUNDING

This work was supported by a Medical Research Council and FAPESP CADDE partnership award (MR/S0195/1), FAPESP grant N° 2018/14389-0, and a John Fell Research Fund (grant N° 005166). NRF is supported by a Wellcome Trust and Royal Society Sir Henry Dale Fellowship (204311/Z/16/Z). DDSC is supported by the

Clarendon Fund and by the Oxford University Zoology Department.

REFERENCES

1. World Health Organization. Critical preparedness, readiness and response actions for COVID-19. [cited 2020 Apr 22]. Available from: <https://www.who.int/publications-detail/critical-preparedness-readiness-and-response-actions-for-covid-19>
2. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol.* 2020 In Press.
3. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nat Med.* 2020;26:450-2.
4. Brasil. Ministério da Saúde. Notificação de casos de doença pelo coronavírus 2019 (COVID-19). [cited 2020 March 15]; Available from: <https://www.saude.gov.br/component/content/article/34-page/9895-coronav%C3%ADrus.html?highlight=WyJjb3JvbmF2XHUwMGVkcVzII0=&Itemid=101>
5. Park DJ, Dudas G, Wohl S, Goba A, Whitmer SL, Andersen KG, et al. Ebola virus epidemiology, transmission, and evolution during seven months in Sierra Leone. *Cell.* 2015;161:1516-26.
6. Gire SK, Goba A, Andersen KG, Sealfon RS, Park DJ, Kanneh L, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science.* 2014;345:1369-72.
7. Holmes EC, Dudas G, Rambaut A, Andersen KG. The evolution of Ebola virus: insights from the 2013-2016 epidemic. *Nature.* 2016;538:193-200.
8. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill.* 2020;25:2000045.
9. Katoh K, Standley DM. MAFFT: iterative refinement and additional methods. In: Russel DJ, editor. *Multiple sequence alignment methods.* New York: Humana; 2014. p.131-46.
10. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59:307-21.
11. Rambaut A. Phylogenetic analysis of nCoV-2019 genomes. [cited 2020 Apr 22]. Available from: <http://virological.org/t/phylogenetic-analysis-of-23-ncov-2019-genomes-2020-01-23/335>
12. World Health Organization. Coronavirus disease (COVID-2019) situation reports. [cited 2020 Apr 22]. Available from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>
13. Candido DS, Watts A, Abade L, Kraemer MU, Pybus OG, Croda J, et al. Routes for COVID-19 importation in Brazil. *J Travel Med.* 2020 In press.

SUPPLEMENTARY MATERIAL

Table S1 - Sequencing statistics for the Brazilian SARS-CoV-2 genomes from this study.

Isolate	Mapped Reads	Average depth coverage	Bases covered >10x	Bases covered >25x	Reference covered (%)
SPBR1	552730	3622.14	29426	29106	96.8966
SPBR2	3461754	5117.28	29849	29845	99.5954

We gratefully acknowledge the authors, originating and submitting laboratories of the sequences from GISAID's EpiFlu™ Database on which this research is based. The list is detailed on Supplementary Table 2, which is available on GitHub (<https://github.com/CADDE-CENTRE/REPOSITORY/blob/master/First%20genomes%20from%20Americas.docx>). All submitters of data may be contacted directly via www.gisaid.org.

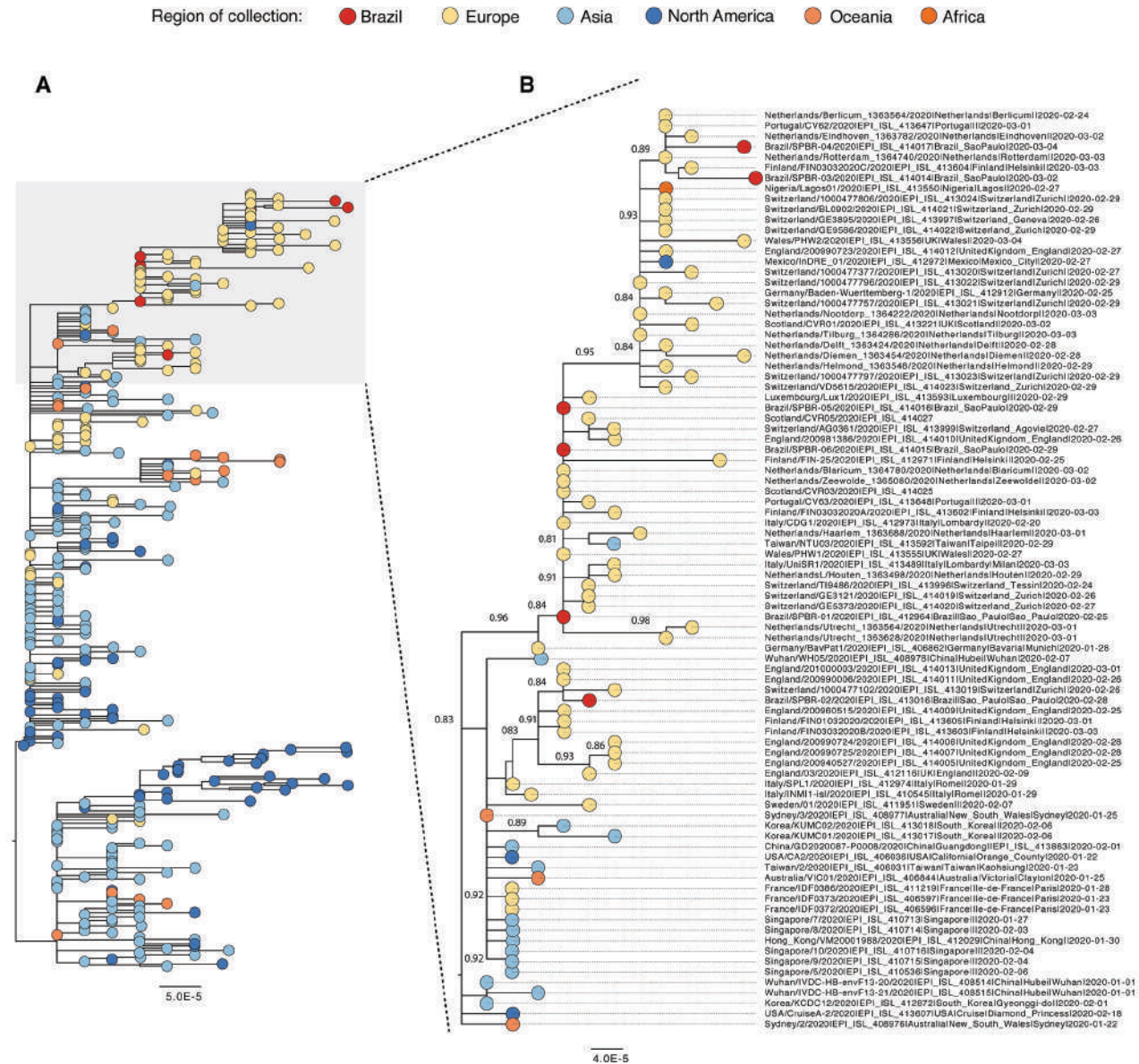


Figure S1 - SARS-CoV-2 phylogeny of the first confirmed Brazilian cases. A) Global SARS-CoV-2 Maximum Likelihood phylogeny including the six first genomes from Brazil. The tree was estimated using all available sequences in GISAID database (n=347) as of the 10th of March 2020. Tips are coloured according to the location of collection. B) Expansion of the clades containing the six Brazilian SARS-CoV-2 genomes (n=88). Tips are coloured according to the location of collection and only approximate likelihood ratio node supports > 0.80 are shown. Tips were labelled according to sequence name, GISAID accession number, place and date of collection.