Viral genetic evidence and host immune response of a small cluster of case reports with two episodes of SARS-CoV-2 infection

Running title: SARS-CoV-2 reinfection in Brazil

Authors

Natalia Fintelman-Rodrigues[1,9], Aline de Paula Dias Da Silva[1,9], Monique Cristina dos Santos[1,9], Felipe B. Saraiva[2], Marcelo Alves Ferreira[9], João Gesto[2,9], Danielle A.S. Rodrigues[3], André Macedo Vale[3], Isaclaudia Gomes de Azevedo[1], Vinícius Cardoso Soares[1], Hui Jiang[5], Hongdong Tan[5], Diogo A. Tschoeke[6], Carolina Q. Sacramento[1,2], Fernando A. Bozza[7,8], Carlos M. Morel[9], Patrícia T. Bozza[1], Thiago Moreno L. Souza[1,9]

Affiliations

1-Laboratório de Imunofarmacologia, Instituto Oswaldo Cruz (IOC), Fundação Oswaldo Cruz (Fiocruz), Rio de Janeiro, RJ, Brazil.

2-Instituto de Tecnologia em Imunobiológicos (Bio-Manguinhos), Fiocruz, Rio de Janeiro, RJ, Brazil

3-Laboratório de Biologia de Linfócitos – LBL, Instituto Biofísica, Universidade Federal do Rio de Janeiro, Rio de Janeiro - RJ – Brazil

4-Programa de Imunologia e Inflamação, Universidade Federal do Rio de Janeiro, UFRJ, Rio de Janeiro, RJ, Brazil.

5- MGI Tech Co., Ltd., Building No.11, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China

6- SAGE/COPPE, UFRJ, Rio de Janeiro, RJ, Brazil.

7-Instituto Nacional de Infectologia Evandro Chagas, (INI), Fundação Oswaldo Cruz (Fiocruz), Rio de Janeiro, RJ, Brazil.

8-D'Or Institute for Research and Education

9-National Institute for Science and Technology on Innovation in Diseases of Neglected Populations (INCT/IDPN), Center for Technological Development in Health (CDTS), Fiocruz, Rio de Janeiro, RJ, Brazil.

Corresponding author: Thiago Moreno Lopes e Souza, PhD

Fundação Oswaldo Cruz (Fiocruz)
Centro de Desenvolvimento Tecnológico em Saúde (CDTS)
Instituto Oswaldo Cruz (IOC)
Pavilhão Osório de Almeida, sala 16
Av. Brasil 4365, Manguinhos, Rio de Janeiro - RJ, Brasil, CEP 21060340
Tel.: +55 21 2562-1311
Email: tmoreno@cdts.fiocruz.br

Abstract

Background: The dynamics underlying severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) reinfection remains poorly understood. We added to the registered case reports of reinfection in USA, Belgium/Netherlands, Ecuador and Hong Kong, a small cluster of individuals with two episodes of 2019 coronavirus disease (COVID-19). Virus genomic analysis and the host immune response were used to characterize this group.

Methods: Four individuals from Rio de Janeiro, Brazil, with clinical manifestations of COVID-19 on March and again in late May of 2020 were studied. Nasopharyngeal swabs were collected for RT-PCR and viral genome sequencing (BGI-MGI-2000). Plasma samples from the acute and convalescent phases of both infection episodes were accessed to document innate and humoral responses.

Findings: After approximately 60 days of the first diagnostic episode of SARS-CoV-2 infection, the four individuals presented new clinical and molecular evidence of COVID-19. Complete SARS-CoV-2 genome sequence provided genetic evidence of reinfection. The individuals presented an enhanced innate response compared to healthy SARS-CoV-2 negative controls. Patients did not develop a neutralizing humoral immunity, possibly remaining susceptible to another episode of COVID-19. The second episode, associated with higher viral loads and clinical symptoms, likely boosted their anti-SARS-CoV-2 humoral response.

Interpretation: SARS-CoV-2 reinfection was fully documented by identification of genetically distinct virus sequences in the first and second episodes for two individuals. The quantity of SARS-CoV-2-associated genetic reads and coverage of virus genome ruled out that the initial RT-PCR results were false positive. The identification that some individuals with mild COVID-19 may have controlled SARS-CoV-2 replication without developing detectable humoral immunity, opens the possibility that reinfection may be more frequent than supposed – but weakly documented.

Funding: This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. Funding was also provided by CNPq, CAPES and FAPERJ through the National Institutes of Science and Technology Program (INCT) to Carlos Morel (INCT-IDPN).

Key-words: Reinfection, SARS-CoV-2, COVID-19, next generation sequencing, humoral response, case report

Research in Context

Evidence before this study

The possibility that SARS-CoV-2 may cause successive infections is a matter of intense debate and concern. So far, there were four documented case reports of SARS-CoV-2 reinfection, by means of partial or complete genomic evidence, among the over 55 million confirmed cases. Neither the innate nor humoral responses were followed-up in depth for these patients.

Added value of this study

We followed up a small cluster of four house-related contactants with two episodes of COVID-19, separated approximately by 60 days apart. Viral genomic evidence shows divergence between the first and second episodes of SARS-CoV-2 detection, indicative of reinfection. The first episode was apparently controlled by the innate immune response, without development of consistent neutralizing humoral immunity. Thus, patients remained susceptible to the second episode of COVID-19. This is the first evidence of SARS-CoV-2 reinfection in Brazil. The data suggests that SARS-CoV-2 reinfection may occur more often than anticipated.

Implication of all available evidence

Confirmations of reinfections imply that first episode of COVID-19 will not necessarily generate adaptive immunity to a second infection, which may contribute to new waves of the pandemics. After primary infection, a still overlooked proportion of patients develops poor adaptative immune response and may be susceptible to reinfection. These results also point out that vaccination may require prime/booster strategies.

Introduction

Confirmed cases of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has surpassed 55 million, along with 1.5 million deaths by 2019 coronavirus disease (COVID-19)[1]. New waves of the pandemics in different northern and southern hemisphere countries are the unequivocal evidence that herd immunity was not fully achieved and that susceptible population is countless [1]. In line with that, there is recent evidence that seroconversion after primary exposure to SARS-CoV-2 may be heterogeneous among the population [2]. Moreover, even for those who seroconvert, the sustainability of the immune response, as judged by IgG level, decay after weeks to months after primary exposure to the new coronavirus [3]. Paradoxically, few documented cases of reinfection by SARS-CoV-2 have been registered [4–7]. Since there is significant laboratory effort to unequivocally document reinfection and public health systems are unlikely prepared to provide molecular diagnosis, virus genome sequencing and serological evidence in a coordinated way during an ongoing pandemic – it is plausible that subsequent episode(s) of infection are currently overlooked. Indeed, respiratory viruses, such as influenza virus and, even other human and veterinary coronaviruses, provoke reinfections throughout the life-span of their hosts [8–10]. Confirmations of reinfections imply that waves of the pandemics do not dependent exclusively on individuals naïve to SARS-CoV-2 exposure. Moreover, documented reinfections may suggest that vaccines will require prime-booster strategies.

Our data here describe the follow-up of four individuals with two episodes of COVID-19. Among them, we fully documented reinfection in two patients, based on virus genomic sequencing. Clinical and molecular diagnosis, along with serological results, make it difficult to rule out that the other two households were not reinfected. These recurrent COVID-19 cases occurred during the last week of March and May, 2020, in Rio de Janeiro, Brazil. The first episode was apparently controlled by innate immune response and did not result in neutralizing humoral immunity to prevent against the second infection. Our study contributes to the better comprehension of the dynamics of the immune and virological responses in mild cases of COVID-19, which might hide individuals susceptible to reinfection.

Material and Methods

Ethics and Study Population

Four outpatients presenting mild self-limiting COVID-19 syndrome search for assistance and diagnosis at their convenience by spontaneous demand from March 2016 to August 2020. For comparisons, we also included 5 SARS-CoV-2-negative control subjects from the same City were the groups of patients live, Rio de Janeiro, Brazil. All patients had SARS-CoV-2 confirmed diagnostic through RT-PCR of nasopharyngeal swab. Sequential peripheral vein blood samples were also obtained. The National Review Board of Brazil approved the study protocol (Comissão Nacional de Ética em Pesquisa [CONEP] 30650420.4.1001.0008), and informed consent was obtained from all participants or patients' representatives.

Measurement of serum anti-SARS-CoV-2 antibodies

For quantitative analysis of anti-SARS-CoV-2 spike protein IgM, IgA and IgG antibodies, we performed the S-UFRJ test, as described previously [11]. Briefly, high binding ELISA plates were coated with 50 μL of SARS-CoV-2 spike protein (4 μg/mL in PBS) and incubated overnight. The coating solution was removed and 100 μL of PBS 1% BSA (blocking solution) was added and the plate was incubated at room temperature (RT) for 1-2 hours. The blocking solution was removed and 50 μL of diluted 1:40 (PBS 1% BSA) patient sera were added, subsequently, the sera were serially three-fold diluted in the plate, which was incubated at RT for 2 hours. Then, the plate was washed with 150 μL of PBS (5x) and 50 μL of 1:10000 goat anti-human IgG, IgA and IgM (Fc)-horseradish peroxidase antibody (Sourthen Biotech, Birmingham, USA) were added, and the plate was incubated for 1.5 hours at RT. The plate was washed again with 150 μL of PBS (5x) and then treated with TMB (3,3 ', 5,5; -tetramethylbenzidine) (Scienco, Brazil) until the reaction was stopped with 50 μL of HCl 1N. The optical density (OD) was read at 450 nm with 655 nm background compensation in a microplate reader (Bio-Rad Laboratories, Inc, California, USA).

Quantification of plasma cytokine levels.

Plasma samples from the acute and convalescent phases of the two episodes were collected in EDTA-containing tubes. Tubes were placed on ice and aliquoted. Commercial Elisa kits from R&D Systems (Minneapolis, MN, USA) were used to measure interferon (IFN)-α, -β and -γ, interleukin (IL)-6, -8 and -10, tumor necrosis factor α (TNF-α) and induced protein 10 (IP10)/C-

X-C motif chemokine ligand 10 (CXCL-10). This panel of mediators provides evidence for host production of antiviral, pro-inflammatory and regulatory responses.

Plaque reduction neutralization test (PRNT)

To determine serum titers to block SARS-CoV-2 infection, miniaturized PRNT was performed. In brief, human serum was heat inactivated (30 min, 56°C) prior to two-fold serial dilutions (from 1:4 to 1:2056). Diluted sera were incubated with 100 plaque forming units (PFU) of SARS-CoV-2 (GenBank # MT710714) for 1 h at 37 °C in 96-multiwell plates. Afterwards, mixtures of sera/virus were incubated with Vero E6 cells (2 x 10^4 cell/well) in 96-well plates during an additional 1h at 37°C. Next, fresh semi-solid medium containing 2.4 % of carboxymethylcellulose (CMC) was added to the wells, and cultures were maintained for 72 h at 37 ºC. Cells were fixed with 10 % formaline for 2 h at room temperature and stained with crystal violet (0.4 %). Endpoint dilution ~~to~~ inhibiting 90% PFU ($PRNT_{90}$) was scored. As a control, to validate each assay, a back-titration of the mock-treated virus was included. Our quality control accepts the final readout of the virus input to be equivalent to 100 ± 20 PFU. Plaque numbers were scored in at least 3 independent experiments with technical replicates by two independent blinded readers, to determine the $PRNT_{90}$. A schematic and representative description of the assay read is presented in Figure S1.

Molecular diagnosis

The total viral RNA was extracted using QIAamp Viral RNA (Qiagen®), according to manufacturer's instructions. Quantitative RT-PCR was performed using GoTaq® Probe qPCR and RT-qPCR Systems (Promega) in a StepOne™ Real-Time PCR System (Thermo Fisher Scientific). Amplifications were carried out in 25 µL reaction mixtures containing 2× reaction mix buffer, 50 µM of each primer, 10 µM of probe, and 5 µL of RNA template. Primers, probes, and cycling conditions used to detect the SARS-CoV-2 RNA were those recommended by the Centers for Disease Control and Prevention (CDC) [12]. The standard curve method was employed for virus quantification, using synthetic RNA for gene N (Microbiologics, Minnesota, USA). The ct values for this target were compared to those obtained with different cell amounts (10^7 to 10^2), for reaction calibration.

Genomic analysis

Total viral RNA from nasopharyngeal swabs was extracted using QIAamp Viral RNA (Qiagen®, Hilden, Germany), with minor modifications [13]. In brief, extraction was performed in 2 ml of sample/lysis buffer (1:1) without RNA carrier, and purified RNA was obtained after binding and elution from a single silica column. For better yields, a 50 ul eluate was repetitively loaded (4x) to the same column. Tests were performed to evaluate if digestion with DNAse I and depletion of rRNA enhanced the quantity/quality of SARS-CoV-2-related reads. Samples negative for SARS-CoV-2 and positive for Zika or chikungunya were included as controls.

An amplicon-based enrichment strategy was carried out with the ATOPlex SARS-CoV-2 Full Length Genome Panel v1.0 (kindly donated by MGI Tech Co., Shenzhen, China), to improve sequencing readout. For library construction, RNA samples were first quantified with the Qubit™ RNA BR Assay Kit (Thermo Fisher Scientific, Foster City, CA), according to manufacturer's instructions. Approximately 5 ng of each sample were then used as inputs to reverse transcription (RT) reactions, followed by two-step multiplex PCR-based genome amplifications and dual adaptor indexing (ie barcoding) using proprietary primer sets. Products were purified with DNA Clean beads at a 5:6 volume ratio and subsequent washing steps with 80% ethanol. Next, individual libraries were quantified with the Qubit™ 1X dsDNA HS Assay Kit (Thermo Fisher Scientific, Foster City, CA) and homogeneously pooled to a total sum of 400 ng, before being submitted to denaturation, circularization and digestion steps. Finally, single- stranded circular DNA library pools were converted to DNA nanoballs by rolling circle amplification and submitted to pair-end sequencing (100 nt) on the MGISEQ-2000 platform (MGI Tech Co., Shenzhen, China).

Genomic sequences were quality-scored, filtered, trimmed and assembled in contigs through a validated workflow for SARS-CoV-2 [14]. Genomes were aligned with MAFFT [15] or ClustalW [16], and phylogenies were constructed with Mega 7.0 [17,18], using the Jukes-Cantor model for Maximum Likelihood estimates, by applying Neighbor-Join and BioNJ algorithms [19]. The tree with the highest log likelihood was used. Alternatively, MrBayes 3.2.7 was used for Bayesian inference with a relaxed clock model with a priori model testing using the G + I nucleotide substitution model, selected by jModelTest v1.6. The tree was visualized and edited with FigTree v.1.4.2 (http://tree.bio.ed.ac.uk). SARS-CoV-2 clades were determined by: https://clades.nextstrain.org [20]. To categorize mutations/polymorphisms, the SARS-CoV-2 reference genome Wuhan-Hu-1 (GISAID EPI ISL #402125) was aligned to out sequences. The

original sequences used in this work are publicly available on https://nextstrain.org/ncov: GISAID EPI ISL #636737, 636834-636838). The dataset included in the analysis contained representative sequences of the emerging clades associated to our sequences, 19A and 20B, as well as sequences form the genome 20A – as a negative control (Table S1).

Results

Patient A, a 54-years old male requested a RT-PCR test for SARS-CoV-2 on March 23rd, because of a recurrent headache on the prior two days. He also had previous contact with a syndromic co-worker, presenting cough and what he described as signals of influenza-like illness. Patient A had detectable viral load in nasopharyngeal swabs (cycle threshold; ct = 27.41) of approximately 10^5 copies/ml (Table 1). Patient B, a 57-years old female with previous history of discoid lupus erythematosus, was tested because of intimate contact with Patient A. She tested positive for COVID-19 on March 24th, with a ct value of approximately 36.31 (~10^3 copies/ml) in the nasopharyngeal swabs (Table 1). Two days afterwards, she presented diarrhea (Table 1).

Patient B is a household of patients C and D, a couple at age of 34 years-old. Patients C and D were not on social isolation because of their job duties. Although Patient C was asymptomatic, he displayed a ct = 35.71 (10^3 copies/ml) on March 25th (Table 1). Patient D was negative after molecular testing on March 26th, but one week later, she presented a detectable viral load of 36.01 (10^3 copies/ml) and started a diarrhea in the following days (Table 1). On March 27th, Patients presented a balanced upregulation of markers of innate immune response, compared to healthy controls (Figure 1), pro-inflammatory mediators (IL-6, IL-8 and TNF-α), regulatory (IL-10) and chemotactic (CXCL-10) signals and IFN-γ. On the other hand, anti-SARS-CoV-2 IgM, IgA or IgG antibodies were not detected in the patients' sera on this same day (Table 1 and Figure S2).

Although we run out of respiratory samples from Patients A and D, we were able to obtain full-length SARS-CoV-2 genome for Patients B and C (Table 1). Complete genome sequencing, with Phred quality score >30, composed of 140.000 to 20.000.000 reads and coverage of 100- to 10.000-times, argues against a false-positive RT-PCR result (Table S2, first column). Patients B and C were respectively affected by SARS-CoV-2 strains associated to emerging clades 19A and 20B in the episode of COVID-19 from March 2020 (Figures 2, S3 and Table 1). The detection of the two lineages indicates that Patients B and C contracted the new coronavirus by

independent means, and not transmitted the virus to each other (Figures 2, S3 and Table 1). These distinct lineages co-circulated in Brazil in March 2020 when multiple introductions of the SARS-CoV-2 occurred [21]. Indeed, because Patient B was on social isolation, her intimate contact with Patient A represented the only risk of infection. The syndromic contact of Patient A that unfolds the COVID-19 testing among the households was a traveler. Thus, the association of Patient B, and by epidemiological linkage Patient A, to emerging clade 19A, which is closer to Wuhan-01, is consistent (Figure 2 and S3). Since Patient C was frequently exposed to various probable sources of contamination due to his work. Indeed, he was infected by an emerging clade 20B virus, the most prevalent variant in Brazil (Figure 2 and S3; PATIENT_B_FIRST_EPISODE and PATIENT_C_FIRST_EPISODE). Patients B and C recovered from the mild COVID-19 episode and were retested on the first half of April, when they presented negative RT-PCR results.

On the last week of May, when COVID-19 cases in Rio de Janeiro were on the peak of pandemics [22], these four individuals sought for medical assistance again, with more exuberant signs and symptoms of SARS-CoV-2 infection than previously (Table 1). During the second episode, they presented fever and/or cough, along with indolence, headache, body ache, anosmia and ageusia. Real time RT-PCR revealed higher viral loads in the nasopharyngeal swabs than at the time of the first infection, with ct values of 21.76 (~10^7 copies/ml), 21.84 (~10^7 copies/ml), 26.38 (~10^5 copies/ml), and 16.87 (~10^9 copies/ml) for patients A, B, C and D, respectively (Table 1).

On June 3rd, a week after the second episode, anti-SARS-CoV-2 immunoglobulins were detected only for patients A and B, along with low to non-neutralizing activity (Table 1 and Figure S2). This serological sample from June is informative that the first episode of COVID-19 was not followed by a sustained neutralizing humoral response, as judged by $PRTN_{90}$ titers (Table 1). Since signals of an effector humoral memory were mild (Table 1), we could speculate that the enhanced production of IFNs and pro-inflammatory mediators led to resolution of the first episode of COVID-19 (Figure 1) – meaning indeed that the patients could still be susceptible to SARS-CoV-2. In fact, since the second episode, most of cytokine levels were kept higher, compared to healthy volunteers (Figure 1).

Serological samples consistent with convalescence phase of the second episode, collected on July 10th, 40 days after the claimed episode of reinfection, revealed that all individuals presented detectable immunoglobulin levels and their best $PRNT_{90}$ results (Table 1 and Figure S2),

declining thereafter on August 10th (Table 1 and Figure S2). These data mean that all four patients were able to develop secondary humoral adaptive responses, which might have contributed to limit viral dissemination and control the infection. In parallel, these individuals continuously presented up-regulated pro-inflammatory markers (Figure 1), which is consistent with boostered response to a second SARS-CoV-2 exposure. In particular, the dichotomic markers of inflammation and regulatory responses, respectively, TNF-α and IL-10, also decreased in August, compared to previous months (Figure 1).

At the second episode, we fully sequenced the SARS-CoV-2 genome from all patients (Tables 1, S2 and Figures 2 and S3). SARS-CoV-2 sequences from the reinfection clustered together, suggesting a household transmission (Figure 2 and S3, Patient_A to Patient_D). The emerging genotype 20B, which is the major variant circulating in Brazil since May 2020, was detected in all samples from the second episodes (Figure 2 and S3 and Table 1). For Patient B, the first and second episodes were associated with the emerging clades 19A and 20B, respectively (Figure 2 and S3). Two episodes provoked by genetically distinct lineages support the notion of reinfection.

For patient C, although both episodes were associated with clade 20B, they clustered apart on the phylogeny with significant statistical support: by 86 % of bootstrap using maximum likelihood (Figure 2) and by Bayesian inference (Figure S3). Besides, there are genetic markers in the SARS-CoV-2 genome that were different in the two episodes of COVID-19 (Table S2). Virus genome from the second and first episodes of Patient C diverges at the genes encoding the non-structural protein (nsp) 3, 3C-like proteinase and exonuclease (Table S2). Besides the genetic variations, the poor development of anti-SARS-CoV-2 serology, between the two episodes of infection, points towards the interpretation of reinfection.

Moreover, Patient A and D, for whom respiratory samples from the first episode were not enough for sequencing analysis, displayed clinical symptomatology, positive RT-PCRs, enhanced cytokine/chemokine levels and epidemiological linkage that make it difficult to rule out that they did not present two events of SARS-CoV-2 infections.

Altogether, compared to previous studies on case reports of reinfection, here we describe a small cluster of individuals who presented laboratory-based evidence of two episodes of COVID-19.

Discussion

Seasonal coronaviruses and influenza A virus are respiratory viruses that may cause reinfection [8,10]. Drifted strains of influenza A virus emerge with mutations to scape immune response [23]. Nevertheless, sustainability of the immune response to influenza lasts for roughly a year for the general population [8], meaning that yearly vaccination would occur even if vaccine-escape mutant did not emerge. Influenza A virus is more mutagenic than coronaviruses, which are equipped with an exonuclease, the nsp14 [24], that function as proofreading enzyme. Of note, in veterinary medicine, domestic mammals have commonly coronavirus reinfection [9]. It is likely that in mammals the adaptive, memory-generating, immunity to coronaviruses is heterogeneously sustainable and some events of infection are controlled at the level of the innate immunity.

We fully documented reinfection in two genetically unrelated individuals in Rio de Janeiro, Brazil, describing patients that presented for two times, in a near two-month interval, clinical and laboratory diagnosis of COVID-19. Virus polymorphisms from the primary and second episodes and negative RT-PCR between the events strengths the argument towards reinfection. Throughout the follow-up of these individuals, generation of neutralizing anti-SARS-CoV-2 titers was achieved just after the second infection, meaning that they were still vulnerable after the primary episode. Although we have not measured the serum levels of anti-SARS-CoV-2 antibodies after the first infection, it is plausible that the first episode was resolved without a relevant engagement of humoral memory.

SARS-CoV-2 reinfection has been described in USA, Ecuador, Belgium/Netherlands and Hong Kong, on the basis of case reports [4–7]. These Brazilian patients configure a small cluster claimed to be reinfected. Similarly to the cases of USA and Ecuador [4,7], reinfection patients in Brazil presented more symptoms than in the first episode. Antibody-dependent enhancement or simply exposure to higher amounts of virus could be the reason to the change from asymptomatic/oligosymptomatic to syndromic. We understand that current literature is not enough to associate the second infection with a more virulent strain. In our study, primary and second infections were caused by a strain carrying the D614G mutation in the spike, which has been associated with higher replication efficiency [25]. Although V125F change in the nsp14 is a non-conservative mutation that may increase the volume in the loop between beta-sheets number 5 and 6, which could affect its methyltransferase activity [26] – it is unlikely to increase the virulence on the second episode. On the other hand, changes in nsp6 [27] and ORF6 [28] may result in viral evasion form innate immunity.

Primary infections of patients B and C were associated to emerging clades 19A and 20B –, indicating that these co-habitants were infected independently. Indeed, while one patient was on social isolation, the other was in an active circulation. The co-circulation of these clades of SARS-CoV-2 is consistent with the COVID-19 databases [29] and the multiple introductions of the new coronavirus in Brazil [21]. In the following months, emerging clade 20B turned to be the most prevalent genotype, representing 60 % of the deposited genomes on GISAID [29]. The detection of clade 20B on the second episode of COVID-19, by the end of May, is associated with the peak of the pandemic in Rio de Janeiro, Brazil [22].

Distinct clades of SARS-CoV-2 were found in the primary and secondary respiratory samples from patient B, supporting the argument on reinfection. For patient C, both the first and second detections of SARS-CoV-2 were associated with clade 20B. Although viral persistence could be imagined in this scenario, SARS-CoV-2 genomic sequences from the first and second episodes do not cluster together in the same branch, such as it did for the immunocompromised individual that shed SARS-CoV-2 for 150 days [30]. Thus, phylogeny does not support the interpretation of persistence, by different methods. By branching apart, SARS-CoV-2 genomes associated to patient C strengthen the chances of a relevant degree of variation [31], pointing towards the direction of reinfection. In the documented case of reinfection in USA, both episodes of molecular diagnosis of COVID-19 were also associated with the same emerging clade [4], but they also clustered apart from each other in the phylogeny, similarly to virus genomes from patient C. Whereas the detection of two episodes of SARS-CoV-2 infection from patient C were separated by over 60 days, prolonged virus shedding in the nasopharyngeal swabs from mild cases lasts for no more than 22 to 46 days [32] – another evidence that reduce the chances of persistence.

Results of SARS-CoV-2 reinfection echo that prime-booster may be necessary to achieve humoral protection and underscore that sustainability of the immune response may be heterogeneous. We documented that these individuals with mild COVID-19 displayed a balanced innate immune response. Although cytokine storm has been associated with severe COVID-19 [33], we interpret that the balanced innate immune response might have led to infection resolution [34]. The natural history of mild COVID-19 described for these individuals may be also representative of many individuals exposed to the first wave of the pandemics, leading to the hypothesis that they would also be susceptible to other episodes of SARS-CoV-2 infections. Perhaps, the rigor to document reinfection through virus genomic sequencing in

disorganized public health systems that succumb to pandemic may have overlooked the frequency of this event. It is thus important to stimulate cohort studies to further quantify the incidence of reinfection.

Considering this study, there will be five documented reports of reinfection. Along with the cases described elsewhere [4–7], the small cluster found here should open the debate whether cases of reinfection are more numerous but poorly documented. This hypothesis could be addressed by cohort studies following up individuals between pandemic waves. Our investigation may help to identify laboratory endpoints for studies on reinfection.

References

1 WHO Coronavirus Disease (COVID-19) Dashboard | WHO Coronavirus Disease (COVID-19) Dashboard. https://covid19.who.int/?gclid=CjwKCAiAtK79BRAIEiwA4OskBvIhZPswIQFwuW7goWU9PiYw-a5jR_WeCMyPg7833I_jrccLzoflCxoCV7EQAvD_BwE (accessed Nov 12, 2020).

2 Ripperger TJ, Uhrlaub JL, Watanabe M, *et al.* Detection, prevalence, and duration of humoral responses to SARS-CoV-2 under conditions of limited population exposure. *medRxiv : the preprint server for health sciences* 2020; : 2020.08.14.20174490.

3 Hueston L, Kok J, Guibone A, *et al.* The Antibody Response to SARS-CoV-2 Infection. *Open Forum Infectious Diseases* 2020; **7**. DOI:10.1093/ofid/ofaa387.

4 Tillett RL, Sevinsky JR, Hartley PD, *et al.* Genomic evidence for reinfection with SARS-CoV-2: a case study. *The Lancet Infectious Diseases* 2020; **0**. DOI:10.1016/s1473-3099(20)30764-7.

5 To KK-W, Hung IF-N, Ip JD, *et al.* Coronavirus Disease 2019 (COVID-19) Re-infection by a Phylogenetically Distinct Severe Acute Respiratory Syndrome Coronavirus 2 Strain Confirmed by Whole Genome Sequencing. *Clinical Infectious Diseases* 2020; published online Aug 25. DOI:10.1093/cid/ciaa1275.

6 Van Elslande J, Vermeersch P, Vandervoort K, *et al.* Symptomatic SARS-CoV-2 reinfection by a phylogenetically distinct strain. *Clinical Infectious Diseases* 2020; published online Sept 5. DOI:10.1093/cid/ciaa1330.

7 Prado-Vivar B, Becerra-Wong M, Guadalupe JJ, *et al.* COVID-19 Re-Infection by a Phylogenetically Distinct SARS-CoV-2 Variant, First Confirmed Event in South America. *SSRN Electronic Journal* 2020; published online Sept 9. DOI:10.2139/ssrn.3686174.

8 Krammer F. The human antibody response to influenza A virus infection and vaccination. Nature Reviews Immunology. 2019; **19**: 383–97.

9 Decaro N, Martella V, Saif LJ, Buonavoglia C. COVID-19 from veterinary medicine and one health perspectives: What animal coronaviruses have taught us. *Research in Veterinary Science* 2020; **131**: 21–3.

10 Galanti M, Shaman J. Direct Observation of Repeated Infections With Endemic Coronaviruses. *The Journal of Infectious Diseases* 2020. DOI:10.1093/infdis/jiaa392.

11 F Alvim RG, Lima TM, S Rodrigues DA, *et al.* An affordable anti-SARS-COV-2 spike protein ELISA test for early detection of IgG seroconversion suited for large-scale surveillance studies in low-income countries. *medRxiv* 2020; : 2020.07.13.20152884.

12 Centers for Disease Control and Prevention C. Real-time RT-PCR Primers and Probes for COVID-19 | CDC. https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel-primer-probes.html (accessed Nov 11, 2020).

13 Metsky HC, Matranga CB, Wohl S, *et al.* Zika virus evolution and spread in the Americas. *Nature* 2017; **546**: 411–5.

14 Cleemput S, Dumon W, Fonseca V, *et al.* Genome detective coronavirus typing tool for rapid identification and characterization of novel coronavirus genomes. *Bioinformatics* 2020; **36**: 3552–5.

15 Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic acids research* 2005; **33**: 511–8.

16 Larkin MA, Blackshields G, Brown NP, *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* 2007; **23**: 2947–8.

17 Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 2018; **35**: 1547–9.

18 Felsenstein J. CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP. *Evolution* 1985; **39**: 783–91.

19 Jukes TH, Cantor CR. Evolution of Protein Molecules, chapter 24. Mammalian Protein Metabolism. New York: Academic Press, 1969.

20 Nextclade. https://clades.nextstrain.org/ (accessed Nov 14, 2020).

21 Candido DS, Claro IM, de Jesus JG, *et al.* Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science* 2020; **369**: 1255–60.

22 Secretaria de Saúde do Estado do Rio de Janeiro. Painel de monitoramento de Covid-19 no Estado do Rio de Janeiro. http://painel.saude.rj.gov.br/monitoramento/covid19.html# (accessed Nov 24, 2020).

23 Jaramillo JM, Ma J, van den Driessche P, Yuan S. Host contact structure is important for the recurrence of Influenza A. *Journal of Mathematical Biology* 2018; **77**: 1563–88.

24 Romano M, Ruggiero A, Squeglia F, Maga G, Berisio R. A Structural View of SARS-CoV-2 RNA Replication Machinery: RNA Synthesis, Proofreading and Final Capping. Cells. 2020; **9**. DOI:10.3390/cells9051267.

25 Groves DC, Rowland-Jones SL, Angyal A. The D614G mutations in the SARS-CoV-2 spike protein: Implications for viral infectivity, disease severity and vaccine design. *Biochemical and Biophysical Research Communications* 2020; published online Nov. DOI:10.1016/j.bbrc.2020.10.109.

26 Krafcikova P, Silhan J, Nencka R, Boura E. Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. *Nature Communications* 2020; **11**. DOI:10.1038/s41467-020-17495-9.

27 Cottam EM, Whelband MC, Wileman T. Coronavirus NSP6 restricts autophagosome expansion. *Autophagy* 2014; **10**: 1426–41.

28 Sarif Hassan S, Pal Choudhury P, Uversky VN, *et al.* Variability of Accessory Proteins Rules the SARS-CoV-2 Pathogenicity in the National Center for Biotechnology Information. 2020. DOI:10.1101/2020.11.06.372227.

29 GISAID. GISAID - Initiative. https://www.gisaid.org/ (accessed Nov 14, 2020).

30 Choi B, Choudhary MC, Regan J, *et al.* Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *The New England journal of medicine* 2020; : NEJMc2031364.

31 Koyama T, Platt D, Parida L. Variant analysis of SARS-cov-2 genomes. *Bulletin of the World Health Organization* 2020; **98**: 495–504.

32 Sun J, Xiao J, Sun R, *et al.* Prolonged persistence of SARS-CoV-2 RNA in body fluids. *Emerging Infectious Diseases* 2020; **26**: 1834–8.

33 de la Rica R, Borges M, Gonzalez-Freire M. COVID-19: In the Eye of the Cytokine Storm. Frontiers in Immunology. 2020; **11**: 558898.

34 Arunachalam PS, Wimmers F, Mok CKP, *et al.* Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science* 2020; **369**: 1210–20.

Acknowledgment

Thanks are due to Dr. Carmen Beatriz Wagner Giacoia Gripp from Oswaldo Cruz Institute (IOC) and Dr. Marco Alberto Medeiros from Biomanguinhos for assessments related to BSL3 facility and Sequencing platform, respectively. We thank Dr. Gonzalo Bello (IOC) Dr.

Dumith Chequer Bou-Habib (IOC), Dr. Willian Provance (CDTS) and Dr. Fabiano Thompson (UFRJ) for insightful discussions. We deeply appreciated the MGI, a partner in the implementation of next generation sequencing through collaborations with Oswaldo Cruz Foundation (Fiocruz), in especial for challenging samples of COVID-19.

Authors Contributions

Clinical Surveillance – FAB, PTB

Patients Enrolment – NFR, FAB

Immune assessments – NFR, CQS, DR, IGA, VCS

Sequencing – NFR, APDS, MCS, FBS, MAF, JG, HJ, HT

Bioinformatics – APDS, MCS, DAT

Study coordination – FAB, PTB, CMM, TMLS

Manuscript preparation and revision –NFR, PTB, AMV, TMLS

All authors revised and approved the manuscript.

Declaration of interests

Authors declare no conflict of interest

Role of the funding source

The funder had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all data in the study and had final responsibility for the decision to submit for publication.

Figures

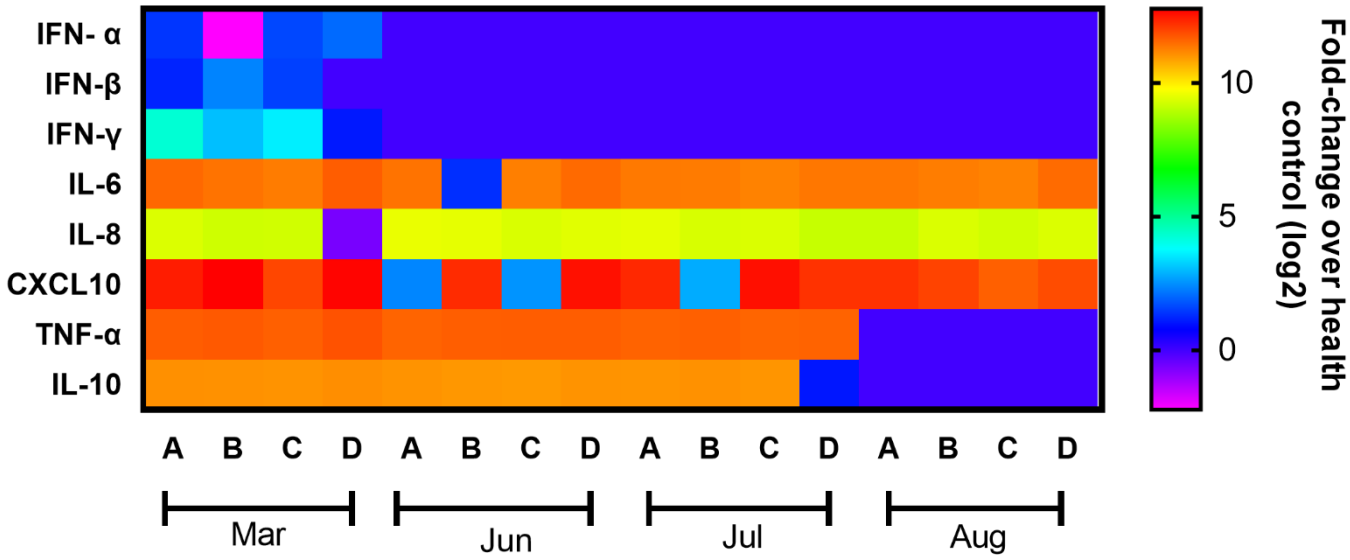

Figure 1 – The profile of innate immune response from patients with two episodes of SARS-CoV-2. The indicated mediators of the innate immunity were measured by ELISA for patients A-D on the indicated months. For comparisons, these molecules were also quantified in the plasma from 5 healthy donors negative for SARS-CoV-2. Heatmap displays the Log_2 ratio of the fold-change from the plasma of the patients over the healthy volunteers. The mean ± SEM for the healthy volunteers were the following: IFN-α = 20.4 ± 4.7 pg/ml, IFN-β = 26.0 ± 3.9 pg/ml, IFN-γ = 27.8 ± 7.8 pg/ml, IL-6 = 13.4 ± 1.7 pg/ml, IL-8 = 137 ± 21.6 pg/ml, IL-10 = 165.4 ± 40.7 pg/ml, TNF-α = 33.8 ± 11.5 pg/ml, CXCL-10 = 61.0 ± 27.3 pg/ml.

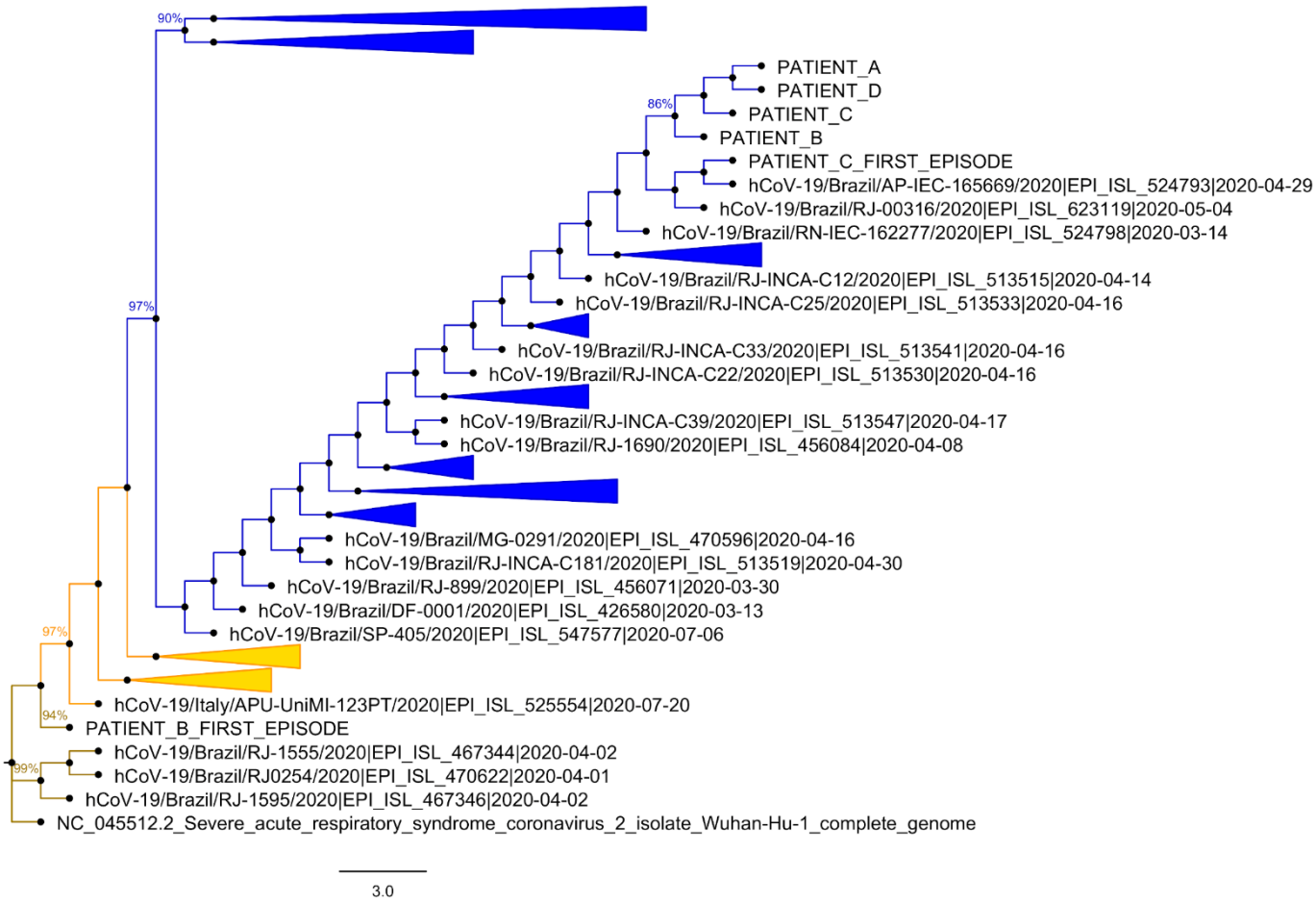

Figure 2 – Phylogenetic analysis of SARS-CoV-2 genomes from reinfected patients. Representative genomes deposited on GISAID (Table S1 and Fig S3) were compared to sequences from virus genomes found in the respiratory samples from the first episode of patient B and C, and second episode of patients A to D. Emerging clades 19A, 20A and 20B are brown, orange and blue, respectively. The evolutionary history was inferred by using the Maximum Likelihood method and Jukes-Cantor model [19]. The tree with the highest log likelihood (-46487.36) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Jukes-Cantor model, and then selecting the topology with superior log likelihood value. There were a total of 29920 positions in the final dataset. Evolutionary analyses were conducted in MEGA 7.0 [17,18]. A total of 1000 bootstraps were used and condensed phylogenetic tree rooted by reference genome Wuhan-Hu-1 (#EPI_ISL_402125) is displayed.

Table 1- Clinical, demographic, virological and serological aspects of the cluster of

576 **Patients.**

General Information	Individual code	Patient A	Patient B	Patient C	Patient D
	Gender	Male	Female	Male	Female
	Age (years-old)	54	57	34	34
	Co-morbidities	None	Discoid lupus erythematosus	None	None
Primary infection	**Data of onset ilness**	March 21st	March 26th	Asymptomatic	March 31st
	Symptons	Headache	Mild diarrhea	No	Mild diarrhea
	RT-PCR (log10; copies/mL) Date	5.12 March 23rd	3.21 March 24th	3.83 March 24th	3.01 April 2nd
	Serology March 27th	IgM/IgA/IgG detectable	IgM/IgA/IgG detectable	IgM/IgA/IgG detectable	IgM/IgA/IgG detectable
	$PRNT_{90}$/25uL March 27th	<1:4	<1:4	<1:4	<1:4
	Sequencing	Not enough sample	emerging clade 19A	emerging clade 20B	Not enough sample
	Accession ID	N/A	EPI_ISL_636834	EPI_ISL_636836	N/A
Secondary Infection	**Data of onset ilness**	May 25th	May 26th	May 27th	May 30th
	Symptons	Fever, Dry cough, tiredness, body ache, anosmia, ageusia	Fever, diarrhea, headache, body ache, anosmia, ageusia	Fever, nausea, tiredness, headache, body ache	Dry cough, diarrhea, tiredness, headache, body ache, anosmia, ageusia,
	RT-PCR (log10; copies/mL) Date	7.31 May 29th	7.42 May 29th	5.18 May 29th	9.61 May 29th
	Serology June 3rd	IgM/IgA/IgG detectable	IgM/IgA/IgG detectable	IgM/IgA/IgG undetectable	IgM/IgA/IgG undetectable
	$PRNT_{90}$/25uL June 3rd	1:16	<1:4	<1:4	<1:4
	Sequencing	Emerging clade 20B	Emerging clade 20B	Emerging clade 20B	Emerging clade 20B
	Accession ID	EPI_ISL_636737	EPI_ISL_636835	EPI_ISL_636837	EPI_ISL_636838
Follow-up	**Serology July 9th**	IgM/IgA/IgG detectable	IgM/IgA/IgG detectable	IgM/IgA/IgG detectable	IgM/IgA/IgG detectable
	$PRNT_{90}$/25uL	1:128	1:32	1:64	1:64

	July 9th				
	Serology Aug 10th	IgM/IgA/IgG detectable	IgM/IgA/IgG detectable	IgM/IgA/IgG detectable	IgM/IgA/IgG detectable
	$PRNT_{90}$/25uL Aug 10th	1:64	1:16	1:8	1:8

Supplementary Material

Viral genetic evidence and host immune response of a small cluster of individuals with two episodes of SARS-CoV-2 infection

Running title: SARS-CoV-2 reinfection in Brazil

Authors

Natalia Fintelman-Rodrigues[1,7], Aline de Paula Dias Da Silva[1,7], Monique Cristina dos Santos[1,7], Felipe B. Saraiva[2], Marcelo Alves Ferreira[7], João Gesto[2,7], Danielle A.S. Rodrigues[3], André Macedo Vale[3], Isaclaudia Gomes de Azevedo[1], Vinícius Cardoso Soares[1], Hui Jiang[5], Hongdong Tan[5], Diogo A. Tschoeke[6], Carolina Q. Sacramento[1,2], Fernando A. Bozza[7,8], Carlos M. Morel[9], Patrícia T. Bozza[1], Thiago Moreno L. Souza[1,9]

Affiliations

1-Laboratório de Imunofarmacologia, Instituto Oswaldo Cruz (IOC), Fundação Oswaldo Cruz (Fiocruz), Rio de Janeiro, RJ, Brazil.

2-Instituto de Tecnologia em Imunobiológicos (Bio-Manguinhos), Fiocruz, Rio de Janeiro, RJ, Brazil

3-Laboratório de Biologia de Linfócitos – LBL, Instituto Biofísica, Universidade Federal do Rio de Janeiro, Rio de Janeiro - RJ – Brazil

4-Programa de Imunologia e Inflamação, Universidade Federal do Rio de Janeiro, UFRJ, Rio de Janeiro, RJ, Brazil.

5- MGI Tech Co., Ltd., Building No.11, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China

6- SAGE/COPPE, UFRJ, Rio de Janeiro, RJ, Brazil.

7-Instituto Nacional de Infectologia Evandro Chagas, (INI), Fundação Oswaldo Cruz (Fiocruz), Rio de Janeiro, RJ, Brazil.

8-D'Or Institute for Research and Education

9-National Institute for Science and Technology on Innovation in Diseases of Neglected Populations (INCT/IDPN), Center for Technological Development in Health (CDTS), Fiocruz, Rio de Janeiro, RJ, Brazil.

Corresponding author: Thiago Moreno Lopes e Souza, PhD

Fundação Oswaldo Cruz (Fiocruz)
Centro de Desenvolvimento Tecnológico em Saúde (CDTS)
Instituto Oswaldo Cruz (IOC)
Pavilhão Osório de Almeida, sala 16
Av. Brasil 4365, Manguinhos, Rio de Janeiro - RJ, Brasil, CEP 21060340
Tel.: +55 21 2562-1311
Email: tmoreno@cdts.fiocruz.br

Figures

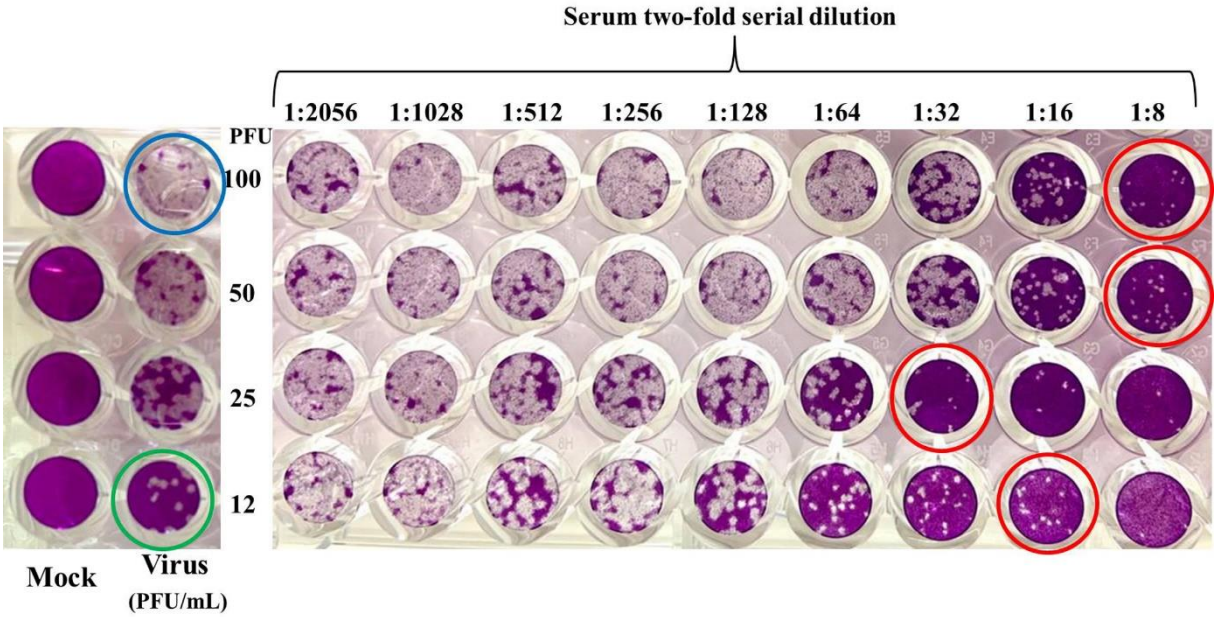

Figure S1 - Representative PRNT. Representative read out of PRNT. Two-fold serial dilutions (from 1:4 to 1:2056) of human sera was incubated in duplicates with approximately 100 plaque forming units (PFU) of SARS-CoV-2. The sera/virus mixture was incubated for 1h at 37 °C and then, added to Vero E6 cells (2 x 10^4 cell/well) in 96-well plates and incubated for an additional 1h at 37°C. Next, medium with 2.4 % CMC was added. After, 72 h at 37 ºC, cells are fixed with 10 % formaline and stained with crystal violet (0.4 %). Mock – uninfected control. Virus (PFU/ml) - Back-titration of the mock-treated virus was included in each experiment, the undiluted virus input incubated with the sera is highlighted by the blue circle, two-fold serial dilution of the virus is shown, the last dilution of the virus input (1:8) produced 12/13 PFU (green circle), validating the assay. The endpoint dilution of the sera capable of neutralizing the virus input (blue circle) by 90 % was expected to produce around 10 PFU, these dilutions are shown by the red circles.

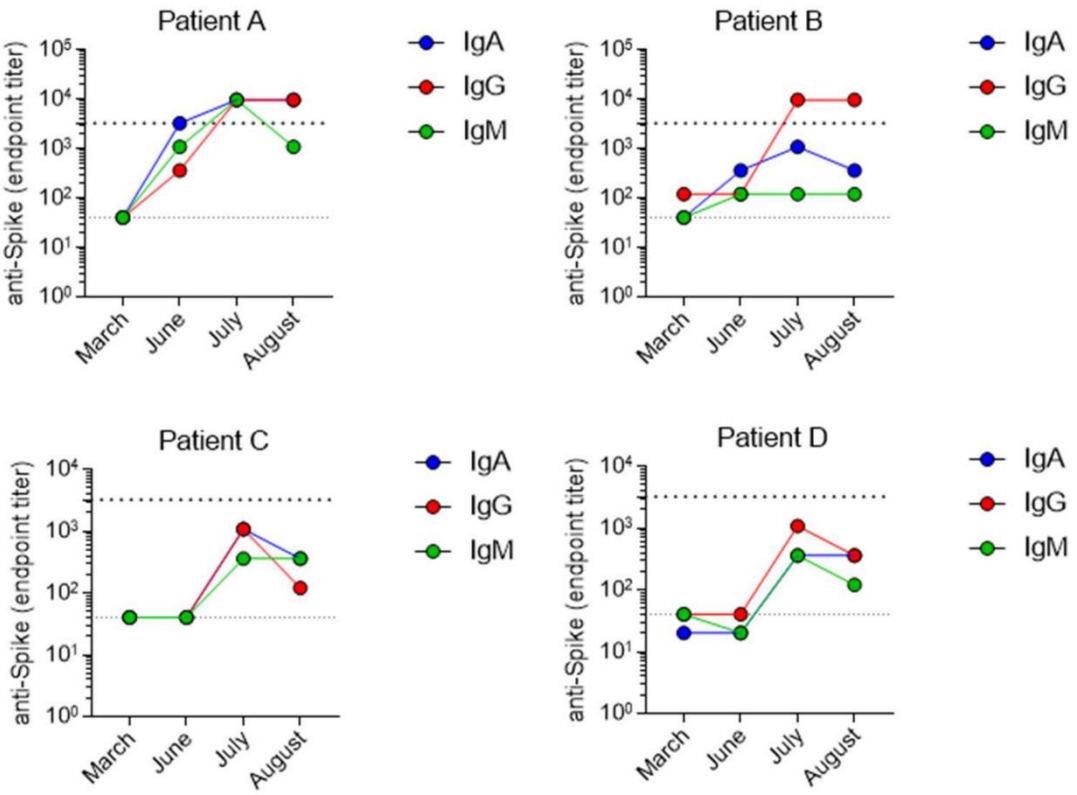

Figure S2 - Quantitative analysis of IgA, IgM and IgG from patients during primary, second infections and two months after second infection. Plasma samples from patients were collected in March, June, July and August for longitudinal detection of anti-Sipke IgM (green), IgA (blue) and IgG (red) antibodies (A-D). The relative levels of antibodies were shown as endpoint titers of patient samples values for the O.D. [mean + 3 standard deviation (X + 3SD)] negative controls on the same ELISA plate. The dashed horizontal line represents the endpoint titer value between 10,000 and 30,000. The samples below the dotted line are considered negative.

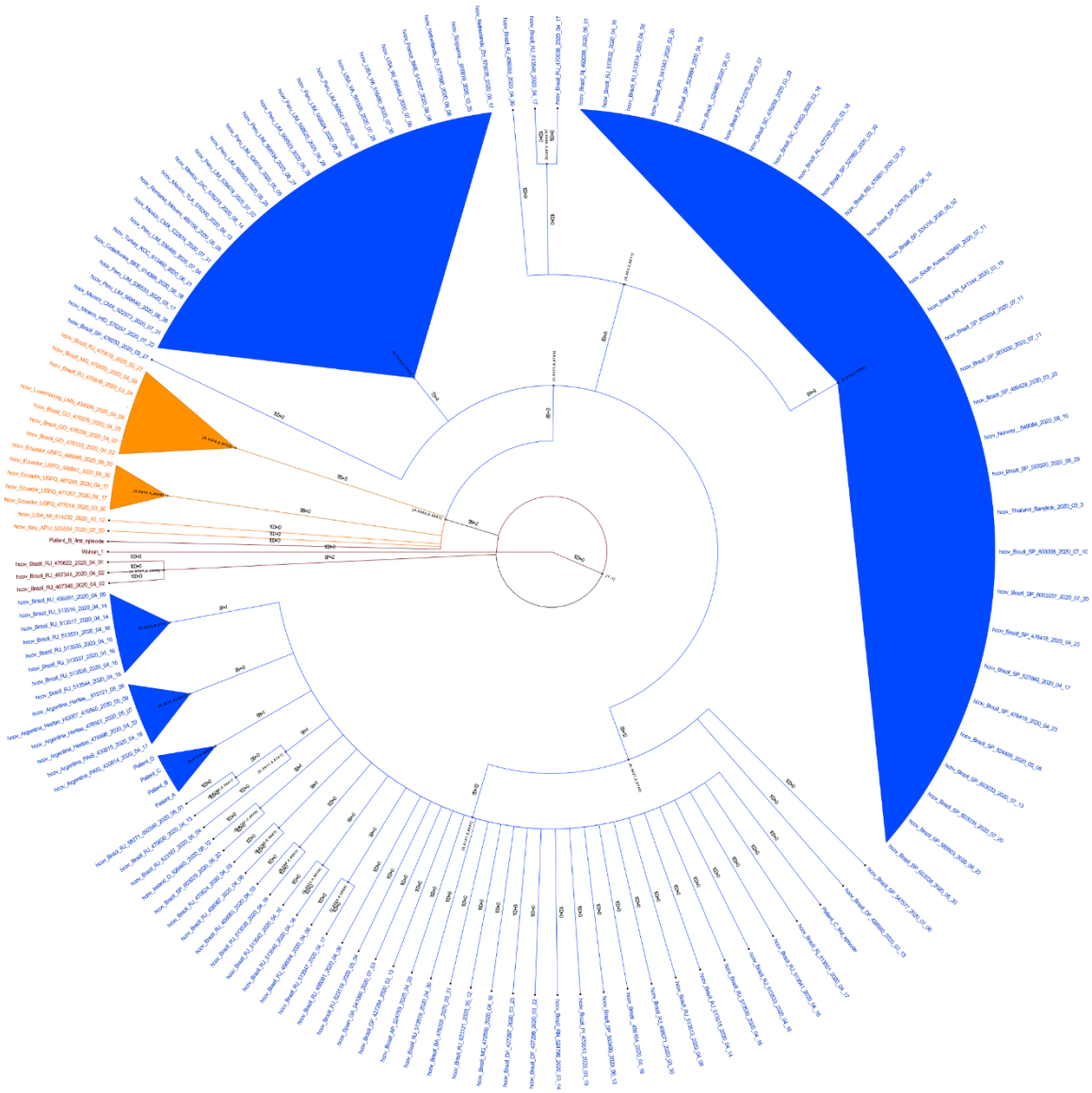

Figure S3- Phylogeny constructed by Bayesian inference, through MrBayes 3.2.7, assuming a relaxed clock model with a priori model testing using the G + I nucleotide substitution model, selected by jModelTest v1.6. Emerging clades 19A, 20A and 20B are brown, orange and blue, respectively. Posterior and anterior probabilities are presented for each branch.

714 **Table S1 – List of access code from sequences used in tree construction to compared to patients virus sequence.**

Number	Name	Accession code
1	hCoV-19/Wuhan/Hu-1/2019	EPI_ISL_402125
2	hCoV-19/Brazil/RJ-UFRJ-9331/2020|EPI_ISL_492036|2020-06-01	EPI_ISL_492036
3	hCoV-19/Brazil/RJ-INCA-C44/2020|EPI_ISL_513551|2020-04-17	EPI_ISL_513551
4	hCoV-19/Brazil/RJ-INCA-C39/2020|EPI_ISL_513547|2020-04-17	EPI_ISL_513547
5	hCoV-19/Brazil/RJ-INCA-C38/2020|EPI_ISL_513546|2020-04-17	EPI_ISL_513546
6	hCoV-19/Brazil/RJ-INCA-C37/2020|EPI_ISL_513545|2020-04-16	EPI_ISL_513545
7	hCoV-19/Brazil/RJ-INCA-C36/2020|EPI_ISL_513544|2020-04-16	EPI_ISL_513544
8	hCoV-19/Brazil/RJ-INCA-C34/2020|EPI_ISL_513542|2020-04-16	EPI_ISL_513542
9	hCoV-19/Brazil/RJ-INCA-C33/2020|EPI_ISL_513541|2020-04-16	EPI_ISL_513541
10	hCoV-19/Brazil/RJ-INCA-C31/2020|EPI_ISL_513539|2020-04-16	EPI_ISL_513539
11	hCoV-19/Brazil/RJ-INCA-C30/2020|EPI_ISL_513538|2020-04-16	EPI_ISL_513538
12	hCoV-19/Brazil/RJ-INCA-C29/2020|EPI_ISL_513537|2020-04-16	EPI_ISL_513537
13	hCoV-19/Brazil/RJ-INCA-C27/2020|EPI_ISL_513535|2020-04-16	EPI_ISL_513535
14	hCoV-19/Brazil/RJ-INCA-C25/2020|EPI_ISL_513533|2020-04-16	EPI_ISL_513533
15	hCoV-19/Brazil/RJ-INCA-C24/2020|EPI_ISL_513532|2020-04-16	EPI_ISL_513532
16	hCoV-19/Brazil/RJ-INCA-C23/2020|EPI_ISL_513531|2020-04-16	EPI_ISL_513531
17	hCoV-19/Brazil/RJ-INCA-C22/2020|EPI_ISL_513530|2020-04-16	EPI_ISL_513530
18	hCoV-19/Brazil/RJ-INCA-C14/2020|EPI_ISL_513517|2020-04-14	EPI_ISL_513517
19	hCoV-19/Brazil/RJ-INCA-C13/2020|EPI_ISL_513516|2020-04-14	EPI_ISL_513516
20	hCoV-19/Brazil/RJ-INCA-C12/2020|EPI_ISL_513515|2020-04-14	EPI_ISL_513515
21	hCoV-19/Brazil/RJ-INCA-C07/2020|EPI_ISL_513514|2020-04-08	EPI_ISL_513514
22	hCoV-19/Brazil/RJ-INCA-C06/2020|EPI_ISL_513513|2020-04-08	EPI_ISL_513513
23	hCoV-19/Brazil/RJ-DCVN1/2020|EPI_ISL_509434|2020-03-23	EPI_ISL_509434
24	hCoV-19/Brazil/RJ-899/2020|EPI_ISL_456071|2020-03-30	EPI_ISL_456071
25	hCoV-19/Brazil/RJ-2072/2020|EPI_ISL_456104|2020-04-16	EPI_ISL_456104
26	hCoV-19/Brazil/RJ-1921/2020|EPI_ISL_456091|2020-04-09	EPI_ISL_456091
27	hCoV-19/Brazil/RJ-1719/2020|EPI_ISL_456088|2020-04-06	EPI_ISL_456088
28	hCoV-19/Brazil/RJ-1702/2020|EPI_ISL_456087|2020-04-08	EPI_ISL_456087
29	hCoV-19/Brazil/RJ-1690/2020|EPI_ISL_456084|2020-04-08	EPI_ISL_456084
30	hCoV-19/Brazil/RJ-1627/2020|EPI_ISL_456083|2020-04-03	EPI_ISL_456083
31	hCoV-19/Brazil/RJ-1595/2020|EPI_ISL_467346|2020-04-02	EPI_ISL_467346
32	hCoV-19/Brazil/RJ-1555/2020|EPI_ISL_467344|2020-04-02	EPI_ISL_467344
33	hCoV-19/Brazil/GO-L19-CD410/2020|EPI_ISL_476333|2020-04-02	EPI_ISL_476333
34	hCoV-19/Brazil/PR-5620/2020|EPI_ISL_541343|2020-03-20	EPI_ISL_541343
35	hCoV-19/Brazil/SP-193/2020|EPI_ISL_523984|2020-04-19	EPI_ISL_523984
36	hCoV-19/Brazil/SP-240/2020|EPI_ISL_524468|2020-05-01	EPI_ISL_524468
37	hCoV-19/Brazil/PE-IAM1126/2020|EPI_ISL_572379|2020-05-07	EPI_ISL_572379
38	hCoV-19/Brazil/SC-L15-CD265/2020|EPI_ISL_476259|2020-03-29	EPI_ISL_476259
39	hCoV-19/Brazil/SP-L14-CD257/2020|EPI_ISL_476253|2020-03-27	EPI_ISL_476253
40	hCoV-19/Brazil/SC-0244/2020|EPI_ISL_470653|2020-03-18	EPI_ISL_470653
41	hCoV-19/Brazil/AL-837/2020|EPI_ISL_427292|2020-03-18	EPI_ISL_427292
42	hCoV-19/Brazil/SP-294/2020|EPI_ISL_527862|2020-03-30	EPI_ISL_527862
43	hCoV-19/Brazil/SP-L5-CAMPI91/2020|EPI_ISL_476419|2020-04-23	EPI_ISL_476419
44	hCoV-19/Brazil/SP-254/2020|EPI_ISL_524469|2020-05-05	EPI_ISL_524469
45	hCoV-19/Brazil/SP-436/2020|EPI_ISL_603033|2020-07-13	EPI_ISL_603033

46	hCoV-19/Brazil/SP-441/2020|EPI_ISL_603038|2020-07-20	EPI_ISL_603038
47	hCoV-19/Brazil/RS-0242/2020|EPI_ISL_470651|2020-03-20	EPI_ISL_470651
48	hCoV-19/Brazil/SP-356/2020|EPI_ISL_547575|2020-06-16	EPI_ISL_547575
49	hCoV-19/Brazil/SP-321/2020|EPI_ISL_534316|2020-05-02	EPI_ISL_534316
50	hCoV-19/South_Korea/KCDC2712/2020|EPI_ISL_522491|2020-07-11	EPI_ISL_522491
51	hCoV-19/Brazil/PR-5621/2020|EPI_ISL_541344|2020-03-19	EPI_ISL_541344
52	hCoV-19/Brazil/SP-437/2020|EPI_ISL_603034|2020-07-11	EPI_ISL_603034
53	hCoV-19/Brazil/SP-433/2020|EPI_ISL_603030|2020-07-11	EPI_ISL_603030
54	hCoV-19/Brazil/UN-HIAE-SP04/2020|EPI_ISL_486429|2020-03-20	EPI_ISL_486429
55	hCoV-19/Norway/3069/2020|EPI_ISL_549084|2020-08-10	EPI_ISL_549084
56	hCoV-19/Brazil/SP-370/2020|EPI_ISL_583500|2020-06-29	EPI_ISL_583500
57	hCoV-19/Brazil/RJ-UFRJ-9331/2020|EPI_ISL_492036|2020-06-01	EPI_ISL_492036
58	hCoV-19/Thailand/Bangkok-0071/2020|EPI_ISL_445380|2020-03-30	EPI_ISL_445380
59	hCoV-19/Brazil/SP-439/2020|EPI_ISL_603036|2020-07-10	EPI_ISL_603036
60	hCoV-19/Brazil/SP-440/2020|EPI_ISL_603037|2020-07-20	EPI_ISL_603037
61	hCoV-19/Brazil/SP-394/2020|EPI_ISL_583503|2020-06-22	EPI_ISL_583503
62	hCoV-19/Brazil/SP-398/2020|EPI_ISL_603028|2020-06-30	EPI_ISL_603028
63	hCoV-19/Brazil/SP-L5-CAMPI77/2020|EPI_ISL_476416|2020-04-23	EPI_ISL_476416
64	hCoV-19/Brazil/SP-283/2020|EPI_ISL_527860|2020-04-17	EPI_ISL_527860
65	hCoV-19/Ecuador/USFQ-004/2020|EPI_ISL_477014|2020-03-30	EPI_ISL_477014
66	hCoV-19/Ecuador/USFQ-020/2020|EPI_ISL_471267|2020-04-17	EPI_ISL_471267
67	hCoV-19/Ecuador/USFQ-039/2020|EPI_ISL_481245|2020-04-17	EPI_ISL_481245
68	hCoV-19/Ecuador/USFQ-1112/2020|EPI_ISL_486847|2020-06-30	EPI_ISL_486847
69	hCoV-19/Ecuador/USFQ-110/2020|EPI_ISL_486845|2020-06-30	EPI_ISL_486845
70	hCoV-19/Italy/APU-UniMI-123PT/2020|EPI_ISL_525554|2020-07-20	EPI_ISL_525554
71	hCoV-19/Brazil/GO-L19-CD413/2020|EPI_ISL_476336|2020-04-02	EPI_ISL_476336
72	hCoV-19/Brazil/MG-0288/2020|EPI_ISL_470593|2020-04-09	EPI_ISL_470593
73	hCoV-19/Brazil/RJ-0251/2020|EPI_ISL_470619|2020-03-27	EPI_ISL_470619
74	hCoV-19/Brazil/GO-0209/2020|EPI_ISL_470576|2020-04-03	EPI_ISL_470576
75	hCoV-19/Argentina/PAIS-A0023/2020|EPI_ISL_430814|2020-04-17	EPI_ISL_430814
76	hCoV-19/Luxembourg/LNS9470500/2020|EPI_ISL_434505|2020-04-08	EPI_ISL_434505
77	hCoV-19/Mexico/HID-InDRE-54/2020|EPI_ISL_576257|2020-07-22	EPI_ISL_576257
78	hCoV-19/Mexico/CMX-INMEGEN-02/2020|EPI_ISL_522873|2020-07-31	EPI_ISL_522873
79	hCoV-19/Peru/LIM-UPCH-0146/2020|EPI_ISL_568540|2020-08-28	EPI_ISL_568540
80	hCoV-19/Peru/LIM-INS-138/2020|EPI_ISL_536533|2020-03-17	EPI_ISL_536533
81	hCoV-19/Peru/LIM-INS-078/2020|EPI_ISL_536478|2020-07-02	EPI_ISL_536478
82	hCoV-19/Peru/LIM-UPCH-0160/2020|EPI_ISL_568553|2020-08-28	EPI_ISL_568553
83	hCoV-19/CotedIvoire/BKE0891/2020|EPI_ISL_614389|2020-08-18	EPI_ISL_614389
84	hCoV-19/Peru/LIM-INS-120/2020|EPI_ISL_536518|2020-05-05	EPI_ISL_536518
85	hCoV-19/Peru/LIM-UPCH-0138/2020|EPI_ISL_568534|2020-08-27	EPI_ISL_568534
86	hCoV-19/Turkey/KOC-IST-OD5/2020|EPI_ISL_613460|2020-06-21	EPI_ISL_613460
87	hCoV-19/USA/WI-UW-1054/2020|EPI_ISL_516480|2020-07-30	EPI_ISL_516480
88	hCoV-19/USA/WI-UW-759/2020|EPI_ISL_495464|2020-07-06	EPI_ISL_495464
89	hCoV-19/Peru/LIM-UPCH-0147/2020|EPI_ISL_568541|2020-08-28	EPI_ISL_568541
90	hCoV-19/USA/MI-MDHHS-SC22181/2020|EPI_ISL_614232|2020-10-12	EPI_ISL_614232
91	hCoV-19/USA/VA-DCLS-1506/2020|EPI_ISL_581508|2020-07-28	EPI_ISL_581508
92	hCoV-19/Peru/LIM-INS-100/2020|EPI_ISL_536499|2020-07-04	EPI_ISL_536499

93	hCoV-19/Mexico/CMX-INMEGEN-03/2020|EPI_ISL_522874|2020-07-31	EPI_ISL_522874
94	hCoV-19/Peru/LIM-UPCH-0127/2020|EPI_ISL_568523|2020-08-28	EPI_ISL_568523
95	hCoV-19/Peru/LIM-UPCH-0129/2020|EPI_ISL_568525|2020-08-28	EPI_ISL_568525
96	hCoV-19/Peru/LIM-UPCH-0128/2020|EPI_ISL_568524|2020-08-28	EPI_ISL_568524
97	hCoV-19/Romania/Mioveni-24095/2020|EPI_ISL_468156|2020-05-08	EPI_ISL_468156
98	hCoV-19/Mexico/TLA-InDRE-57/2020|EPI_ISL_576260|2020-08-13	EPI_ISL_576260
99	hCoV-19/Mexico/ZAC-InDRE-72/2020|EPI_ISL_576275|2020-08-14	EPI_ISL_576275
100	hCoV-19/France/BRE-BR9068/2020|EPI_ISL_613557|2020-09-06	EPI_ISL_613557
101	hCoV-19/Singapore/1110/2020|EPI_ISL_605819|2020-10-25	EPI_ISL_605819
102	hCoV-19/Netherlands/ZH-EMC-552/2020|EPI_ISL_577980|2020-09-08	EPI_ISL_577980
103	hCoV-19/Netherlands/ZH-EMC-607/2020|EPI_ISL_578035|2020-09-17	EPI_ISL_578035
104	hCoV-19/Brazil/SP-405/2020|EPI_ISL_547577|2020-07-06	EPI_ISL_547577
105	hCoV-19/Brazil/DF-0001/2020|EPI_ISL_426580|2020-03-13	EPI_ISL_426580
106	hCoV-19/Brazil/SP-345/2020|EPI_ISL_583495|2020-06-13	EPI_ISL_583495
107	hCoV-19/Brazil/PI-0239/2020|EPI_ISL_470613|2020-03-19	EPI_ISL_470613
108	hCoV-19/Brazil/RN-IEC-162277/2020|EPI_ISL_524798|2020-03-14	EPI_ISL_524798
109	hCoV-19/Brazil/RJ-INCA-C34/2020|EPI_ISL_513542|2020-04-16	EPI_ISL_513542
110	hCoV-19/Brazil/DF-891/2020|EPI_ISL_427298|2020-03-22	EPI_ISL_427298
111	hCoV-19/Brazil/DF-862/2020|EPI_ISL_427297|2020-03-23	EPI_ISL_427297
112	hCoV-19/Brazil/SP-399/2020|EPI_ISL_603029|2020-06-22	EPI_ISL_603029
113	hCoV-19/Ireland/D-NVRL-72IRL12139/2020|EPI_ISL_528465|2020-08-12	EPI_ISL_528465
114	hCoV-19/Brazil/MG-0291/2020|EPI_ISL_470596|2020-04-16	EPI_ISL_470596
115	hCoV-19/Brazil/RJ-00318/2020|EPI_ISL_623121|2020-05-12	EPI_ISL_623121
116	hCoV-19/Brazil/BA-L17-CD359/2020|EPI_ISL_476305|2020-03-31	EPI_ISL_476305
117	hCoV-19/Brazil/RJ-INCA-C181/2020|EPI_ISL_513519|2020-04-30	EPI_ISL_513519
118	hCoV-19/Brazil/AP-IEC-165669/2020|EPI_ISL_524793|2020-04-29	EPI_ISL_524793
119	hCoV-19/Brazil/RJ-0263/2020|EPI_ISL_470630|2020-04-13	EPI_ISL_470630
120	hCoV-19/Brazil/RJ-00364/2020|EPI_ISL_623167|2020-05-04	EPI_ISL_623167
121	hCoV-19/Brazil/RJ-UFRJ-58271/2020|EPI_ISL_492048|2020-06-01	EPI_ISL_492048
122	hCoV-19/Brazil/DF-615i/2020|EPI_ISL_427294|2020-03-13	EPI_ISL_427294
123	hCoV-19/Spain/GA-IBV-98006079/2020|EPI_ISL_541066|2020-07-03	EPI_ISL_541066
124	hCoV-19/Brazil/RJ-00316/2020|EPI_ISL_623119|2020-05-04	EPI_ISL_623119
125	hCoV-19/Argentina/PAIS-A0024/2020|EPI_ISL_430815|2020-04-18	EPI_ISL_430815
126	hCoV-19/Argentina/Heritas_HG001/2020|EPI_ISL_476496|2020-04-22	EPI_ISL_476496
127	hCoV-19/Argentina/Heritas_HG006/2020|EPI_ISL_476561|2020-05-07	EPI_ISL_476561
128	hCoV-19/Argentina/Heritas_HG007/2020|EPI_ISL_476565|2020-05-09	EPI_ISL_476565
129	hCoV-19/Argentina/Heritas-HG023/2020|EPI_ISL_615121|2020-05-26	EPI_ISL_615121
130	hCoV-19/Brazil/RJ-1466/2020|EPI_ISL_456081|2020-04-06	EPI_ISL_456081
131	hCoV-19/Brazil/RJ0272/2020|EPI_ISL_470638|2020-04-17	EPI_ISL_470638
132	hCoV-19/Brazil/RJ0256/2020|EPI_ISL_470624|2020-04-03	EPI_ISL_470624
133	hCoV-19/Brazil/RJ0254/2020|EPI_ISL_470622|2020-04-01	EPI_ISL_470622
134	hCoV-19/Brazil/RJ0251/2020|EPI_ISL_470619|2020-03-27	EPI_ISL_470619
135	hCoV-19/Brazil/RJ0248/2020|EPI_ISL_470616|2020-03-24	EPI_ISL_470616
136	hCoV-19/Brazil/RJ-0720/2020	EPI_ISL_636836
137	hCoV-19/Brazil/RJ-01020/2020	EPI_ISL_636834
138	hCoV-19/Brazil/RJ-06020/2020	EPI_ISL_636737
139	hCoV-19/Brazil/RJ-01020-2/2020	EPI_ISL_636835

140	hCoV-19/Brazil/RJ-0720.2/2020	EPI_ISL_63683
141	hCoV-19/Brazil/RJ-0920/2020	EPI_ISL_63683

715

716

Table S2 - Genetic characteristics of the SARS-CoV-2 sequences

Characteristics of SARS-CoV-2 sequences	**Patient B (first episode) (% quasispecies)**	**Patient C (first episode) (% quasispecies)**	**Patient A (% quasispecies)**	**Patient B (% quasispecies)**	**Patient C (% quasispecies)**	**Patient D (% quasispecies)**
Number of reads and phred quality score	>140K - Q35	>2.600K - Q36	>472K - Q33	>24.400K - Q36	>15.600K - Q36	>20.000K - Q36
Depth of Coverage, average ± SD	103.5 ± 1.24	2018.85 ± 31.53	350.85 ± 4.04	18426.125 ± 279.91	11753.53 ± 257.32	10061.58 ± 3701.20
Mutations						
NSP2 (AA)						
G1522A (none)	X					
NSP3 (AA)						
A6866G (N1383D)	X					
C3037T (none)	X	X (24.7)	X	X	X	X
C6021T (P1101L)	X (25.44)					
4151_4152delTT (L478X)		X				
C7164T (T1482I)		X (25.1)				
T7082A (S1455T)			X (25.3)			
A7384T (Q1555H)			X (25.3)			
NSP4 (AA)						
C9569T (P339S)			X (25.3)			
3C-like proteinase (AA)						
A10904G (S284G)		X (25.1)				
NSP6 (AA)						
C11514T (T181I)			X	X	X	X
RNA-dependent RNA Polymerase (AA)						
G15406T (A656S)	X					
C14408T (P323L)	X (74.6)	X	X	X	X	X
A14836G (I466V)	X (24.77)					
Helicase (AA)						
A17105G (H290R)			X (25.3)			
3'-5'-Exonuclease						
G18412T (V125F)			X	X	X	X
G18180T (K47N)						X
18180_18181insTG (K47_D48insX)						X
EndoRNAse						
A20265G (none)	X (25.44)					
2'-O-Ribose Methyltransferase						
A21415T (K53*)	X (24.76)					
Surface glycoprotein						
A23403G (D614G)	X	X	X	X	X	X
T22619G (W353G)			X (25.3)			
Membrane glycoprotein						
G26795T (M91I)	X					
G27112A (S197N)	X (25.01)					
A26555N (E11X)	X (25.44)					
G26556N /A26557N/ G26558N (E12X)	X (25.44)					
C26559N (L13X)	X (25.44)					

27184_27228del TACAGTAAGTGACAA CAGATGTTTCATCTCGTTGACTTTC AGGTT (V221X)				X (24.7)		
ORF6						
T27299C (133T)		X	X	X (75.3)	X	X
27184_27228delTACAGTAAGTGACA AC AGATGTTTCATCTCGTTGACTTTCA GGTT (M1_V9del)				X (24.7)		
27196_27228delCAACAGATGTTTCA TCTC GTTGACTTTCAGGTT (M1_V9del)					X (24.54)	
27197_27226delAACAGATGTTTCAT CTC GTTGACTTTCAGG (M1_Q8del)					X (25.56)	
27197_27226delAACAGATGTTTCAT CTC GTTGACTTTCAGG (V9X)					X (25.56)	
A27313G (K38E)						X
Nucleocapside						
G28881A/G28882G (R203K)		X	X	X	X	X
G28883C (G204R)		X	X	X	X	X
T29148C (I292T)		X	X	X	X	X

739