




# Genomic evidence of SARS-CoV-2 reinfection cases in southern Brazil

Ana Paula Muterle Varela<sup>1</sup> · Fernando Hayashi Sant'Anna<sup>1</sup> · Aníusca Vieira dos Santos<sup>1</sup> · Janira Prichula<sup>2</sup> · Juliana Comerlato<sup>1</sup> · Giovana Tavares dos Santos<sup>1</sup> · Eliana Wendland<sup>1,3</sup> 

Received: 29 June 2022 / Accepted: 13 October 2022 / Published online: 3 January 2023  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2022

## Abstract

Cases of reinfection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have been reported worldwide. We investigated reinfection cases in a set of more than 30,000 samples, and the SARS-CoV-2 genomes from selected samples from four patients with at least two positive diagnoses with an interval  $\geq 45$  days between tests were sequenced and analyzed. Comparative genomic and phylogenetic analysis confirmed three reinfection cases and suggested that the fourth one was caused by a virus of the same lineage. Viral sequencing is crucial for understanding the natural course of reinfections and for planning public health strategies for management of COVID-19.

## Introduction

During the COVID-19 pandemic, cases of reinfection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have been occurring worldwide [1–3]. According to the Centers for Disease Control and Prevention, reinfection is defined as at least two SARS-CoV-2 infection episodes caused by viruses of different lineages confirmed by sequencing or a second positive test result obtained using a specimen collected  $>90$  days after the first one, except in the case of severely immunocompromised individuals [4]. However, this definition might be changed when more information has been obtained.

Re-detection of viral RNA in patients has been associated with either reinfection or reactivation (incomplete virus

elimination leading to positive results after a negative test) [5, 6]. To distinguish between these possibilities, reinfection and/or reactivation should preferably be confirmed by viral genome sequencing to evaluate if the viruses causing primary and secondary infections are of different lineages [7, 8].

Understanding the natural course of reinfection and reactivation is pivotal for developing control strategies for COVID-19. We therefore used high-throughput sequencing of samples from four putative cases of COVID-19 reinfection from a database composed of more than 30,000 tests from a region of southern Brazil between March and November 2020, before the beginning of the national vaccination campaign. These genome sequences were analyzed to identify the mutational patterns of the viruses from the first and second infections and reconstruct their evolutionary history.

---

Handling Editor: Diego G. Diel.

---

Ana Paula Muterle Varela and Fernando Hayashi Sant'Anna contributed equally to this work.

---

✉ Eliana Wendland  
elianawend@gmail.com

- <sup>1</sup> Hospital Moinhos de Vento, PROADI-SUS, Porto Alegre, Brazil
- <sup>2</sup> Department of Microbiology and Ophthalmology, Harvard Medical School and Mass Eye and Ear Infirmary, Boston, USA
- <sup>3</sup> Department of Community Health, Federal University of Health Sciences of Porto Alegre (UFCSA), Porto Alegre, Brazil

## Materials and methods

The Clinical Epidemiology Laboratory (Epiclin) is located at the Federal University of Health Sciences of Porto Alegre (UFCSA) and funded by Associação Hospitalar Moinhos de Vento (AHMV). The Institutional Review Board of Moinhos de Vento Hospital approved this study under protocol number 32149620.9.0000.5330, following procedures from national guidelines for ethics committees. In the Epiclin, 31,973 tests for SARS-CoV-2 were carried out between April and December

2020. Naso- and oropharynx swab samples were obtained from four regions belonging to the Northeast and Metropolitan mesoregions of Rio Grande do Sul, southern Brazil. We identified (by name and date of birth) four symptomatic patients who had experienced at least two distinct episodes of COVID-19 with an interval of at least 45 days and tested positive based on amplification of the three viral genes that are targeted by the Allplex™ SARS-CoV-2 Assay (Seegene, South Korea) (Supplementary Fig. S1).

Sequencing libraries were constructed using a QIAseq SARS-CoV-2 Kit (QIAGEN, Hilden, Germany). Sequencing was performed using an Illumina MiSeq instrument and a MiSeq Reagent Kit v3 (600 cycles) (Illumina, USA) as described by Sant'Anna et al. [9].

A variant-calling pipeline, available at [https://github.com/fhsantanna/variant\\_calling\\_pipeline](https://github.com/fhsantanna/variant_calling_pipeline), was built using the Snakemake scheme. Reads were preprocessed using Fastp version 0.23.2, performing quality trimming and removing sequencing adapters. Subsequently, preprocessed reads were mapped to the SARS-CoV-2 archetype genome (MN908947) using BWA version 0.7.17-r1188. Next, optical duplicates were removed using MarkDuplicates from GATK package version 4.2.0.0. Deduplicated reads were then trimmed for ARTIC primers using Ivar version 1.3.1. Trimmed reads were assembled using Samtools version 1.15, and variants were called using Ivar with a minimum read depth of 5. Variants were filtered using Ivar with default parameters. Finally, consensus sequences were generated using Ivar, with a minimum read depth of 5. Consensus sequences were evaluated using the COVID-19 genome annotator (<http://giorgilab.unibo.it/coronannotator/>). Mutations that were not found in the variant-calling step were removed manually from the consensus sequence.

For phylogenetic reconstruction, all available SARS-CoV-2 sequences from the year 2020 from the cities where the putative reinfection cases occurred were used as references. These sequences were downloaded from GISAID in July 2021 and analyzed together with the genome sequences from this study. The sequence Wuhan/2019 was also included in the dataset. Sequences were aligned using MAFFT version 7.480 with the FFT-NS-2 option with default parameters. A phylogenetic tree was built using IQ-Tree version 2.1.4-beta, with the best-fit model determined automatically (option -m MFP) and ultrafast bootstrapping with 1,000 replicates. Next, a timetree inference and a "mugration" model using discrete PANGO lineages were conducted using Treetime version 0.8.1. Subsequently, the tree was plotted using a script written in R (ggtree library), coloring the branches according to the PANGO lineages.

## Results and discussion

Among the 31,973 samples in our collection, we found 11 putative reinfection cases, and in four of these cases, all of the viral targets were amplified in the diagnostic assay. This is consistent with previous studies showing reinfection to be a rare event within a short timeframe (~one year) [10–12].

The four patients were between 33 and 76 years old, symptomatic, and unvaccinated for SARS-CoV-2, and three of them were male (Table 1). The first and last samples from patients P1, P2, and P3 were collected approximately three months apart (September and November 2020). The four samples from patient P4 were collected on 5, 18, and 25 May and on 2 July 2020 (~60 days between the first and fourth collections). The third sample from P4 (25 May 2020) gave a negative result in the RT-PCR test.

We sequenced all nine of the positive samples, but the data from one of them was of insufficient quality for further analysis. We conducted a comparative analysis of eight genome sequences. For generating the consensus sequences of the viral genomes, 3276 to 1,662,544 reads were mapped to the reference sequence (Table 1). The breadth of coverage ranged from 37.2% to 99.98%, and the mean depth of coverage ranged from 10.61- to 6791.66-fold (Table 1).

The timetree shown in Figure 1 presents the relationships between the episodes of infection in a temporal framework and highlights their lineage groups. Two of the four cases (P1 and P3) exhibited infections with distinct lineages: B.1.1.28 and B.1.1.33 in the first (sample 1, S1) and second infection (S2), respectively. For patient P2, the sequence obtained from sample S1 grouped with the B.1.1.28 lineage, but the S2 sequence did not group with the B.1.1.28 or B.1.1.33 lineage. In the fourth case (P4), three positive samples (S1, S2, and S4) were sequenced, and two sequences (S1 and S4) were of sufficient quality for phylogenetic reconstruction. Both sequences clustered within the B.1.1.33 lineage. The period from March to November 2020 represents the first wave of the COVID-19 pandemic in southern Brazil. During that time, B.1.1.33 and B.1.1.28, descendants of the B.1 lineage, were the main circulating lineages [9, 13, 14].

Considering only regions covered in both viral genomes, the comparison between sequences from samples S1 and S2 of patients P1, P2, and P3, showed a total of two, seven, and 12 distinct nucleotide substitutions, and they shared one, two, and four nucleotide (nt) substitutions, respectively (Fig. 2).

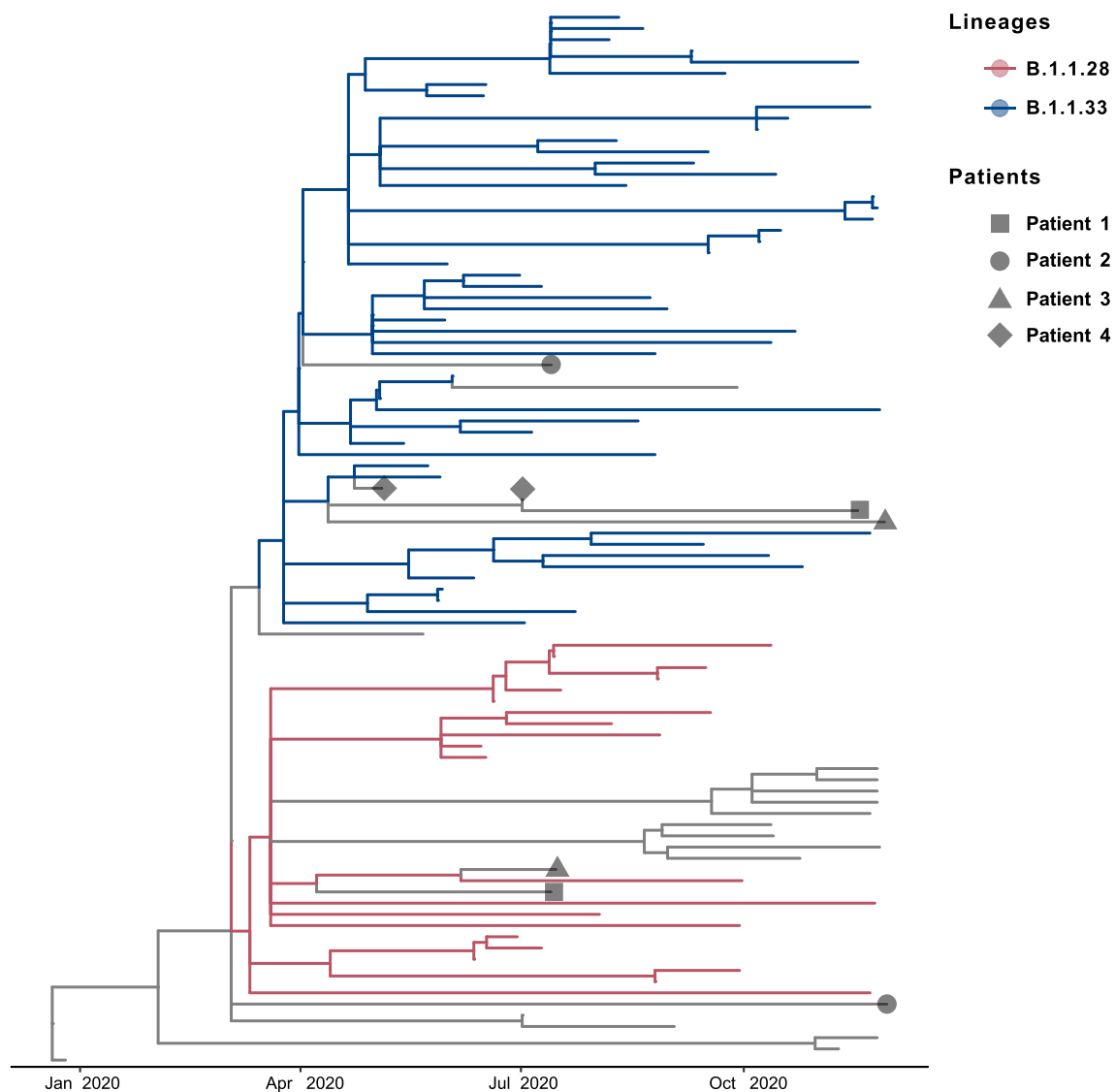
Some of these substitutions resulted in amino acid (aa) changes in the encoded viral proteins. From patient P1, sample S1 contained the substitution ORF8:R115L.

**Table 1** Data for samples included in this study

Patient	Sample (S)	ID	Sex	Age	Start date of symptoms*	Sample collection date	Test result	E	C <sub>t</sub> value	RdRP	N	Mean depth of coverage	No. of mapped reads	Breadth of coverage (%)	Accession number/source
P1	S1	7063CVL	M	76	7/10/2020	7/14/2020	2019-nCoV Detected	26.98	29.14	30	1451.95	363843	99.14	EPI_ ISL_13051164	
	S2	31600CVL			11/9/2020	11/18/2020	2019-nCoV Detected	39.03	37.56	37.11	10.61	3276	37.2	Supplementary information#	
P2	S1	7212CVL	M	56	7/8/2020	7/14/2020	2019-nCoV Detected	28.42	30.27	31.4	1112.76	284647	98.31	EPI_ ISL_13051165	
	S2	34661CVL			11/26/2020	11/30/2020	2019-nCoV Detected	33.62	33.56	33.15	49.59	35371	82.3	Supplementary information#	
P3	S1	7591CVL	M	33	7/11/2020	7/16/2020	2019-nCoV Detected	20.58	21.95	24.37	6791.66	1662544	99.98	EPI_ ISL_13051166	
	S2	34712CVL			11/29/2020	11/29/2020	2019-nCoV Detected	37.12	38.83	38.07	714.56	185474	88.83	EPI_ ISL_13051162	
P4	S1	0288CVL	F	60	4/29/2020	5/5/2020	2019-nCoV Detected	27.6	30.14	31.79	5464.01	1327513	99.90	EPI_ ISL_13051161	
	S2	0548CVL			5/17/2020	5/18/2020	2019-nCoV Detected	33.21	34.75	36.02	NA	NA	NA	NA	
	S3	0839CVL			5/21/2020	5/25/2020	Negative	NA	NA	NA	NA	NA	NA	NA	NA
	S4	4699CVL			7/1/2020	7/2/2020	2019-nCoV Detected	32.82	35.44	35.32	121.68	31801	91.93	EPI_ ISL_13051163	

\*Self-reported date, registered in GAL (Gerenciador de Ambiente Laboratorial - Laboratory Environment Manager)

#Sequences displayed quality too low for GISAID and GenBank submission. These sequences are available in FASTA format in Supplementary Information.



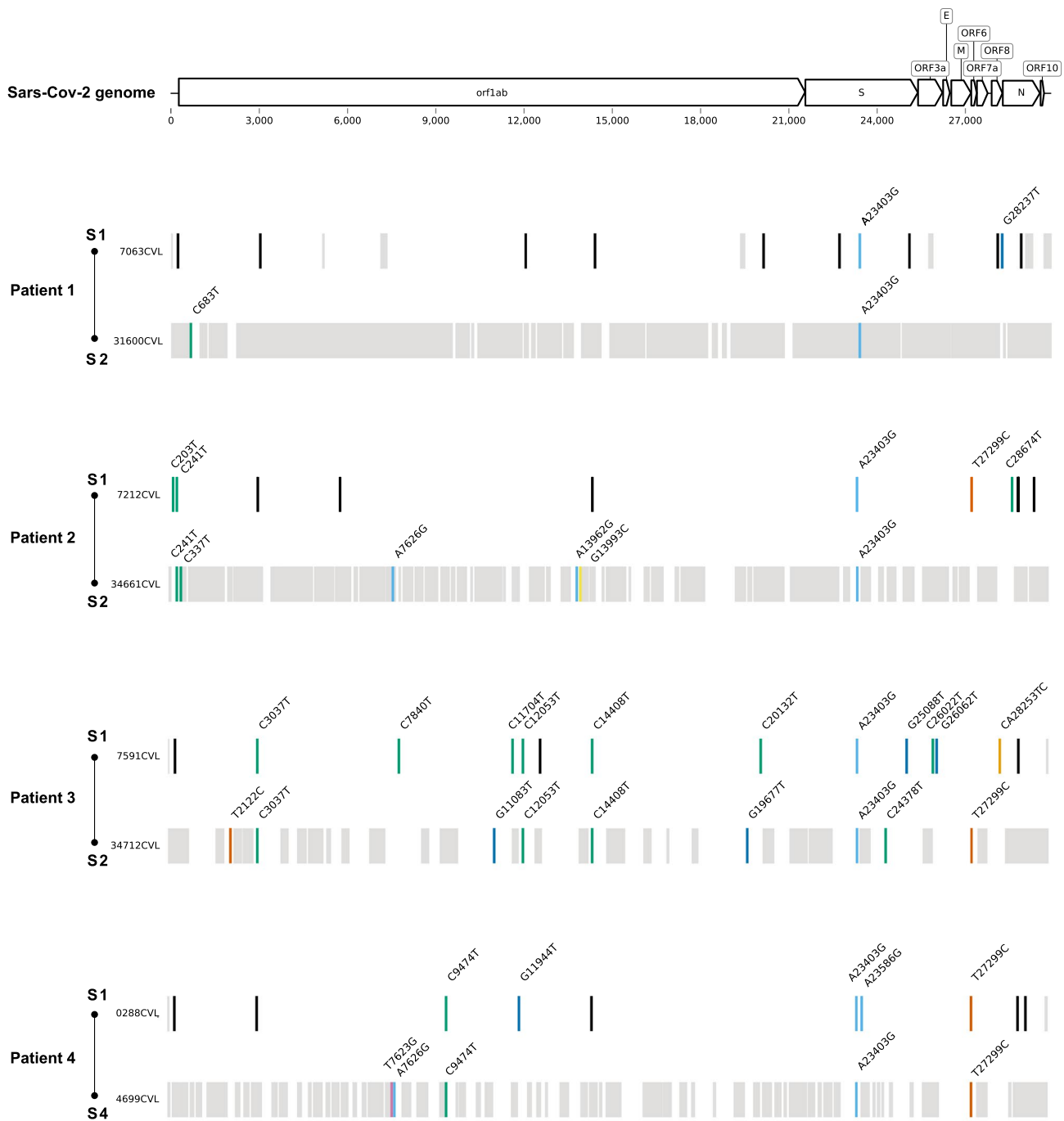
**Fig. 1** Time-scaled phylogenetic tree of SARS-CoV-2 reinfection cases. Geometrical shapes indicate the patients who were reinfected (legend). Internal branches are colored according to the most probable PANGO lineage of the last common ancestor of the clade (legend).

From P2, S1 contained ORF6:I33T and N:A134V and S2 contained NSP3:N1636S and NSP12b:A176P. From P3, S1 contained NSP15:A171V, S:V1176F, ORF3a:G224C, and ORF8:I121L, and S2 contained NSP6:L37F, NSP15:Q19H, S:S939F, and ORF6:I33T.

The sequences of the first (S1) and fourth (S4) samples from patient P4 showed four distinct nt substitutions, resulting in the aa changes NSP7:Q34H, S:Q675R, NSP3:V1635G, and NSP3:N1636S (Fig. 2). The phylogenetic reconstruction showed that these sequences were in the outermost clade of the B.1.1.33 lineage. The S3 sample from this patient was negative, suggesting that patient P4 was reinfected after that sample was collected. However,

we cannot rule out a reactivation, since we did not have access to the clinical data and outcomes of the patients, and both sequences clustered in the same lineage [4, 6]. Another limitation of this study is the low quality of two of the eight genome sequences determined, which were not accepted in a genomic database (GISAD and GenBank; Supplementary Information). The sequence quality can be associated with a low viral copy number, as suggested by high cycle threshold ( $C_t$ ) values.

An individual's immune response might be deficient in episodes of reinfection, allowing the second virus to evade the limited and transitory immunity induced by the primary infection [1, 15, 16]. Decreased immunity to SARS-CoV-2



**Fig. 2** Mutation profile of the genome sequences of each patient's samples. Nucleotide substitutions are represented by colored bars. Black bars indicate regions not covered in both genome sequences,

while grey bars indicate positions not included in the consensus sequence due to low coverage depth.

over time selects immune escape variants, which can be a threat to vaccine strategies [7, 17, 18].

Here, we describe three cases of confirmed reinfection and one case of suspected reinfection in southern Brazil

using genomic and phylogenetic analysis. Genomic monitoring is a crucial tool for differentiating between reactivation and reinfection and for providing information to guide public health systems.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00705-022-05648-8>.

**Acknowledgements** We thank Dr. Vanessa Mattevi, Dr. Marília Zandoná, and Tiago Fetzner for providing technical support.

**Funding** This study was financed by Hospital Moinhos de Vento through the Program for Supporting the Institutional Development of the Public Health System (PROADI-SUS), supported by the Ministry of Health of Brazil [grant number 25000.012591/2021-21].

**Data deposition** SARS-CoV-2 sequences determined in this study were submitted to the GISAID database under the accession numbers shown in Table 1.

## Declarations

**Conflict of interest** The authors declare no conflicts of interest.

**Ethical approval** The authors confirm that no ethical approval was required.

## References

- Gularte JS, da Silva MS, Demoliner M et al (2021) Reinfection cases by closely related SARS-CoV-2 lineages in Southern Brazil. *Braz J Microbiol* 52:1881–1885
- Gupta V, Bhojar RC, Jain A et al (2021) Asymptomatic reinfection in 2 healthcare workers from India with genetically distinct severe acute respiratory syndrome coronavirus 2. *Clin Infect Dis* 73:e2823–e2825
- To KK-W, Hung IF-N, Ip JD et al (2021) Coronavirus disease 2019 (COVID-19) re-infection by a phylogenetically distinct severe acute respiratory syndrome coronavirus 2 strain confirmed by whole genome sequencing. *Clin Infect Dis* 73:e2946–e2951
- Centers for Disease Control and Prevention (2021) Coronavirus disease 2019 (COVID-19) case definition. <https://ndc.services.cdc.gov/case-definitions/coronavirus-disease-2019-2021/>. Accessed 24 Jun 2022
- Dao TL, Hoang VT, Gautret P (2020) Recurrence of SARS-CoV-2 viral RNA in recovered COVID-19 patients: a narrative review. *Eur J Clin Microbiol Infect Dis* 40:13–25
- Tang X, Musa SS, Zhao S, He D (2021) Reinfection or reactivation of severe acute respiratory syndrome coronavirus 2: a systematic review. *Front Public Health* 9:663045
- European Centre for Disease Prevention and Control (2020) Reinfection with SARS-CoV-2: considerations for public health response. In: European Centre for Disease Prevention and Control. <https://www.ecdc.europa.eu/en/publications-data/threat-assessment-brief-reinfection-sars-cov-2>. Accessed 3 Jun 2022
- Pan American Health Organization Interim guidelines for detecting cases of reinfection by SARS-CoV-2. <https://www.paho.org/en/documents/interim-guidelines-detecting-cases-reinfection-sars-cov-2>. Accessed 3 Jun 2022
- Sant’Anna FH, Muterle Varela AP, Prichula J et al (2021) Emergence of the novel SARS-CoV-2 lineage VUI-NP13L and massive spread of P.2 in South Brazil. *Emerg Microbes Infect* 10:1431–1440
- Murchu EO, Byrne P, Carty PG et al (2022) Quantifying the risk of SARS-CoV-2 reinfection over time. *Rev Med Virol* 32:e2260
- Mao Y, Wang W, Ma J et al (2021) Reinfection rates among patients previously infected by SARS-CoV-2: systematic review and meta-analysis. *Chin Med J* 135:145–152
- Massachi J, Donohue KC, Kelly JD (2021) Severe acute respiratory syndrome coronavirus 2 reinfection cases corroborated by sequencing. *Am J Trop Med Hyg* 105:884–889
- Varela APM, Prichula J, Mayer FQ et al (2021) SARS-CoV-2 introduction and lineage dynamics across three epidemic peaks in Southern Brazil: massive spread of P.1. *Infect Genet Evol* 96:105144
- Franceschi VB, Caldana GD, de Menezes MA et al (2021) Genomic Epidemiology of SARS-CoV-2 in Esteio, Rio Grande do Sul, Brazil. *BMC Genom*. <https://doi.org/10.1101/2021.01.21.21249906>
- Harrington D, Kele B, Pereira S et al (2021) Confirmed reinfection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Variant VOC-202012/01. *Clin Infect Dis* 73:1946–1947
- Tillett RL, Sevinsky JR, Hartley PD et al (2021) Genomic evidence for reinfection with SARS-CoV-2: a case study. *Lancet Infect Dis* 21:52–58
- Jain VK, Iyengar K, Garg R, Vaishya R (2021) Elucidating reasons of COVID-19 re-infection and its management strategies. *Diabetes Metab Syndr* 15:1001–1006
- Reynolds SL, Kaufman HW, Meyer WA et al (2022) Duration of protection against SARS-CoV-2 reinfection and associated risk of reinfection assessed with real-world data. *medRxiv*. <https://doi.org/10.1101/2022.02.25.22271515>

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