Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

International Journal of Infectious Diseases 122 (2022) 444–448

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

Short Communication

SARS-CoV-2 intra-host evolution during prolonged infection in an immunocompromised patient

Erika Giorgia Quaranta[1,*], Alice Fusaro[1], Edoardo Giussani[1], Valeria D'Amico[1], Maria Varotto[1], Matteo Pagliari[1], Maria Teresa Giordani[2], Maira Zoppelletto[3], Francesca Merola[3], Antonio Antico[3], Paola Stefanelli[4], Calogero Terregino[1], Isabella Monne[1]

[1] Division of Comparative Biomedical Sciences, Istituto Zooprofilattico Sperimentale delle Venezie, Legnaro, Padova, Italy
[2] Infectious Diseases Unit, Alto Vicentino Hospital, ULSS 7 Pedemontana, Santorso, Vicenza, Italy
[3] San Bassiano Hospital, ULSS 7 Pedemontana, Bassano del Grappa, Vicenza, Italy
[4] Department of Infectious Diseases, Istituto Superiore di Sanità, Roma, Italy

ARTICLE INFO

Article history:
Received 1 April 2022
Revised 9 June 2022
Accepted 14 June 2022

Keywords:
SARS-CoV-2
Immunocompromised patient
Intra-host evolution
Prolonged infection
Sequencing

ABSTRACT

Objectives: Intra-host SARS-CoV-2 evolution during chronic infection in immunocompromised hosts has been suggested as being the possible trigger of the emergence of new variants. *Methods:* Using a deep sequencing approach, we investigated the SARS-CoV-2 intra-host genetic evolution in a patient with HIV over a period of 109 days. *Results:* Sequencing of nasopharyngeal swabs at three time points demonstrated dynamic changes in the viral population, with the emergence of 26 amino acid mutations and two deletions, 57% of them in the Spike protein. Such a combination of mutations has never been observed in other SARS-CoV-2 lineages detected so far. *Conclusion:* Our data confirm that persistent infection in certain immunocompromised individuals for a long time may favor the dangerous emergence of new SARS-CoV-2 variants with immune evasion properties.

Introduction

During the SARS-CoV-2 pandemic, several cases of prolonged infections were reported in immunosuppressed patients (Dolan et al., 2021). Most of these cases revealed an intra-host viral evolution, which allows the virus to accumulate mutations faster than during normal inter-host transmission (Avanzato et al., 2020; Leung et al., 2022). During these prolonged infections, SARS-CoV-2 can acquire mutations at key epitopes in the Spike (S) protein, potentially affecting virus replication, infectivity, and antigenicity, which are common to the variants of concern (VOCs) (European Centre for Disease Prevention and Control, 2022). Although the mechanisms underlying the emergence of new variants are still widely discussed, accelerated viral evolution in immunocompromised patients seems to be involved in the genesis of VOCs, as suggested for the Alpha variant (Hill et al., 2022.). This may also have been the case for the more recent Omicron variant, whose excess number of mutations in the S gene suggests it may have originated in a patient with immunocompromised conditions chronically infected with SARS-CoV-2 (Ma et al., 2022).

In this study, we performed an in-depth analysis of the complete genome of SARS-CoV-2 collected at three different time points from a patient with HIV over a period of 109 days.

Methods

Case description

A patient was diagnosed HIV-1 positive with AIDS and wasting syndrome in October 2006. Antiretroviral therapy (ART) was immediately started with azidothymidine (AZT) + 3TC + nevirapine when the clinical picture was characterized by a CD4 count of 52 cells/ml and an HIV-1 load (VL) of 83,157 cv/ml. In March 2007, the CD4 count was 184 cells/ml, and VL was 141 cv/ml, but after a few months, the patient interrupted the therapy, and the CD4 count dropped to 5 cells/ml, with a VL greater than 500,000 on September 12, 2007. The patient restarted and interrupted

* Corresponding author: Erika Giorgia Quaranta, Laboratory for Viral Genomics and Trascriptomics, Istituto Zooprofilattico Sperimentale delle Venezie (IZSVe), Viale dell'Università 10, 35020 Legnaro (Padova), Italy.
E-mail address: equaranta@izsvenezie.it (E.G. Quaranta).

https://doi.org/10.1016/j.ijid.2022.06.023

AZT + Abacavir + Lopinavir treatment several times from 2007 to 2020. On November 5, 2020, the patient tested positive for SARS-CoV-2 by real-time reverse transcription–PCR (rRT-PCR) (Allplex™ SARS-CoV-2 Assay, Seegene Inc., Seoul, Korea) and the nasopharyngeal swabs collected on 15, 22, and 30 November 2020 still confirmed the presence of SARS-CoV-2 infection. During this time, the patient was not under any ART therapy, and clinical data regarding this period are lacking. On June 5, 2021 (time point one), the patient went to the hospital with a mild fever and watery diarrhea that had persisted for a few weeks, with a weightloss of 20 kg. Once again, the patient tested positive by rRT-PCR for SARS-CoV-2, and blood tests demonstrated a very low CD4 count (3 cells/ml) and a very high HIV-1 load of 558,000 cv/ml. ART was immediately restarted with Dolutegravir + Abacavir + Lamivudine, and soon after, the patient was discharged on a domestic isolation regimen. The patient received two doses of SARS-CoV-2 mRNA (Pfizer) Comirnaty* 195 FL multidose vaccine, the first on June 3, 2021, and the second on July 9, 2021. Testing for SARS-CoV-2 was performed on June 28, 2021 (time point two), and on September 22, 2021 (time point three), with the patient still resulting positive.

The patient was treated with monoclonal antibodies (Bamlanizumab 700 mg + Etesivimab 1400 mg ev) on September 30, 2021.

RNA extraction, next-generation sequencing, and bioinformatics analysis

RNA purification was performed on the three nasopharyngeal swabs using the QIAamp Viral RNA kit (Qiagen) according to the manufacturer's protocol. The RNA from each sample was reverse transcribed using the LunaScript RT SuperMix (New England Biolabs), and 5 μL of cDNA was amplified with the ARTIC primer set (ARTIC nCoV-2019 V3 Panel, Integrated DNA Technologies) and the Q5 Hot Start High Fidelity DNA Polymerase (New England Biolabs). The following conditions were used: 98°C for 30 seconds, 95°C for 15 seconds, and 63°C for 5 minutes (35-40 cycles), with a 4°C holding step. Amplicons were purified using 0.8x AMPure XP beads (Beckman Coulter) and quantified using the Qubit dsDNA HS Assay kit (ThermoFisher Scientific). Libraries were prepared with the Illumina DNA Prep kit (Illumina) according to the manufacturer's instructions, and quality was checked using the Bioanalyzer High Sensitivity DNA kit (Agilent Technologies). Libraries were sequenced on Illumina platforms in 150 paired-end mode. Read quality was assessed using FastQC (Andrews, 2010). Cutadapt (Martin, 2011) was employed to perform quality filtering and adapter trimming. Reads were aligned against the Wuhan-Hu-1 SARS-CoV-2 genome (GenBank accession number **NC_045512.2**) using BWA-MEM (Li, 2013). iVar (Grubaugh et al., 2019) was used to remove PCR primers. GATK4 (Van der Auwera et al., 2020) was used to correct potential errors, improve alignment and recalibrate base quality. Variants call was performed with LoFreq* (Wilm et al., 2012). Consensus sequences were produced with a script developed in-house. Single nucleotide polymorphisms and indels with a frequency >50% were reported in the consensus, whereas positions with a coverage <10 were masked with the letter N, which indicates any base according to the International Union of Pure and Applied Chemistry annotation.

Results

SARS-CoV-2 sequencing was performed on nasopharyngeal swabs collected from the patient on time points one (June 5, 2021), two (June 28, 2021), and three (September 22, 2021). Unfortunately, they were the only samples available for sequencing at the time of writing. The raw data are available on NCBI Sequence Read Archive with accession number PRJNA837407 and consensus sequences were deposited in the Global Initiative on Sharing All Influenza Data under the following ID: **EPI_ISL_2927997** (time point one), **EPI_ISL_3006795** (time point two) and **EPI_ISL_4968925** (time point three). All the samples produced high-quality data, yielding the near-complete genome. Specifically, we obtained 1,784,831 SARS-CoV-2 reads for time point one, 1,608,528 for time point two, and 3,223,925 for time point three, which allowed us to produce respectively 99.73%, 99.46%, and 99.67% of the genome, with an average coverage of 7,806X for time point one, 6,599X for time point two and 14,140X for time point three (Figure 1A). These sequences formed a monophyletic group within the AH.3 lineage (Pangolin v4.0.6, pangolin-data version v1.8, O'Toole et al., 2021), a lineage that has been circulating in a few European countries since the second half of October 2020 (Figure 1C). They showed 24-26 amino acid substitutions compared with the Wuhan-Hu-1 reference sequence (GenBank accession number **NC_045512.2**), 10 of which are distinctive of the AH.3 lineage (Figure 1B). Over 109 days, 35 nucleotide positions underwent mutations, 26 resulting in amino acid changes, and two different deletions appeared (Figure 2). Among protein-coding changes, eight were observed in Open Reading Frame 1ab (ORF1ab), 16 in S protein (14 mutations and two deletions), one in ORF3a, one in Membrane (M) protein, and two in Nucleocapsid (N) protein. Similarly to what is reported in other studies (Leung et al., 2022; Weigang et al., 2021), several amino acid changes arising from such intra-host viral evolution have also been observed in VOCs and/or have been associated with effects on viral fitness. The **S13I** mutation in the S protein, for example, is typical of the Epsilon variant and is responsible for a marked reduction in the binding of N-terminal domain-directed neutralizing antibodies *in vitro* (McCallum et al., 2021). Similarly, **del144**, peculiar to the Alpha variant, is believed to ease antibody resistance (Wang et al., 2021). Although **del242/246** has neither been associated with any VOC nor with other particular biological effects, it should be highlighted that deletions in this region of the S protein have been associated with the alteration of the epitope structure and consequently a reduced immunoreactivity (Klinakis et al., 2021). Remarkably, at residue E484, we observed the appearance of **E484A** mutation at time point two and **E484K** at time point three, although **E484A** continued to be present at a lower frequency (20%) (Figure 1D). E484 is a crucial residue, as substitutions at this position have been noticed in many rapidly spreading VOCs (Laurini et al., 2021) and are associated with reduced sensitivity to neutralization by convalescent human sera (Liu et al., 2021). In addition, it is worth mentioning that the **E484A** mutation, typical of the Omicron variant, detected in the sample collected on June 28, 2021 (time point two), had never been observed in Italy. Two other mutations in the receptor-binding domain, **F490L,** and **S494P**, are respectively linked to a reduced neutralization sensitivity and an increased binding to angiotensin-converting enzyme 2, thereby increasing overall viral virulence (Chakraborty, 2021; Li et al., 2020). Focusing on the 37 nucleotide positions where the three consensus sequences differed (35 mutations and two deletions), we found that 95% of the observed mutations/deletions were present as minority variants even in samples where they were not detected at the consensus level (Figure 2). For example, mutations **F490L, S494P,** and **Q677K** in the S protein were present with a frequency of 89.70-89.79% at time point one, then dropped to a frequency of 16.43-18.51% at time point two and returned to be the prevalent mutation (with a frequency of 63.57-89.02%) at time point three (Figure 1D). This observation suggests that a mixed population was present in the patient as early as time point one. Unfortunately, the lack of sequences from previous samples makes it impossible to determine whether these multiple variants may have resulted from pre-

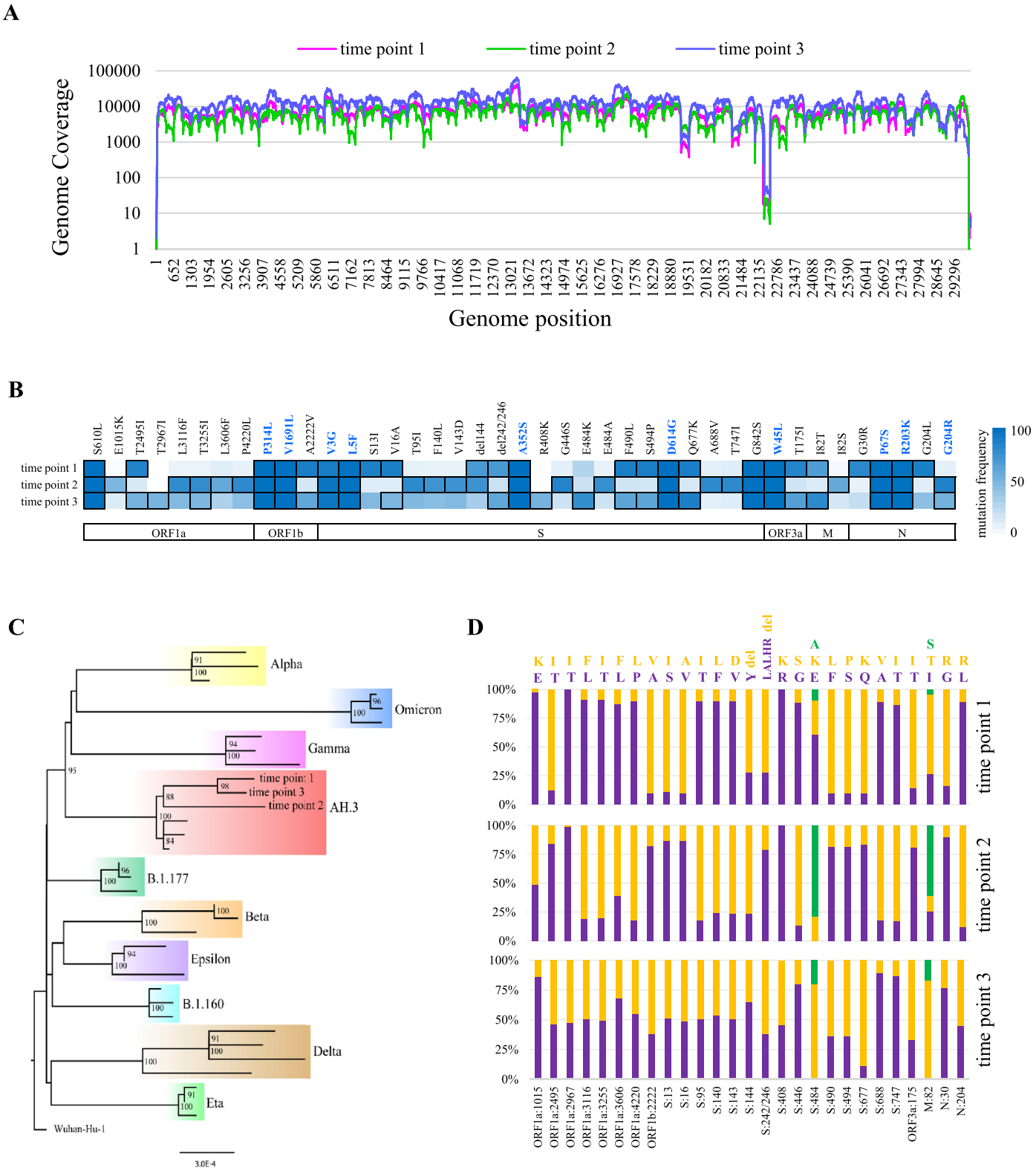

Figure 1. SARS-CoV-2 mutation accumulation during chronic infection in a patient with HIV over 109 days. **(A)** Genome coverage profiles for the three nasopharyngeal swabs analyzed in this study. Y-axis is in logarithmic scale. **(B)** Schematic representation of amino acid mutations found in patient swabs compared with the Wuhan-Hu-1 reference sequence (GenBank accession number **NC_045512.2**). The heatmap summarizes the variant frequencies; bordered squares represent consensus mutations (frequency >50%), and the mutations written in blue are those typical of the AH.3 lineage. **(C)** Phylogenetic analysis of the sequences obtained from patient swabs at the three time points. The sequences were aligned to a set of representative SARS-CoV-2 genome sequences belonging to AH.3 lineage and the main lineages and VOCs identified so far. The maximum-likelihood phylogenetic tree was constructed with IQ-TREE (GTR+F+R2) and rooted on the Wuhan-Hu-1 reference genome. Ultra-fast bootstrap supports are indicated above the nodes. **(D)** Histograms representing the amino acid frequency at the positions that differ between the three time points. Amino acids are indicated above the columns for each position. VOCs = variants of concern.

Genome position	Reference base	Alt base	time point 1	time point 2	time point 3	amino acid change
346	C	T	9.13%	71.97%	20.55%	
3308	G	A	2.58%	51.07%	13.81%	ORF1a:E1015K
5144	C	T	89.01%	14.24%	68.57%	
7531	A	G	9.15%	82.13%	41.59%	
7749	C	T	87.14%	15.95%	54.12%	ORF1a:T2495I
9165	C	T	0.00%	0.00%	52.86%	ORF1a:T2967I
9611	C	T	8.75%	80.81%	49.57%	ORF1a:L3116F
9857	C	T	85.39%	17.62%	47.79%	
10029	C	T	8.75%	80.09%	50.84%	ORF1a:T3255I
11083	G	T	12.30%	60.51%	32.28%	ORF1a:L3606F
12484	C	T	6.28%	75.20%	43.89%	
12924	C	T	10.03%	82.31%	45.41%	ORF1a:P4220L
14292	C	T	9.27%	81.14%	45.20%	
16887	C	T	91.28%	16.56%	49.88%	
20132	C	T	90.15%	17.46%	62.16%	ORF1b:A2222V
21600	G	T	88.58%	13.12%	49.00%	S:S13I
21609	T	C	89.76%	12.96%	51.35%	S:V16A
21846	C	T	10.22%	81.87%	49.65%	S:T95I
21980	T	C	9.67%	75.37%	46.42%	S:F140L
21989	GTTT	G	9.70%	76.35%	49.40%	S:V143D + S:del144
21990	TTTA	T	71.73%	9.34%	35.05%	S:del144
22284	TACTTGCTTTACATAG	T	71.72%	21.01%	61.82%	S:del242/246
22785	G	A	0.00%	0.00%	54.23%	S:R408K
22898	G	A	11.01%	86.36%	20.15%	S:G446S
23012	G	A	30.15%	21.17%	79.70%	S:E484K
23013	A	C	9.10%	78.83%	20.15%	S:E484A
23030	T	C	89.79%	18.51%	63.57%	S:F490L
23042	T	C	89.77%	18.17%	63.88%	S:S494P
23591	C	A	89.70%	16.43%	89.02%	S:Q677K
23625	C	T	10.69%	82.14%	10.61%	S:A688V
23802	C	T	13.09%	82.39%	13.09%	S:T747I
25916	C	T	85.58%	18.75%	67.08%	ORF3a:T175I
26767	T	C	68.99%	13.75%	82.93%	M:I82T
26767	T	G	4.23%	60.63%	16.93%	M:I82S
28271	A	C	17.75%	87.15%	73.16%	
28273	A	T	77.43%	11.70%	20.89%	
28361	G	C	83.35%	10.20%	23.42%	N:G30R
28884	G	T	89.22%	12.07%	45.08%	N:R204L

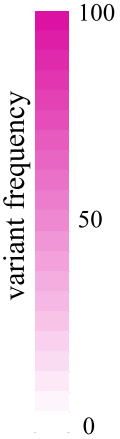

Figure 2. Genome variants identified in nasopharyngeal swabs collected from the patient at three time points in comparison to the Wuhan-Hu-1 reference sequence (GenBank accession number **NC_045512.2**) in the nucleotide positions that differed between the three consensus sequences. When present, amino acid mutations are indicated. The rectangle marks two different mutations present at the same nucleotide position (26,767; Membrane protein).

existing co-infection or intra-host virus evolution of a single virus strain.

Conclusion

Through the ultra-deep sequencing of the SARS-CoV-2 population at multiple time points, we demonstrated that a highly divergent variant could rapidly emerge in an HIV-positive patient with SARS-CoV-2 persistent infection. In less than four months, the virus accumulated 26 amino acid mutations and two deletions, 57% of which were located in the S protein. Many of these mutations are involved in SARS-CoV-2 ability to evade recognition by the immune system and are indeed found in the major VOCs. The depletion of CD4 unfitted the patient for clearing SARS-CoV-2. This case, added to the ones previously reported (Avanzato et al., 2020; Leung et al., 2022), highlights the crucial role of SARS-CoV-2-positive patients with HIV and low CD4 count as possible sources of new variants.

Declaration of Competing Interest

The authors have no competing interests to declare.

Funding

This study was partially funded by the Italian Ministry of Health through the project CCM 2020 - Title: “Caratterizzazione molecolare del virus pandemico SARS-CoV-2 in Italia” and the Ricerca Corrente Istituto Zooprofilattico Sperimentale delle Venezie 10/19. Part of the genomic monitoring which led to these results has also been supported by the Resolution of the Regional Council of Veneto n. 805 on June 22, 2021.

Acknowledgments

The authors thank all researchers who have shared genome data openly through the Global Initiative on Sharing All Influenza Data (GISAID). The authors would like to thank the technical and

scientific representatives of the Ministry of Health: Dr. Stefania D'Amato and Dr. Michela Sabbatucci. We extend our sincere appreciation to all the operators of the Unità Locale Socio Sanitaria 7 Pedemontana P.O. Bassano del Grappa engaged in the sampling and initial processing of specimens. We also thank Ms. Francesca Ellero from the Istituto Zooprofilattico Sperimentale delle Venezie for editing the manuscript.

Ethical approval

Informed consent was obtained from the patient.

References

Andrews S. FastQC. A quality control tool for high throughput sequencing data, 2010, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed 9 May 2022).

Avanzato VA, Matson MJ, Seifert SN, Pryce R, Williamson BN, Anzick SL, Barbian K, Judson SD, Fischer ER, Martens C, Bowden TA, de Wit E, Riedo FX, Munster VJ. Case study: prolonged infectious SARS-CoV-2 shedding from an asymptomatic immunocompromised individual with cancer. Cell 2020;183:1901–12 e9.

Chakraborty S. Evolutionary and structural analysis elucidates mutations on SARS–CoV2 spike protein with altered human ACE2 binding affinity. Biochem Biophys Res Commun 2021;534:374–80.

Dolan SA, Mulcahy Levy J, Moss A, Pearce K, Butler M, Jung S, Dominguez SR, Mwangi E, Maloney K, Rao S. SARS-CoV-2 persistence in immunocompromised children. Pediatr Blood Cancer 2021;68:e29277.

European Centre for Disease Prevention and Control, 2022. SARS-CoV-2 variants of concern, 2022, https://www.ecdc.europa.eu/en/covid-19/variants-concern (accessed 9 March 2022).

Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, Goes De Jesus J, Main BJ, Tan AL, Paul LM, Brackney DE, Grewal S, Gurfield N, Ka K, Rompay V, Isern S, Michael SF, Coffey LL, Loman NJ, Andersen KG. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol 2019;20:8.

Klinakis A, Cournia Z, Rampias T. N-terminal domain mutations of the spike protein are structurally implicated in epitope recognition in emerging SARS-CoV-2 strains. Comput Struct Biotechnol J 2021;19:5556–67.

Laurini E, Marson D, Aulic S, Fermeglia A, Pricl S. Molecular rationale for SARS–CoV-2 spike circulating mutations able to escape bamlanivimab and etesevimab monoclonal antibodies. Sci Rep 2021;11:20274.

Leung WF, Chorlton S, Tyson J, Al-Rawahi GN, Jassem AN, Prystajecky N, Masud S, Deans GD, Chapman MG, Mirzanejad Y, Murray MCM, Wong PHP. COVID-19 in an immunocompromised host: persistent shedding of viable SARS–CoV-2 and emergence of multiple mutations: a case report. Int J Infect Dis 2022;114:178–82.

Hill V, Plessis L du, Peacock TP, Aggarwal D, Colquhoun R, Carabelli AM, Ellaby N, Gallagher E, Groves N, Jackson B, McCrone J, O'Toole Á, Price A, Sanderson T, Scher E, Southgate J, Volz E, The COVID-19 genomics UK (COG-UK) consortium, Barclay W, Barrett J, Chand M, Connor T, Goodfellow I, Gupta R, Harrison E, Loman N, Myers R, Robertson D, Pybus O, Rambaut A. The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK. bioRxiv, 8 March 2022. https://www.biorxiv.org/content/10.1101/2022.03.08.481609v1 (accessed 9 May 2022).

Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, 2013, https://arxiv.org/abs/1303.3997v2 (accessed 9 May 2022).

Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, Zhao C, Zhang Q, Liu H, Nie L, Qin H, Wang M, Lu Q, Li X, Sun Q, Liu J, Zhang L, Li X, Huang W, Wang Y. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. Cell 2020;182:1284–94 e9.

Liu Z, VanBlargan LA, Bloyet LM, Rothlauf PW, Chen RE, Stumpf S, Zhao H, Errico JM, Theel ES, Liebeskind MJ, Alford B, Buchser WJ, Ellebedy AH, Fremont DH, Diamond MS, Whelan SPJ. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. Cell Host Microbe 2021;29:477–88 e4.

Ma W, Yang J, Fu H, Su C, Yu C, Wang Q, de Vasconcelos ATR, Bazykin GA, Bao Y, Li M. Genomic perspectives on the emerging SARS-CoV-2 omicron variant. Genomics Proteomics Bioinformatics 2022.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 2011;17:10–12.

McCallum M, Bassi J, de Marco A, Chen A, Walls AC, di Iulio J, Tortorici MA, Navarro MJ, Silacci-Fregni C, Saliba C, Sprouse KR, Agostini M, Pinto D, Culap K, Bianchi S, Jaconi S, Cameroni E, Bowen JE, Tilles SW, … Veesler D. SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern. Science 2021;373:648–54.

O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, Colquhoun R, Ruis C, Abu-Dahab K, Taylor B, Yeats C, Du Plessis L, Maloney D, Medd N, Attwood SW, Aanensen DM, Holmes EC, Pybus OG, Rambaut A. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol 2021;7:veab064.

Van der Auwera GA, O'Connor BD. Genomics in the Cloud: using Docker, GATK, and WDL in Terra. O'Reilly Media; 2020.

Wang R, Zhang Q, Ge J, Ren W, Zhang R, Lan J, Ju B, Su B, Yu F, Chen P, Liao H, Feng Y, Li X, Shi X, Zhang Z, Zhang F, Ding Q, Zhang T, Wang X, Zhang L. SARS-CoV-2 variants resist antibody neutralization and broaden host ACE2 usage. bioRxiv 2021 03.09.434497.

Weigang S, Fuchs J, Zimmer G, Schnepf D, Kern L, Beer J, Luxenburger H, Ankerhold J, Falcone V, Kemming J, Hofmann M, Thimme R, Neumann-Haefelin C, Ulferts S, Grosse R, Hornuss D, Tanriver Y, Rieg S, Wagner D, … Kochs G. Within-host evolution of SARS-CoV-2 in an immunosuppressed COVID-19 patient as a source of immune escape variants. Nat Commun 2021;12:6405.

Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, Khor CC, Petric R, Hibberd ML, Nagarajan N. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res 2012;40:11189–201.