

CASE REPORT

Open Access



# Case reports of persistent SARS-CoV-2 infection outline within-host viral evolution in immunocompromised patients

Luca Ruotolo<sup>1</sup>, Silvia Silenzi<sup>2</sup>, Beatrice Mola<sup>1</sup>, Margherita Ortalli<sup>1,2</sup>, Tiziana Lazzarotto<sup>1,2\*</sup> and Giada Rossini<sup>2</sup>

## Abstract

**Background** SARS-CoV-2 is responsible for the ongoing global pandemic, and the continuous emergence of novel variants threatens fragile populations, such as immunocompromised patients. This subgroup of patients seems to be seriously affected by intrahost viral changes, as the pathogens, which are keen to cause replication inefficiency, affect the impaired immune system, preventing efficient clearance of the virus. Therefore, these patients may represent an optimal reservoir for the development of new circulating SARS-CoV-2 variants. The following study aimed to investigate genomic changes in SARS-CoV-2-positive immunocompromised patients over time.

**Methods** SARS-CoV-2-positive nasopharyngeal swabs were collected at different time points for each patient (patient A and patient B), extracted and then analyzed through next-generation sequencing (NGS). The resulting sequences were examined to determine mutation frequencies, describing viral evolution over time.

**Case presentation** Patient A was a 53-year-old patient with onco-hematological disease with prolonged infection lasting for 51 days from May 28th to July 18th, 2022. Three confirmed SARS-CoV-2-positive samples were collected on May 28th, June 15th and July 4th. Patient B was 75 years old and had onco-hematological disease with prolonged infection lasting for 146 days. Two confirmed positive SARS-CoV-2 samples were collected at the following time points: May 21st and August 18th.

**Conclusion** Heat map construction provided evidence of gain and/or loss of mutations over time for both patients, suggesting within-host genomic evolution of the virus. In addition, mutation polymorphisms and changes in the SARS-CoV-2 lineage were observed in Patient B. Sequence analysis revealed high mutational pattern variability, reflecting the high complexity of viral replication dynamics in fragile patients.

**Keywords** SARS-CoV-2, COVID-19, NGS, Immunocompromised patients, Variants of concern, Intrahost variability

## Introduction

Since its first appearance in 2019, SARS-CoV-2 has continued to spread and evolve worldwide, showing important genetic changes which have allowed the emergence of new variants of concern (VOCs). These factors still have a significant impact on the management of the COVID-19 pandemic, as they often provide new viral adaptation skills, such as potential increases in transmissibility, escape from both natural and vaccine-induced immunity and decreased sensitivity to monoclonal antibody therapies [1, 2]. Despite newly acquired features, the

\*Correspondence:

Tiziana Lazzarotto  
tiziana.lazzarotto@unibo.it

<sup>1</sup> Microbiology Unit, DIMEC, Alma Mater Studiorum Università di Bologna, Bologna, Italy

<sup>2</sup> Microbiology Unit, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

origin of SARS-CoV-2 variants has not yet been defined due to numerous factors involved in infectious mechanisms, including a high rate of RNA replication error, viral load, disease severity, disease progression, treatment outcome, drug resistance and cell tropism [3, 4]. It has been reported that SARS-CoV-2 has a similar rate of mutation to other coronaviruses, such as SARS-CoV and MERS-CoV [5], for which within-host diversity was also reported in infected patients during these viral outbreaks [6]. Several hypotheses have been raised regarding the origin of highly divergent SARS-CoV-2 variants, namely, undetected circulation in some geographical regions, a zoonotic origin, and emergence in immunocompromised patients with prolonged SARS-CoV-2 infections [7, 8]. Notably, it is very hard to definitively prove any of these hypotheses. Understanding of the dynamics of these infections and how they correlate with global evolutionary patterns would be of great interest but mostly relies on isolated case reports [7, 9]. One of the most efficient practices for carrying out such research is next-generation sequencing. In addition to its main role in outbreak investigations and studies on drug resistance, NGS provides deep analysis of viral genetic sequences for the determination of intrahost viral population complexity for a large plethora of pathogens [10] because of the production of high-information data outputs.

The present study investigated within-host SARS-CoV-2 variability in nasopharyngeal swabs from two different immunocompromised patients with persistent COVID-19 hospitalized at the *IRCCS Azienda Ospedaliero-Universitaria di Bologna* during 2022. Genomic analyses were conducted using NGS technology coupled with bioinformatic tools to establish the course of viral evolution in patients with prolonged disease and to study the occurrence of specific mutations.

## Methods

This study was conducted at the Center of Regional Reference for Microbiological Emergences (CRREM) laboratory, *IRCCS Azienda Ospedaliero-Universitaria di Bologna*, Italy. Patients were selected based on clinician requests for SARS-CoV-2 variants to establish appropriate medical treatment. Samples were collected anonymously and traced with unique internal codes from the post diagnostic phase. Sequencing was performed both for informational purposes for clinicians with patients in care and in the broader context of territorial monitoring of SARS-CoV-2 circulation. Requirements for sample analysis were based on Cycle threshold (Ct) values < 30, when possible, for at least one or more target genes, depending on the diagnostic assay (Allplex SARS-CoV-2 Assay, Seegene, Simplexa COVID-19 Direct, Diasorin, Xpert Xpress COV-2, Cepheid).

Patient A was a 53-year-old patient with onco-hematological disease with prolonged infection lasting for 51 days from May 28th to July 18th, 2022. Confirmed SARS-CoV-2-positive samples were collected at three time points: on May 28th (day 1, Ct: 21), June 15th (day 19, Ct: 22) and July 4th (day 38, Ct: 22) (Additional file 1).

Patient B was 75 years old and had onco-hematological disease with prolonged infection lasting for 146 days. The first positivity was reported on April 23rd, 2022, with an antigen test (not available), while the last positive molecular swab was reported on September 16th, 2022. The last nasopharyngeal swab tested negative on October 11th. Two confirmed positive SARS-CoV-2 samples were collected at the following time points: May 21st (day 1, Ct: 29) and August 18th (day 86, Ct: 22) (see Additional file 2).

SARS-CoV-2 viral RNA was extracted from nasopharyngeal swabs using a manual QIAamp RNA viral kit (QIAGEN, Hilden, Germany). For both Patient A and B, all selected SARS-CoV-2-positive samples were extracted and subjected to whole-genome sequencing analysis via next-generation sequencing (NGS) technology (Illumina). Libraries were prepared according to the Illumina COV-IDSeq assay (96 samples) library preparation kit (Illumina, USA) following an amplicon-based approach. The libraries were quantified using a Qubit dsDNA High Sensitivity (Thermo Fisher Scientific, USA) assay kit according to the manufacturer's protocol. Then, the libraries were loaded and sequenced using an Illumina MiSeq system. The sequencing results of the FASTQ files were analyzed by BaseSpace Onsite Hub software through the DRAGEN COVID Lineages Application tool (Illumina). This approach allowed sequence trimming and alignment of consensus sequences to the SARS-CoV-2 reference genome (NC\_045512) to determine variant lineages and clades using the Pangolin and NextClade algorithms.

To monitor the accumulation of mutations throughout the whole genome over time, FAST-Q files were submitted to the Stanford SARS-CoV-2 analysis software to produce CodFreq files [11] using the provided pipeline with default settings [12]. CodFreq files were uploaded to the Input Sequence Reads section [13] and run with default output settings (minimum read depth  $\geq 10$ ; nucleotide mixture threshold  $\leq 0.05\%$ ; mutation detection threshold  $\geq 10\%$ ; see Additional files 3 and 4). Output tables were processed by a custom R-Script (RStudio, version 4.2.2; see Additional files 5 and 6) to compare the frequency percentages of a given mutation between the various time points, providing a representation of the results through heat maps. The relevant mutations resulting from the analysis were selected if the minimum allele frequency (MAF) was greater than 5% between at least two different time points [4, 14, 15]. A phylogenetic

analysis of 201 SARS-CoV-2 sequences collected in 2022 was performed. Consensus sequences were first aligned to SARS-CoV-2 reference (NC\_045512) with Unipro UGENE MUSCLE Alignment tool (v44.0) and a Maximum Likelihood tree was constructed with IQ-Tree web server [16]. The best-fitting substitution model was automatically determined (TIM+F+I chosen according to Bayesian Information Criterion, BIC) and the tree was calculated with 1000 bootstrap replicates. Branch support was approximated using the Shimodaira–Hasegawa [SH]-aLRT method (1000 replicates). The tree was rooted to the reference sequence NC\_045512 and its visualization was realized with iTOL Interactive Tree of Life [17].

## Case presentation

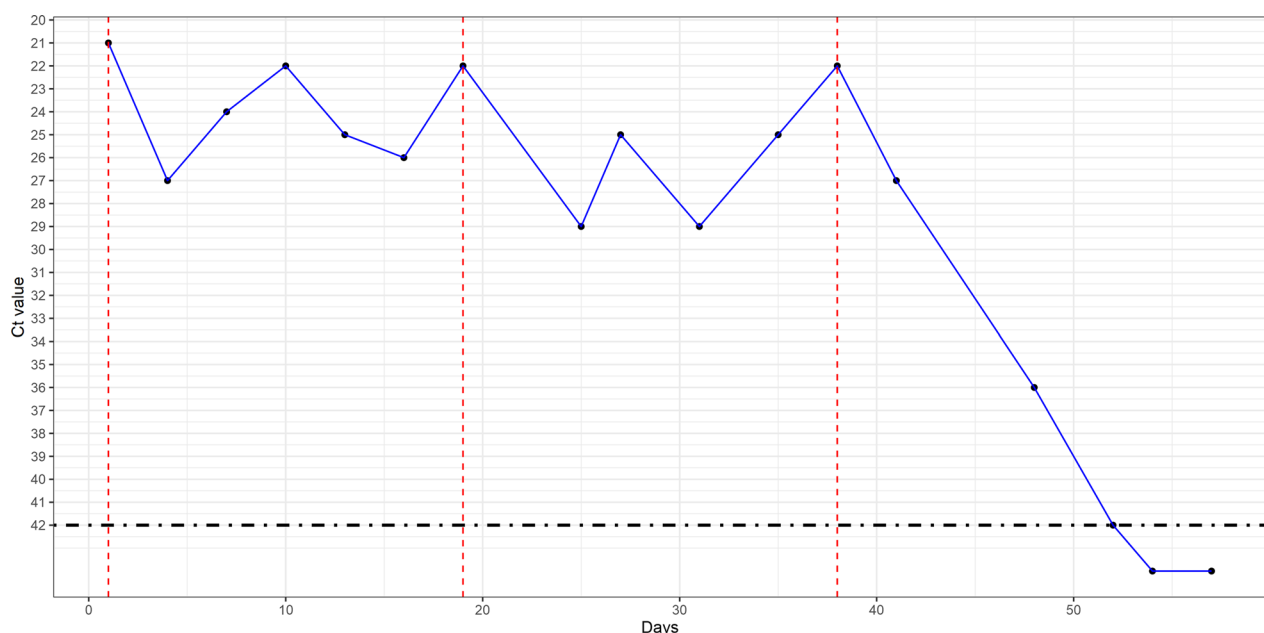
### Patient A

Patient A was a 53-year-old patient with onco-hematological disease with prolonged infection lasting for 51 days (Fig. 1).

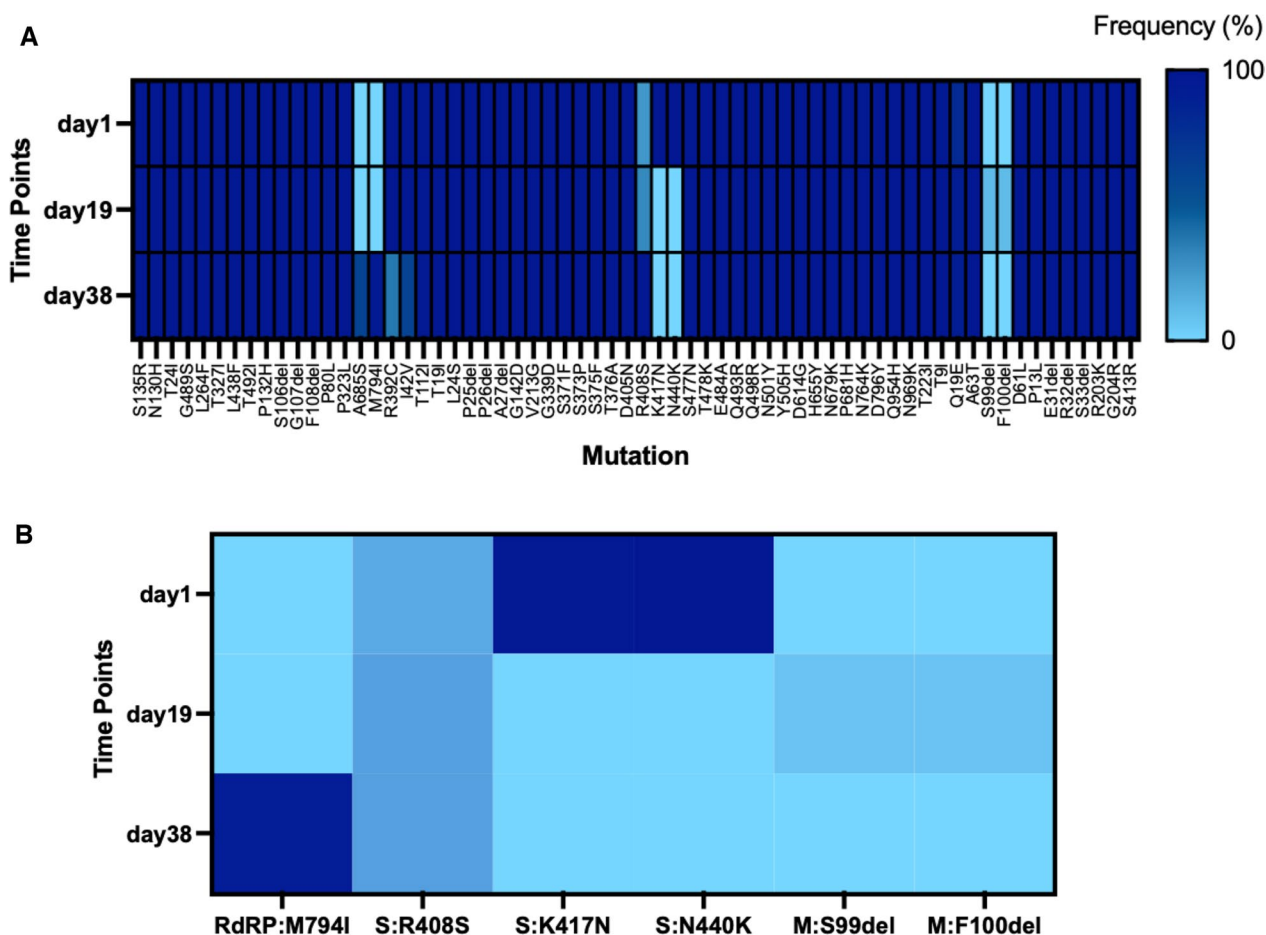
Analyses of consensus sequences at three longitudinally collected time points (day 1, day 19 and day 38) indicated that the virus strains belonged to BA.2 Omicron lineage. Day 1 was collected on May 28th, day 19 on June 15th and day 38 on July 4th, 2022.

The consensus sequences from the day 1 sample presented 55 mutations (29 of which were located within the S gene) compared to the Wuhan-1 reference strain, while in the day 19 and day 38 samples, 54 mutations

(27 located in the S gene) were detected in comparison to the reference strain. Overall, the vast majority of mutations that occurred throughout day 1 (mean = 99.5;  $\sigma = 0.69$ ), day 19 (mean = 99.2;  $\sigma = 0.64$ ) and day 38 (mean = 99.4;  $\sigma = 0.80$ ) did not present high variability over time, with some notable exceptions (Fig. 2). The comparison of consensus sequences over the three time points revealed that two de novo mutations in the RdRP gene, A685S and M794I, emerged at the last time point (day 38) with a frequency of 62.4% and 95.4%, respectively, suggesting a within-host viral evolution attempt. Both the mutations are not common in SARS-CoV-2 Omicron variants [18]. A slight increasing trend in the frequency of R408S in the S gene (25.2% on day 1, 30.6% on day 19 and 31.1% on day 38) and of Q19E in the M gene (80.6% on day 1, 92.0% on day 19 and 91.3% on day 38) was recorded. Both these mutations then emerged as high prevalent in successive Omicron lineages in the global population [18]. Conversely, S99del and F100del mutations in the M gene showed variability in their frequency, with a total absence on day 1 and day 38 and a low frequency on day 19 (11.2% and 10.9%, respectively). K417N and N440K mutations in the S gene, common in several Omicron variants [18], were detected only on day 1 (frequency > 99%) and then disappeared on following days 19 and 38 (Table 1).



**Fig. 1** Evolution of prolonged SARS-CoV-2 infection in Patient A. The blue line represents the trend of viral load based on Ct values, while black points indicate sampling days. The dotted and dashed black lines mark the thresholds for which the swab was considered negative ( $Ct > 42$ ); samples with  $40 < Ct < 42$  were considered to have low positivity. The dashed red lines indicate the time points selected for sequencing analysis



**Fig. 2** Heat map showing mutation frequencies in longitudinally collected Patient A samples. Genes and mutations are reported on the x-axis with “gene: mutation” wording, while time points are on the y-axis. The frequency is displayed through a color gradient ranging from light blue (for low mutations) to dark blue (for high mutations). **A** Heat map showing that the general representation of the detected mutations allows the discrimination of sites with low variability from those with high variability in mutation frequency. **B** Mutations presenting greater intrahost variability (MAF > 5% between at least two different time points) throughout three longitudinally collected samples from the same individual

**Table 1** Mutation frequency (%) in SARS-CoV-2 genome sequences collected at three time points from Patient A

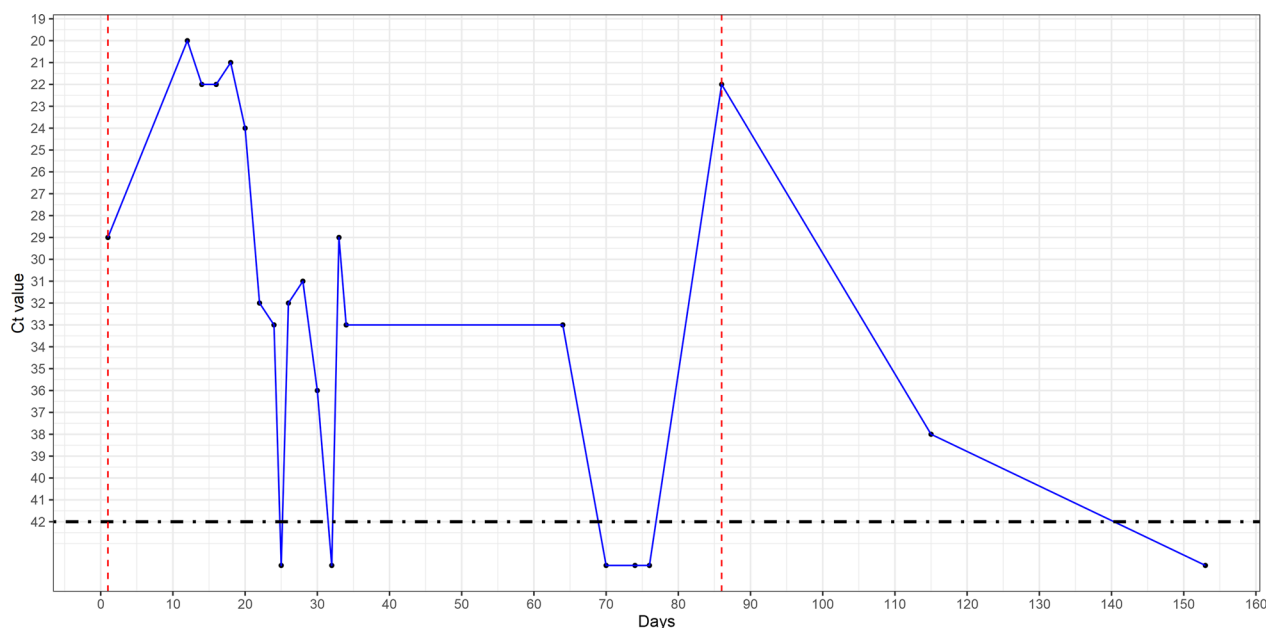
Gene	Mutation	Day 1 (%)	Day 19 (%)	Day 38 (%)
M	Q19E	80.6	92.0	91.3
M	S99del	0	11.2	0
M	F100del	0	10.9	0
RdRP	A685S	0	0	62.4
RdRP	M794I	0	0	95.4
S	R408S	25.2	30.6	31.1
S	K417N	99.0	0	0
S	N440K	100	0	0

Mutations were selected considering a minimum variation of 5% (minimum allele frequency cutoff value) between at least two samples

**Patient B**

Patient B was 75 years old with onco-hematological disease experiencing prolonged SARS-CoV-2 infection lasting for 146 days. The viral load, very high in the first 20 days and then gradually decreasing, showed a swinging trend (Fig. 3). Two nasopharyngeal samples collected on May 22nd, 2022 (day 1) and on August 18th, 2022 (day 86) were sequenced and analyzed.

The consensus sequences produced by NGS allowed the identification of two different Omicron sublineages at two different time points: the BA.2 lineage was assigned to sample from day 1, while the day 86 sample was recognized as the BA.2.1 lineage. These results may indicate the occurrence of an intrahost evolution of SARS-CoV-2. Compared to reference strain sequence, a total of 55 mutations were detected in samples collected on day 1, 31 of which were in the S gene. On the other hand, 61 mutations were detected



**Fig. 3** Evolution of prolonged SARS-CoV-2 infection in Patient B. The blue line represents the trend of infection based on the E gene Ct values reported in the diagnostic center, while the black points indicate sampling days. The dotted and dashed black lines mark the thresholds for which the swab was considered negative ( $Ct > 42$ ); samples with  $40 < Ct < 42$  were considered to have low positivity. The dashed red lines indicate the time points selected for genomic analyses

in day 86 samples, 34 of which were inside the S gene. Overall, the vast majority of mutations that occurred throughout day 1 (mean = 99.6;  $\sigma = 0.42$ ) and day 86 (mean = 99.7;  $\sigma = 0.33$ ) did not present high variability over time, with some notable exceptions. To investigate the temporal evolution of viral populations between the two time points, the frequency of genomic variations during 146 days of infection was analyzed (Fig. 4).

Compared to the first collected sample (day 1), in the sample collected on day 86, 10 de novo mutations have arisen: T35I in nsp9; Y324C in nsp13; P217H, N440K, K444N, L455F, V642G, and D1153Y in the S gene; T30I in the E gene and A182V in the N gene. For each mutation, the frequency on day 1 was 0%, while the frequency on day 86 was  $> 80\%$  (Table 2). The mutations Q19E in E protein and R408S in S protein have been shown to increase their frequency over time, from 43.9 to 90.6% and from 24.7 to 32.9, respectively (Table 2).

Four de novo deletions in S protein (F374del, S375del, T376del and Y144del) emerged in the sample collected at day 86 (Table 2).

Most of these mutations and deletions are considered very rare and have not been subsequently seen at high prevalence in the global population [18]. The only mutations that were found at high prevalence in global population were N440K and R408S in S protein together with Q19E in M protein (Table 2) [18].

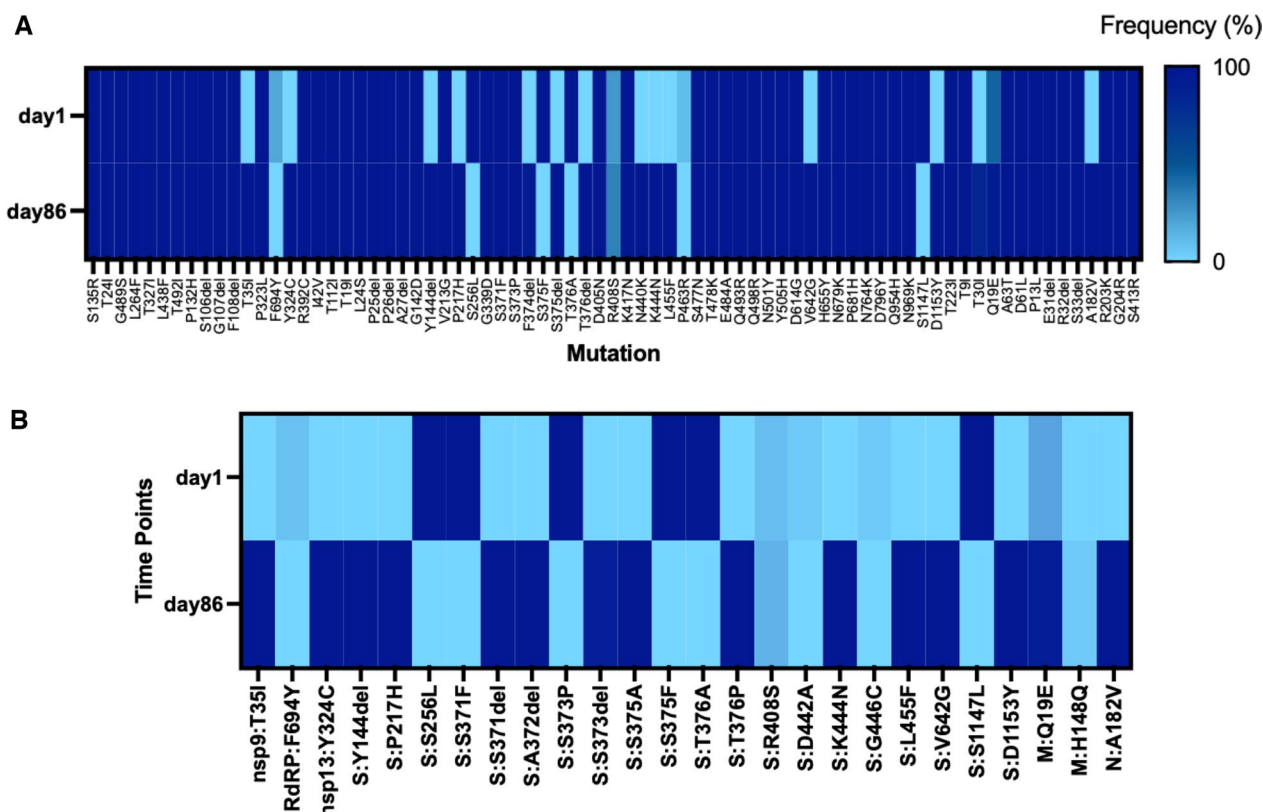
Furthermore, by comparing the two time points, we also observed the disappearance of four mutations present on day 1 but not on day 86: F694Y in the RdRP gene (from 18.4% to 0%) and S256L, P463R, and S1147L in the S gene. (Table 2).

Both Patient A and Patient B underwent phylogenetic analysis to explore the magnitude of intrahost viral diversity at the reported time points (Fig. 5, Additional files 7, 8).

Every analyzed sample was located in the corresponding Omicron BA.2 clade, as reported in previous results. Phylogenetic analysis revealed within-host viral variability in both Patient A and Patient B. Specifically, there was strong homology between the day 1 and day 19 time points for Patient A, while major diversity emerged only later on day 38. Similarly, day 1 of Patient B was similar to that of Patient A at the first time points (day 1 and day 19), as they were classified as BA.2 Omicron lineage but an important viral evolution was discovered on day 86, which is also considered a SARS-CoV-2 lineage switch from BA.2 to BA.2.1.

## Conclusion

Since its first appearance, the viral genome of SARS-CoV-2 has undergone consistent changes because of both natural selection occurring during infection and the effect of the immune system. Consequentially, the emergence of new variants has been globally documented



**Fig. 4** Heatmap showing mutation frequencies in longitudinally collected Patient B samples. Genes and mutations are reported on the x-axis with “gene: mutation” wording, while time points are on the y-axis. The frequency is displayed through a color gradient from light blue for low-frequency mutations to dark blue for high-frequency mutations. **A** Heat map showing that the general representation of the detected mutations allows the discrimination of sites with low variability from those with high variability in mutation frequency. **B** Mutations presenting greater intrahost variability (MAF > 5% between at least two different time points) throughout two longitudinally collected samples from the same individual

throughout the pandemic. A leading and widely discussed hypothesis suggested the prolonged infection in immunocompromised patients as the potential source of new variants contributing to global spread [19, 20]. The current study provided insight into the potential within-host evolution of SARS-CoV-2 in two immunocompromised patients, Patient A and Patient B, with prolonged infection lasting for 51 and 146 days, respectively, and as in other numerous case reports, we also can provide documentation of mutations accumulation over time, mostly in spike.

The analysis of viral genomic sequences in different samples collected longitudinally, suggested the possibility of the emergence of new mutations inside the

immunocompromised host that only rarely are then found at high prevalence in the general population. We observed the emergence of several de novo mutations that are non-shared between the two patients: T35I in nsp9; Y324C in nsp13; P217H, S256L, K444N, L455F, V642G, and D1153Y in S; T30I in E; A182V in N; A685S and M794I in RdRP. Most of these mutations are considered very rare, arising most probably under therapeutical pressure and that have not been subsequently seen at high prevalence in the global population. For instance, K444N and L455F in S have been associated with reduced susceptibility to several monoclonal antibodies [20–24], and T30I in E has been indicated as a potential marker of long-term SARS-CoV-2 infections, being one of the most frequent

**Table 2** Mutation frequency (percentage) within two time points available for Patient B. Mutations were selected considering a minor allele frequency > 5% between at least two samples

Gene	Mutation	Day 1 (%)	day8 6 (%)
E	T30I	0	82.1
M	Q19E	43.9	90.6
N	A182V	0	99.3
nsp13	Y324C	0	99.7
nsp9	T35I	0	99.4
RdRP	F694Y	18.4	0.0
S	Y144del	0	99.6
S	P217H	0	99.7
S	S256L	99.5	0
S	F374del	0	91.2
S	S375del	0	99.9
S	T376del	0	99.8
S	N440K	0	100
S	R408S	24.7	32.9
S	K444N	0	100
S	L455F	0	100
S	P463R	10	0
S	V642G	0	99.6
S	S1147L	99.6	0
S	D1153Y	0	99.8

The data analysis led to the observation of different behaviors in terms of the presence or absence of mutations

occurring arising mutation in persistent infection in immunocompromised patients but is absent from the global phylogeny [25]. On the other hand, N440K is a well-known region binding domain (RBD) mutation frequently observed in several Omicron lineages [26] and is associated to a better viral fitness by improving the binding to the human ACE2 receptor [27]. N440K was observed to disappear over time in Patient A and, at contrary, to increase its frequency in Patient B. The sudden loss or gain of globally recognized mutations in the SARS-CoV-2 Omicron lineages, such as N440K (loss in Patient A, gain in Patient B) and K417N (loss in Patient A) has yet to be investigated, as it is unclear

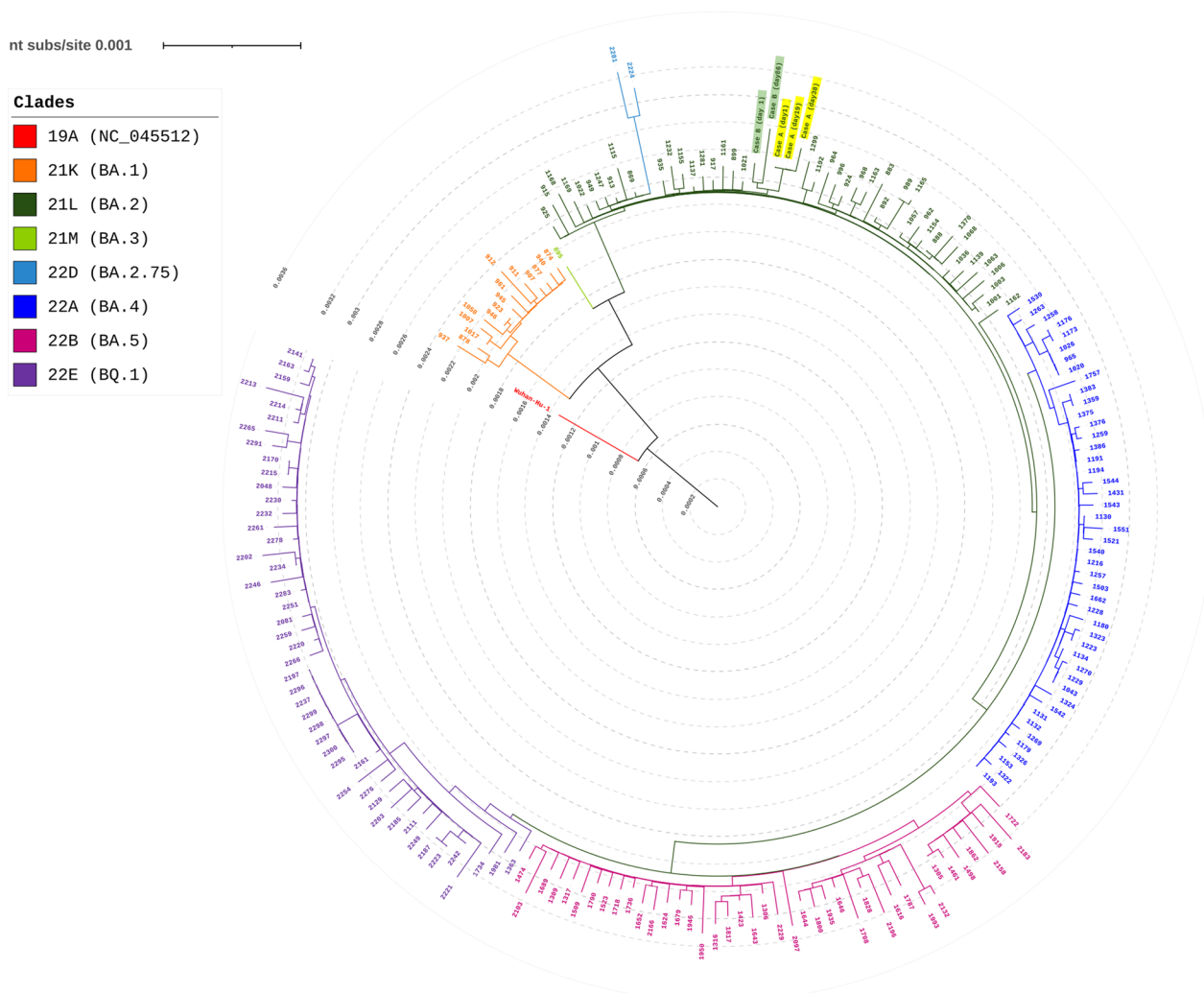
whether these unusual events are random or related to unknown mechanisms of adaptation to clinically fragile hosts.

In samples from both patients, Q19E in M and R408S in S exhibited an increase in frequency over time; both these mutations have been detected in subsequent Omicron lineage sequences at high prevalence. Furthermore, de novo deletions were identified in spike: F374del, S375del, T376del and Y144del. Interestingly, Y144del is part of a group of N-terminal domain (NTD) deletions between positions 141–146 occurring in the alpha and omicron BA.1 and BA.2 variants. These deletions are associated with resistance to several NTD-binding neutralizing monoclonal antibodies, although they do not appear to reduce the neutralizing activity of plasma from convalescent or vaccinated individuals [28, 29]. However, the presence of Y144del in BA.2 has not been broadly reported, although is common subsequently in XBB lineages.

Despite the promising and informative results, our study has some limitations. First, we included only two immunocompromised patients and only few samples for each patient could be longitudinally evaluated. Second, many clinical and therapeutic data for the two patients were not available. These limitations do not allow drawing absolute conclusions about findings and, therefore, results cannot be generalized.

Nevertheless, description of the mutational trend in these two immunocompromised patients, together to other case reports and broader cohort of patients, may provide some valid suggestions to elucidate the very complex process of virus evolution in immunocompromised patients.

Our data, even if collected from only two patients, are in line with results from Raglow Z. et al. obtained from a much larger cohort of patients, showing how some mutations, mostly in spike but also in other genomic regions, are rarely observed in global sequencing data but are strongly associated with escape from therapeutics. These observations highlight the complexity and



**Fig. 5** Phylogenetic tree obtained through the analysis of 201 SARS-CoV-2 sequences collected during 2022, including the considered cases at each time point. Patient A and B time points are highlighted with yellow and light green label background, respectively. Except for the analyzed cases, all samples are reported with an internal system ID, which consists of progressive numbering for each patient’s nasopharyngeal swab (NPS). Clades classification is provided by NextClade consensus sequence analysis report

the importance of an extensive sequencing data monitoring with different approaches to provide insights into future evolutionary patterns of SARS-CoV-2 [20].

**Abbreviations**

- NGS Next-generation sequencing
- NPS Nasopharyngeal swabs
- COVID-19 Coronavirus disease 2019
- SARS-CoV-2 Severe acute respiratory syndrome coronavirus 2
- Ct Cycle threshold
- BIC Bayesian Information Criterion
- RBD Region binding domain
- NTD N-terminal domain
- MAF Multiple allele frequency

**Supplementary Information**

The online version contains supplementary material available at <https://doi.org/10.1186/s12985-024-02483-y>.

- Additional file 1
- Additional file 2
- Additional file 3
- Additional file 4
- Additional file 5
- Additional file 6
- Additional file 7
- Additional file 8

**Acknowledgements**

Not applicable.

**Author contributions**

Conceptualization, GR; methodology, GR; software, LR; validation, GR; formal analysis, LR, BM; investigation, LR, SS and BM; resources, GR and TL; data curation, LR; writing—original draft preparation, LR, BM; writing—review



and editing, GR and SS; visualization, BM, SS and LR; supervision, GR; project administration, GR; funding acquisition, TL.

### Funding

This research was supported by EU funding within the NextGenerationEU-MUR PNRR Extended Partnership initiative on Emerging Infectious Diseases (Project No. PE00000007, INF-ACT).

### Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its additional files. Additional file 1: Patient A—Table of Ct.csv; Additional file 2: Patient B—Table of Ct.csv; Additional file 3: Patient A—Stanford Analysis.csv; Additional file 4: Patient B—Stanford Analysis.csv; Additional file 5: Patient A—analysis & plots.R; Additional file 6: Patient B—analysis & plots.R; Additional file 7: Phylogenetic tree—focus on immunocompromised patients.pdf; Additional file 8: Phylogenetic tree—focus on immunocompromised patients.png.

### Declarations

#### Ethical approval and consent to participate

Patient consent was waived. Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

Received: 29 May 2024 Accepted: 21 August 2024

Published online: 03 September 2024

### References

- Lee J-S, Yun KW, Jeong H, Kim B, Kim MJ, Park JH, et al. SARS-CoV-2 shedding dynamics and transmission in immunosuppressed patients. *Virulence*. 2022;13:1242–51. <https://doi.org/10.1080/21505594.2022.2101198>.
- Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature*. 2021;593:130–5. <https://doi.org/10.1038/s41586-021-03398-2>.
- Voloch CM, De Silva Francisco R Jr, De Almeida LGP, Brustolini OJ, Cardoso CC, Gerber AL, et al. Intra-host evolution during SARS-CoV-2 prolonged infection. *Virus Evol*. 2021;7:veab078. <https://doi.org/10.1093/ve/veab078>.
- Fuhrmann L, Jablonski KP, Beerenwinkel N. Quantitative measures of within-host viral genetic diversity. *Curr Opin Virol*. 2021;49:157–63. <https://doi.org/10.1016/j.coviro.2021.06.002>.
- Wang Y, Wang D, Zhang L, Sun W, Zhang Z, Chen W, et al. Intra-host variation and evolutionary dynamics of SARS-CoV-2 populations in COVID-19 patients. *Genome Med*. 2021;13:30. <https://doi.org/10.1186/s13073-021-00847-5>.
- Al Khatib HA, Benslimane FM, Elbasher IE, Coyle PV, Al Maslamani MA, Al-Khal A, et al. Within-host diversity of SARS-CoV-2 in COVID-19 patients with variable disease severities. *Front Cell Infect Microbiol*. 2020;10:575613. <https://doi.org/10.3389/fcimb.2020.575613>.
- Harari S, Miller D, Fleishon S, Burstein D, Stern A. Using big sequencing data to identify chronic SARS-Coronavirus-2 infections. *Nat Commun*. 2024;15(1):648. <https://doi.org/10.1038/s41467-024-44803-4>.
- Avanzato VA, Matson MJ, Seifert SN, Pryce R, Williamson BN, Anzick SL, et al. Case study: prolonged infectious SARS-CoV-2 shedding from an asymptomatic immunocompromised individual with cancer. *Cell*. 2020;183:1901–1912.e9. <https://doi.org/10.1016/j.cell.2020.10.049>.
- Pérez-Lago L, Aldámiz-Echevarría T, García-Martínez R, Pérez-Latorre L, Herranz M, Sola-Campoy P, et al. Different within-host viral evolution dynamics in severely immunosuppressed cases with persistent SARS-CoV-2. *Biomedicines*. 2021;9:808. <https://doi.org/10.3390/biomedicines9070808>.
- Knyazev S, Hughes L, Skums P, Zelikovskiy A. Epidemiological data analysis of viral quasispecies in the next-generation sequencing era. *Brief Bioinform*. 2021;22:96–108. <https://doi.org/10.1093/bib/bbaa101>.
- Codon frequency table—Stanford Coronavirus Antiviral & Resistance Database (CoVDB). Last updated on Aug 30, 2021. <https://covdb.stanford.edu/page/codfreq/>. Accessed 9 June 2023.
- SARS-CoV-2 NGS Pipeline—Stanford Coronavirus Antiviral & Resistance Database (CoVDB). Last updated on May 14, 2024. <https://covdb.stanford.edu/sierra/sars2/ngs2codfreq/>. Accessed 9 June 2023.
- SARS-CoV-2 Sequence Reads (NGS) Analysis—Stanford Coronavirus Antiviral & Resistance Database (CoVDB). Last updated on May 14, 2024. <https://covdb.stanford.edu/sierra/sars2/by-reads/>. Accessed 9 June 2023.
- Shen Z, Xiao Y, Kang L, Ma W, Shi L, Zhang L, et al. Genomic diversity of severe acute respiratory syndrome-coronavirus 2 in patients with coronavirus disease 2019. *Clin Infect Dis Off Publ Infect Dis Soc Am*. 2020;71:713–20. <https://doi.org/10.1093/cid/ciaa203>.
- Lythgoe KA, Hall M, Ferretti L, De Cesare M, MacIntyre-Cockett G, Trebes A, et al. SARS-CoV-2 within-host diversity and transmission. *Science*. 2021;372:eabg0821. <https://doi.org/10.1126/science.abg0821>.
- Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res*. 2016;44(W1):W232–5. <https://doi.org/10.1093/nar/gkw256>.
- Leticia I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021;49(W1):W293–6. <https://doi.org/10.1093/nar/gkab301>.
- Gangavarapu K, Latif AA, Mullen JL, Alkuzweny M, Huffbauer E, Tsueng G, et al. Outbreak.info genomic reports: scalable and dynamic surveillance of SARS-CoV-2 variants and mutations. *Nat Methods*. 2023. <https://doi.org/10.1038/s41592-023-01769-3>.
- Marques AD, Graham-Wooten J, Fitzgerald AS, Sobel Leonard A, Cook EJ, Everett JK, et al. SARS-CoV-2 evolution during prolonged infection in immunocompromised patients. *mBio*. 2024;15(3):e0011024. <https://doi.org/10.1128/mbio.00110-24>.
- Raglow Z, Surie D, Chappell JD, Zhu Y, Martin ET, Kwon JH, et al. SARS-CoV-2 shedding and evolution in patients who were immunocompromised during the omicron period: a multicentre, prospective analysis. *Lancet Microbe*. 2024;5(3):e235–46. [https://doi.org/10.1016/S2666-5247\(23\)00336-1](https://doi.org/10.1016/S2666-5247(23)00336-1).
- Baum A, Fulton BO, Wloga E, Copin R, Pascal KE, Russo V, et al. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science*. 2020;369:1014–8. <https://doi.org/10.1126/science.abd0831>.
- Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JC, et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife*. 2020;9:e61312. <https://doi.org/10.7554/eLife.61312>.
- Starr TN, Greaney AJ, Addetia A, Hannon WW, Choudhary MC, Dingens AS, et al. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science*. 2021;371:850–4. <https://doi.org/10.1126/science.abf9302>.
- Muecksch F, Weisblum Y, Barnes CO, Schmidt F, Schaefer-Babajew D, Wang Z, et al. Affinity maturation of SARS-CoV-2 neutralizing antibodies confers potency, breadth, and resilience to viral escape mutations.

- Immunity. 2021;54:1853-1868.e7. <https://doi.org/10.1016/j.immuni.2021.07.008>.
25. Wilkinson SAJ, Richter A, Casey A, Osman H, Mirza JD, Stockton J, et al. Recurrent SARS-CoV-2 mutations in immunodeficient patients. *Virus Evol.* 2022;8(2):e050. <https://doi.org/10.1093/ve/veac050>.
  26. Wang Z, Schmidt F, Weisblum Y, Muecksch F, Barnes CO, Finkin S, et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature.* 2021;592:616–22. <https://doi.org/10.1038/s41586-021-03324-6>.
  27. Kullappan M, Mary U, Ambrose JM, Veeraraghavan VP, Surapaneni KM. Elucidating the role of N440K mutation in SARS-CoV-2 spike—ACE-2 binding affinity and COVID-19 severity by virtual screening, molecular docking and dynamics approach. *J Biomol Struct Dyn.* 2023;41(3):912–29. <https://doi.org/10.1080/07391102.2021.2014973>.
  28. Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, et al. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell.* 2020;182:1284-1294.e9. <https://doi.org/10.1016/j.cell.2020.07.012>.
  29. McCallum M, De Marco A, Lempp FA, Tortorici MA, Pinto D, Walls AC, et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell.* 2021;184:2332-2347.e16. <https://doi.org/10.1016/j.cell.2021.03.028>.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.