Received: 10 January 2024 | Accepted: 14 May 2024

DOI: 10.1002/jmv.29708

RESEARCH ARTICLE

JOURNAL OF MEDICAL VIROLOGY WILEY

Persistence of SARS-CoV-2 infection and viral *intra*- and *inter*-host evolution in COVID-19 hospitalized patients

Grazia Pavia[1] | Angela Quirino[1] | Nadia Marascio[1] | Claudia Veneziano[2,3] | Federico Longhini[4] | Andrea Bruni[4] | Eugenio Garofalo[4] | Marta Pantanella[1] | Michele Manno[1] | Simona Gigliotti[1] | Aida Giancotti[1] | Giorgio Settimo Barreca[1] | Francesco Branda[5] | Carlo Torti[6,7] | Salvatore Rotundo[8] | Rosaria Lionello[8] | Valentina La Gamba[8] | Lavinia Berardelli[8] | Sara Palma Gullì[8] | Enrico Maria Trecarichi[8] | Alessandro Russo[8] | Camillo Palmieri[2] | Carmela De Marco[2,3] | Giuseppe Viglietto[2,3] | Marco Casu[9] | Daria Sanna[10] | Massimo Ciccozzi[5] | Fabio Scarpa[10] | Giovanni Matera[1]

[1]Unit of Clinical Microbiology, Department of Health Sciences, "Magna Græcia" University Hospital, Catanzaro, Italy

[2]Department of Experimental and Clinical Medicine, "Magna Graecia" University of Catanzaro, Catanzaro, Italy

[3]Interdepartmental Center of Services (CIS), Molecular Genomics and Pathology, "Magna Græcia" University of Catanzaro, Catanzaro, Italy

[4]Unit of Anesthesia and Intensive Care, Department of Medical and Surgical Sciences, "Magna Graecia" University, Catanzaro, Italy

[5]Unit of Medical Statistics and Molecular Epidemiology, Università Campus Bio-Medico di Roma, Rome, Italy

[6]Dipartimento di Scienze di Laboratorio e Infettivologiche, Fondazione Policlinico Universitario "A. Gemelli" IRCCS, Rome, Italy

[7]Dipartimento di Sicurezza e Bioetica, Università Cattolica del Sacro Cuore, Rome, Italy

[8]Unit of Infectious and Tropical Disease, Department of Medical and Surgical Sciences, "Magna Græcia" University Hospital, Catanzaro, Italy

[9]Department of Veterinary Medicine, University of Sassari, Sassari, Italy

[10]Department of Biomedical Sciences, University of Sassari, Sassari, Italy

Correspondence

Angela Quirino, Unit of Clinical Microbiology, Department of Health Sciences, "Magna Græcia" University Hospital, 88100 Catanzaro, Italy.
Email: quirino@unicz.it

Fabio Scarpa, Department of Biomedical Sciences, University of Sassari, Viale San Pietro 43b, 07100 Sassari, Italy.
Email: fscarpa@uniss.it

Funding information

Regione Calabria, Grant/Award Number: COVID19@UMG POR Calabria-FESR/FSE 2014-2020 D.D.R.C. n. 4584 – Azione 10.5.12.

Abstract

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) persistence in COVID-19 patients could play a key role in the emergence of variants of concern. The rapid *intra*-host evolution of SARS-CoV-2 may result in an increased transmissibility, immune and therapeutic escape which could be a direct consequence of COVID-19 epidemic currents. In this context, a longitudinal retrospective study on eight consecutive COVID-19 patients with persistent SARS-CoV-2 infection, from January 2022 to March 2023, was conducted. To characterize the *intra*- and *inter*-host viral evolution, whole genome sequencing and phylogenetic analysis were performed on nasopharyngeal samples collected at different time points. Phylogenetic reconstruction revealed an accelerated SARS-CoV-2 *intra*-host evolution and emergence of antigenically divergent variants. The Bayesian inference and principal coordinate analysis analysis showed a host-based genomic structuring

Fabio Scarpa and Giovanni Matera equally contributed to the study.

Institution at which the work was performed: "Magna Graecia" University of Catanzaro and University of Sassari.

among antigenically divergent variants, that might reflect the positive effect of containment practices, within the critical hospital area. All longitudinal antigenically divergent isolates shared a wide range of amino acidic (aa) changes, particularly in the Spike (S) glycoprotein, that increased viral transmissibility (K417N, S477N, N501Y and Q498R), enhanced infectivity (R346T, S373P, R408S, T478K, Q498R, Y505H, D614G, H655Y, N679K and P681H), caused host immune escape (S371L, S375F, T376A, K417N, and K444T/R) and displayed partial or complete resistance to treatments (G339D, R346K/T, S371F/L, S375F, T376A, D405N, N440K, G446S, N460K, E484A, F486V, Q493R, G496S and Q498R). These results suggest that multiple novel variants which emerge in the patient during persistent infection, might spread to another individual and continue to evolve. A pro-active genomic surveillance of persistent SARS-CoV-2 infected patients is recommended to identify genetically divergent lineages before their diffusion.

KEYWORDS

inter-host viral genetic evolution, intra-host viral genetic evolution, phylodynamic analysis, phylogenetic analysis, SARS-CoV-2 immune escape mutations, SARS-CoV-2 persistence

1 | INTRODUCTION

Since late 2019, the rapid spread of Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) marked the onset of an unprecedented global health scenario, with 770 million confirmed cases and 6.9 million deaths.[1] The emergence of the first SARS-CoV-2 "variant of concern" (VOC), B.1.1.7 (Alpha), shaped the landscape of Coronavirus Disease 2019 (COVID-19).[2,3] Notably, one of the striking features was its relatively extended phylogenetic branch length, with a lack of intermediate genetic variants prior its detection.[3,4] Moreover, the considerable number of amino acid (aa) substitutions and their rapid accumulation stood in contrast to the *intra*- and *inter*-host evolution typically observed in acute respiratory infection.[5,6] Similar to the Alpha VOC, the autumn of 2021 witnessed the appearance of the Omicron complex lineages (BA.1, BA.2, BA.3, BA.4, and BA.5), signifying a distinct phase in the ongoing pandemic.[7–10] Interestingly, these lineages exhibited a longer branch length compared to preceding VOCs, incorporating over 50 mutations across the entire genome.[10,11] A substantial number of these mutations, classified as mutations of interest (MOIs) and mutations of concern (MOCs), were concentrated in the N-terminal domain (NTD) and receptor binding domain (RBD) of the Spike (S) glycoprotein. This led to an enhanced binding affinity with the ACE2 receptor and an unprecedented capacity to evade neutralizing antibodies (nAbs) and therapeutic monoclonal antibodies (mAbs).[12–16] It became apparent swiftly that viral replication during persistent infection could be the most plausible hypothesis to explain the origin of VOCs.[3,17–19] Despite several reported cases of human-animal and animal-human transmission,[20,21] there is no evidence that such events led to increased evolutionary rates or contributed to community transmission. Conversely, in persistent SARS-CoV-2 infected patients both an accelerated *intra*-host evolution and heightened genetic diversity of the virus were observed, with lineage-defining mutations closely resembling those found in VOCs.[22–26] Immune system conditions, anti-SARS-CoV-2 vaccination, therapies, and various clinical risk factors collectively exert selective pressure on the virus, causing it to evolve at a rate significantly higher than the acute course of infection.[26] Thus, divergent variants emerging within and between persistently infected SARS-CoV-2 hosts may have a substantial impact on future scenarios of the COVID-19 epidemic.

Despite the current classification of SARS-CoV-2 infection as endemic, there remains insufficient knowledge about the frequency and specific circumstances under which antigenically divergent lineages might emerge. Consequently, ongoing genomic monitoring of the "reservoir" represented by persistently infected SARS-CoV-2 patients is essential to detect and contain divergent lineages which could impact on public health.

In this context, we initiated and carried out a retrospective, longitudinal study focusing on nasopharyngeal positive swabs for SARS-CoV-2 in patients with persistent infection. The primary objective of this investigation was to delineate the *intra*- and *inter*-host genetic diversity as well as the evolutionary dynamics of SARS-CoV-2 during persistent infection. As a secondary objective, we examined the mutational patterns of viral variants to identify aa changes that could enhance viral fitness and facilitate spread.

2 | MATERIALS AND METHODS

2.1 | Study design

This longitudinal, retrospective study was conducted at the 'R. Dulbecco' University Hospital of Catanzaro from January 2022 to

March 2023. The study workflow is depicted in Figure 1. The experimental protocol received approval from the Ethical Committee of Calabria Region in Meeting No. 116 held on April 20th, 2023. Despite the data being fully anonymized and deidentified, written informed consent was obtained from patients or their next of kin before publication.

2.2 | Patients

We included all consecutive adult (i.e., ≥18 years old) patients with a persistent SARS-CoV-2 infection with one or more of the following clinical conditions: acute respiratory failure, Acute Respiratory Distress Syndrome, virus-induced septic shock or multiple organ dysfunction.[27] Inclusion and exclusion criteria are shown in Figure 1.

Persistent cases were defined as chronic SARS-CoV-2 infected patients with nasopharyngeal viral shedding duration >17 days.[28–31] The duration of SARS-CoV-2 positivity was calculated by counting days from the date of the initial clinical diagnosis documented in hospital records. As some patients had prolonged infection before hospital admission, time zero was defined as either the date of the first positive SARS-CoV-2 test for the current episode of infection or the most recent positive nasopharyngeal swab for SARS-CoV-2.

Demographic and clinical data for the enrolled patients were gathered from hospital records and documented in an anonymized database. To assess viral evolution and delineate *intra*- and *inter*-host viral genetic diversity, molecular and phylogenetic analyses were conducted on longitudinal nasopharyngeal samples collected for routine care at various time points (days) throughout persistent chronic SARS-CoV-2 infection (Figure 1).

2.3 | Molecular SARS-CoV-2 diagnostics

SARS-CoV-2 routine diagnostics was performed by real-time reverse-transcriptase polymerase-chain-reaction (rRT-PCR) from naso-pharyngeal swabs collected in tubes containing UTM® Universal Transport Medium™ (COPAN Diagnostics Inc.) by Allplex SARS-CoV-2 Assay (Arrow Diagnostics Srl). The diagnostic protocol revealed the presence of three SARS-CoV-2 genes: Envelope (E), RNA-dependent-RNA polymerase (RdRp), and Nucleocapsid (N). According to the manufacturer's instructions, a Cycle threshold (Ct) value < 40 was considered positive. For SARS-CoV-2 whole genome sequencing

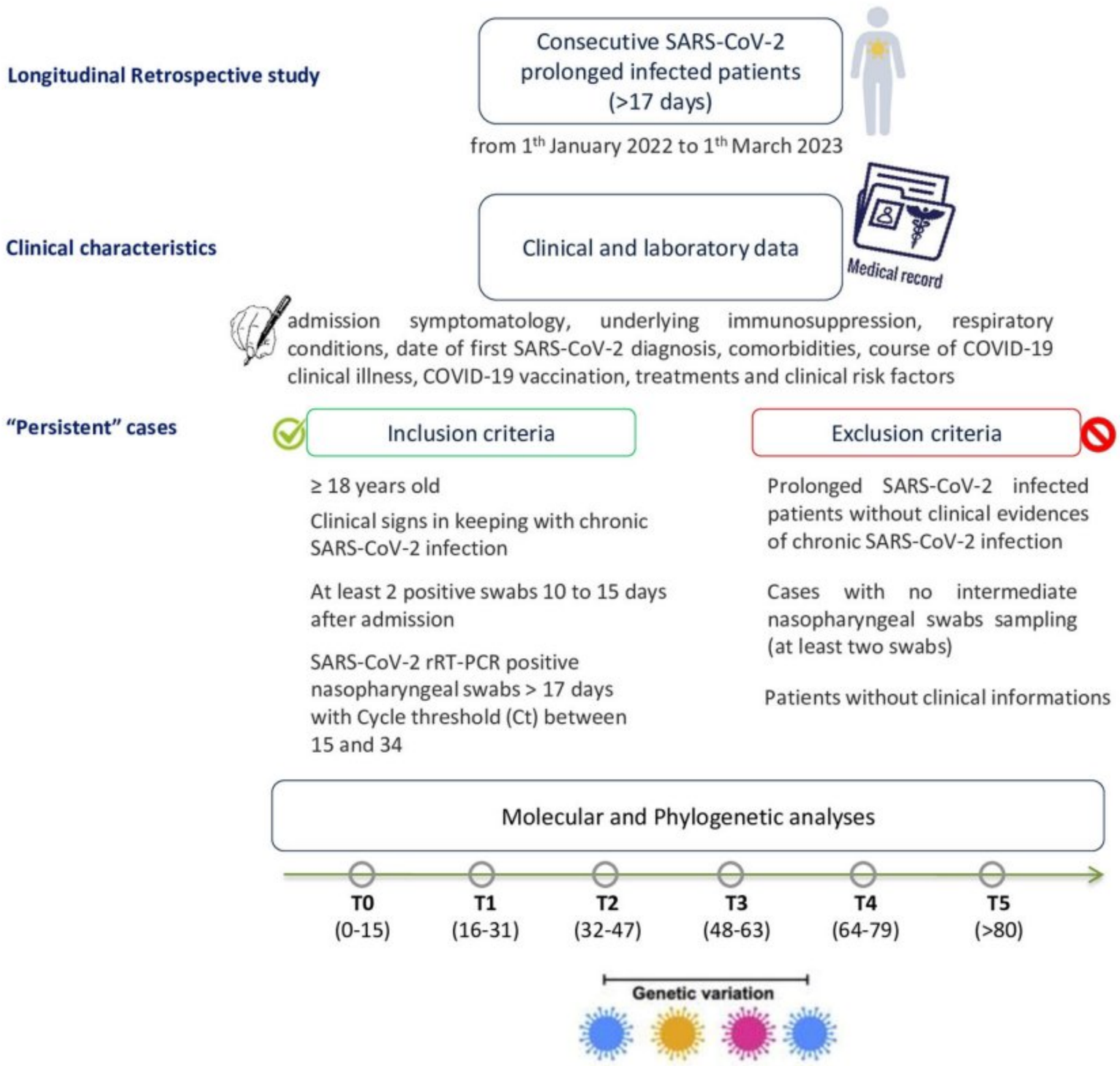


FIGURE 1 Workflow of study.

(WGS), viral RNA was extracted from 500 µL of nasopharyngeal swab sample media, using the automated system NUCLISENS® easyMAG® (bioMérieux). Each sample was re-tested for SARS-CoV-2 rRT-PCR to confirm the Ct values of each viral gene (ranging from 15 to 34 Ct values) and to verify the quality of sample storage at −80°C.

2.4 | SARS-CoV-2 WGS

WGS was performed by the Ion S5™ next-generation sequencing system (Thermofisher™ Scientific). Firstly, the concentration of viral RNA extracted were measured by the Qubit® 4.0 Fluorometer (ThermoFisher™ Scientific). Seven µL of viral RNA were retro-transcribed by Invitrogen™ SuperScript™ VILO cDNA Synthesis Kit (ThermoFisher™ Scientific, Waltham, MA, USA). Libraries were prepared using the Ion AmpliSeq SARS-CoV-2 Research Panel (ThermoFisher™ Scientific, Waltham, MA, USA), which consists of 237 amplicons ranging from 125 to 275 bp in length across the SARS-CoV-2 genome. Library preparation was performed manually according to the Ion AmpliSeq Library Kit Plus (ThermoFisher™ Scientific, Waltham, MA, USA) protocol. The final concentration of prepared cDNA libraries was determined on the Agilent 4200 System by the Agilent High Sensitivity DNA Assay (Agilent Technologies), following manufacturer's recommendations. Barcoded libraries were diluted to 30 pM and then loaded onto the Ion ChefTM Instrument (ThermoFisher™ Scientific) for emulsion PCR, enrichment, and loading onto the Ion S5 530 chip. To visualize metrics and quality of run sequencing, we performed the SARS-CoV-2_coverageAnalysis plugin in Ion Torrent Suite Software.

2.5 | Bioinformatic analysis

2.5.1 | Genome assembly

The assembly with reference process was performed by using the software Bowtie 2 v2.4.2 including a new filter that targeted reads with a nonzero QSEQ filter field.[32,33] Reads were then aligned and mapped to the reference genome Wuhan-Hu-1 (NC_045512.2) using an end-to-end read alignment approach (basic command line for bowtie2 [options]* −x < bt2-idx> {−1 < m1> −2 < m2> | −U < r > | --interleaved <i> | --sra-acc < acc> | b < bam > } - S [< sam >]). Consensus genomes were obtained by using the software Unipro UGENE v.35.[34] To visualize metrics and quality of run sequencing, we performed the SARS-CoV-2_coverageAnalysis plugin in Ion Torrent Suite Software. To evaluate the frequency of mutations variant calling was done with the plugin SARS_CoV_2_variantCaller v5.16.0.5 in Ion Torrent Suite Software.

2.5.2 | Sequence analyses and phylogenetic reconstruction

The initial evolutionary analysis of the isolated SARS-CoV-2 genomes was performed by constructing a data set comprising 120 European SARS-CoV-2 genomes, with 10 genomes representing each Omicron sub-lineage that circulated between January 2022 and March 2023: BA.1, BA.2, BA.3, BA.4, BA.5, BQ. 1, BA.2.75, CH1.1, EG.5, XBB, XBB 1.5, XBB 1.16.

Phylogenetic reconstruction has been performed following the methodology outlined by Scarpa et al.[35] Genomes were aligned using the L-INS-I algorithm within Mafft 7.471.[36] Subsequent refinement of alignments involved manual editing using Unipro UGENE v.35.[34] The software jModeltest 2.1.1 was employed, conducting a maximum likelihood optimized search, to identify the most suitable probabilistic model for genome evolution.[37] The software MrBayes 3.2.7 was utilized to explore the phylogenomic relationships among variants and isolates.[38] Two independently runs were executed, each comprising four Metropolis-coupled Markov-chain Monte Carlo (MCMCMC) simulations, involving one cold and three heated chains. These simultaneous runs spanned 5,000,000 generations, with tree sampling occurring every 1000 generations. The initial 25% of the 10,000 sampled trees were discarded as burn-in. Nodes exhibiting a posterior probability higher than 0.95 were considered as statistically supported. The convergence of chains after the runs was tested verifying that the Average Standard Deviation of Split Frequencies approached to 0,[38] and the Potential Scale Reduction Factor hovered around 1.[39] The phylogenetic tree was drawn and visualized by using FigTree 1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

Finally, a principal coordinate analysis (PCoA) was carried out through the utilization of the software GenAlEx 6.5 on the 27 newly sequenced genomes.[40] This analysis aimed to identify potential subdivisions within genetic clusters and assess the genetic diversity among the genomes.

This PCoA reconstruction was established on a pairwise *p*-distance matrix derived from genetic data. The genetic matrix was obtained using the R package APE,[41] which is implemented in the R statistical environment (www.cran.r-project.org).

2.5.3 | SARS-CoV-2 typing and mutational pattern analysis

Consensus sequences were analyzed with the Coronavirus Typing Tool v.1.25 on the Genome Detective Platform to generate the lineage and clade information.[42] The evaluation of genomic mutations in comparison to the reference sequence Wuhan-Hu-1 (NC_045512.2) was conducted by consulting the Stanford Coronavirus Resistance Database (CoV-RDB; https://covdb.stanford.edu) and the CoVsurver Mutation App (CoVsurver; https://gisaid.org/database-features/covsurver-mutations-app/).[43]

2.6 | Data availability

The newly generated WGS were submitted to GenBank® database.[44] All sequences can be retrieved from GenBank® database under accession numbers PP232072-PP232098.

3 | RESULTS

3.1 | Patients characteristics

Eight consecutive patients with persistent SARS-CoV-2 infection were included in the analysis. Baseline characteristics, clinical features and treatments are summarized in Table 1. The median age was 71 years (range: 57–83 years), of which 75% were male. The observed mean duration of nasopharyngeal viral shedding was 48.8 days (Standard Deviation: 21.14 days). Of these patients, seven (88%) were admitted to the Intensive Care Unit (ICU), while one to the Infectious Disease Unit (IDU). Among the eight patients, six (75%) required invasive mechanical ventilation (IMV) upon hospital admission. Of these, one patient underwent Extra-Corporeal Membrane Oxygenation (ECMO) due to refractory hypoxemia. Additionally, one patient received noninvasive ventilation (NIV), and another one received high flow nasal cannula (HFNC) oxygen therapy.

The 6 (75%) patients received three doses of COVID-19 vaccination, half of whom died; the remaining two (critically ill) patients (PZ_B and PZ_C patients) were unvaccinated and showed a long-term SARS-CoV-2 infection of 56 and 28 days, respectively (Table 1). Six patients (75%) received a 10-day course of remdesivir (RMD) antiviral therapy, either alone (3/6, 50%) or in combination with tixagemivab/cilgavimab (TIX/CIL) mAbs, casirivimab/imdevimab (CAS/IMD) mAbs, and/or tocilizumab (TCZ) mAbs (Table 1). All patients underwent active treatment involving high doses of corticosteroids for a minimum of 10 days. Various underlying primary clinical conditions were observed, including cardiological diseases (38%), autoimmune diseases (25%), hematologic malignancies (25%), and chronic diseases such as chronic cerebral vasculopathy (CCV). Comorbidities encompassed metabolic syndrome (38%), chronic kidney disease (CKD) (25%), premalignant clonal plasma cell disorders (25%), and cardiological disease in one patient (Table 1). In the ICU group, the majority of patients succumbed to septic shock (57%), whereas one patient (PZ_C patient) died for cardiogenic shock.

3.2 | Longitudinal sample collection and genotyping

To genotype the viral isolates in each persistent patient, the entire genome of SARS-CoV-2 of 27 longitudinal respiratory samples from eight persistent SARS-CoV-2 patients was sequenced (Table 2). The median depth of WGS was 1792 (range: 24.3–2664), with median genome coverage 98% of the viral genome. A median number of 295,119 reads (range: 84,613–405,380) was produced for each sample. According to Nextstrain analysis, was showed an Omicron lineages distribution consisting of BA.5 (22B clade) in 75% (6/8) of patients, followed by BA.1 (21K clade) and BA.2 (21L clade) in one patient, respectively (Table 2).

3.3 | Phylogenomic reconstruction

To investigate the evolutionary relationships between our 27 longitudinal isolates, sampled from eight persistent SARS-CoV-2 infected patients (Table 2) and the most frequent European SARS-CoV-2 lineages circulating in January 2022 to March 2023, phylogenomic reconstruction was carried out using MrBayes. In particular, the data set included our 27 viral sequenced genomes aligned and compared to 120 complete European SARS-CoV-2 genomes with high coverage downloaded from the GISAID reference database (called reference sequences), belonging to BA.1, BA.2, BA.3, BA.4, BA.5, BQ. 1, BA.2.75, CH1.1, EG.5, XBB, XBB 1.5, XBB 1.16 Omicron sublineages. Among them were also included all SARS-CoV-2 surveillance sequences of Omicron sublineages, that were isolated by the Microbiology Unit and subsequently sequenced by NGS method in the Laboratory of Molecular Genomics and Pathology.[8,9] The Phylogenetic tree reveals well-supported branches (posterior probabilities = 1). The patients PZ_A, PZ_C, PZ_D, PZ_E, PZ_F and PZ_H clustered with BA.5 reference sequences; while PZ_B and PZ_G belonged with BA.1 and BA.2, respectively (Supporting Information S1: Supplementary Figure 1).

3.4 | Intra-host SARS-CoV-2 evolution

The phylogenetic tree generated with MrBayes which exclusively includes our 27 longitudinal sequences is presented in Figure 2. It is noteworthy that each primary cluster in the phylogenetic tree is monophyletic, encompassing viral strains isolated from the same persistent SARS-CoV-2 patients. A significant host-related structure was identified among the 27 viral strains, independently of the different clinical characteristics of the patients. The relative phylogenetic branch length between the initial and final isolates within each highlights the rapid accumulation of genomic mutations within the host over the duration of the persistent infection. Notably, the unvaccinated patient PZ_B showed the longest branch length in last isolate (T3, 56 days of SARS-CoV-2 persistence), compared to the first isolate at T0 and other isolates, suggesting a high viral evolution rate within the host (Figure 2).

3.5 | Inter-host SARS-CoV-2 evolution

In the bidimensional graphical representation, each genome is depicted as a data point, and the relative distances among points

TABLE 1 Demographic and clinical characteristics of persistent SARS-CoV-2 infected patients.

N.	ID Patient	Age	Sex	Unit	Death	Viral persistence (days)	COVID-19 vaccination (Yes/No)	COVID-19 treatment	Biologic therapy	Primary disease	Comorbidity	Oxygen supply system
1	PZ_A	82	M	ICU	No	32	Yes	RMD	None	NTSEMI, APE	CKD, CHL	HFNC
2	PZ_B	69	M	ICU	Yes	56	No	none	None	CVD	DM2	IMV
3	PZ_C	77	F	ICU	Yes	28	No	RMD	None	MS	CKD, Epilepsy	IMV
4	PZ_D	62	F	ICU	Yes	60	Yes	TCZ + TIX/ CIL + RMD	Anti-CD20 mAbs+ IVIG	RA	MGUS, CVD	IMV, ECMO
5	PZ_E	65	M	ICU	Yes	45	Yes	RMD + TIX/CIL	None	CLL	CVD	IMV
6	PZ_F	83	M	ICU	Yes	47	Yes	RMD	None	CCV	DM2, CVD	IMV
7	PZ_G	57	M	IDU	No	92	Yes	RMD + CAS/IMD	Anti-CD20 mAbs	NHL	MGUS, DM2, Obesity, Thyroiditis	NIV
8	PZ_H	69	M	ICU	No	30	Yes	None	None	CVD	none	IMV

Abbreviations: APE, acute pulmonary edema; CAS, casirivimab; CCV, chronic cerebral vasculopathy; CHL, combined hyperlipidemia; CKD, chronic kidney disease; CLL, chronic lymphocytic leukemia; CVD, cardiovascular disease; DM2, diabetes mellitus type 2; ECMO, extracorporeal membrane oxygenation; HFNC, high flow nasal cannula; ICU, intensive care unit; IDU, infectious disease unit; IMD, imdevimab; IMV, invasive mechanical ventilation; IVIG, Intravenous immunoglobulin; MGUS, monoclonal gammopathy of undetermined significance; MS, multiple sclerosis; NHL, Non-Hodgkin lymphoma; NIV, full-face noninvasive ventilation; NTSEMI, non ST-segment elevation myocardial infarction; RA, rheumatic arthritis; RMD, remdesivir; TCZ, tocilizumab; TIX/CIL, tixagemivab/cilgavimab.

TABLE 2 Longitudinal retrospective collected nasopharyngeal swabs from eight persistent SARS-CoV-2 patients.

No.	ID Patient	First SARS-CoV-2 diagnosis	Admission date	Nasopharyngeal swabs sampling (Time points, day*)						PANGO lineage	Nextrain clade
				T0	*T1*	*T2*	*T3*	*T4*	*T5*		
1	PZ_A	08 Jan 2023	12 Jan 2023	12	23	32	—	—	—	BA.5	22B (Omicron)
2	PZ_B	21 Mar 2022	21 Mar 2022	7	18	41	56	—	—	BA.1	21K (Omicron)
3	PZ_C	30 Oct 2022	06 Nov 2022	15	—	28	—	—	—	BA.5	22B (Omicron)
4	PZ_D	09 Sep 2022	04 Oct 2022	—	26	39	52	64	—	BA.5	22B (Omicron)
5	PZ_E	10 Jan 2023	12 Jan 2023	—	18	32	48	—	—	BA.5	22B (Omicron)
6	PZ_F	03 Jan 2023	19 Jan 2023	—	16	32	47	—	—	BA.5	22B (Omicron)
7	PZ_G	09 Apr 2022	21 Apr 2022	2	20	34	48	64	79	BA.2	21 L (Omicron)
8	PZ_H	08 Jul 2022	15 Jul 2022	8	16	—	—	—	—	BA.5	22B (Omicron)

*Days after the first clinical diagnosis referred to in hospital records.

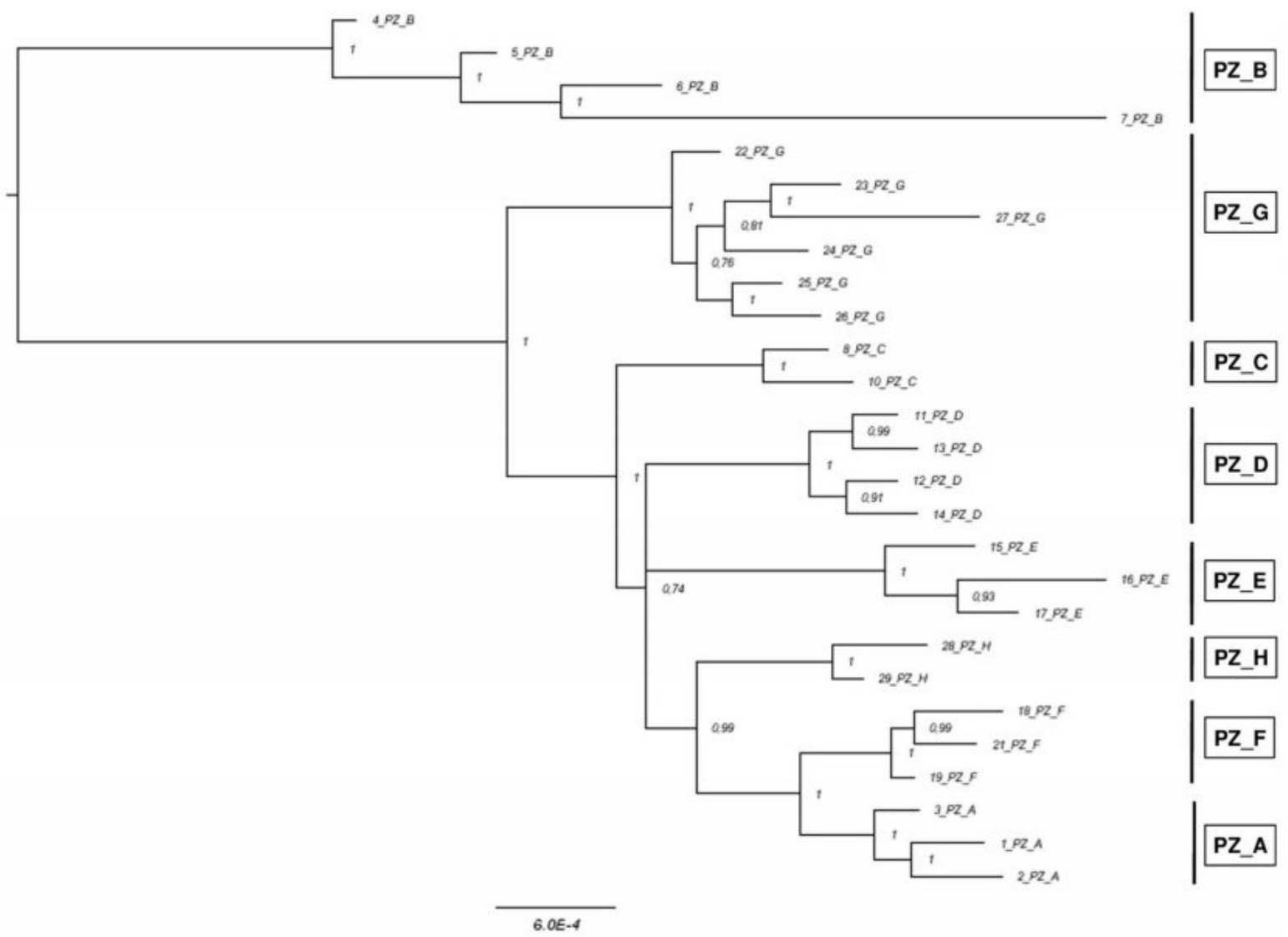


FIGURE 2 Phylogenetic tree of 27 longitudinal respiratory samples from eight persistent SARS-CoV-2 patients. The phylogenetic tree shows the genetic relationship between the whole genomes included in this study.

reflect the genetic similarity or difference between samples. This analysis, based on genetic distance matrix constructed with *p*-distance method, revealed the same primary subdivision observed through Bayesian inference (Figure 2). In the PCoA evolutionarily clades in the tree clustered together *per* host, forming unique groups. Consistent with the phylogenetic tree (Figure 2), the unvaccinated patient PZ_B exhibited high genetic divergence compared to other patients and among first and last

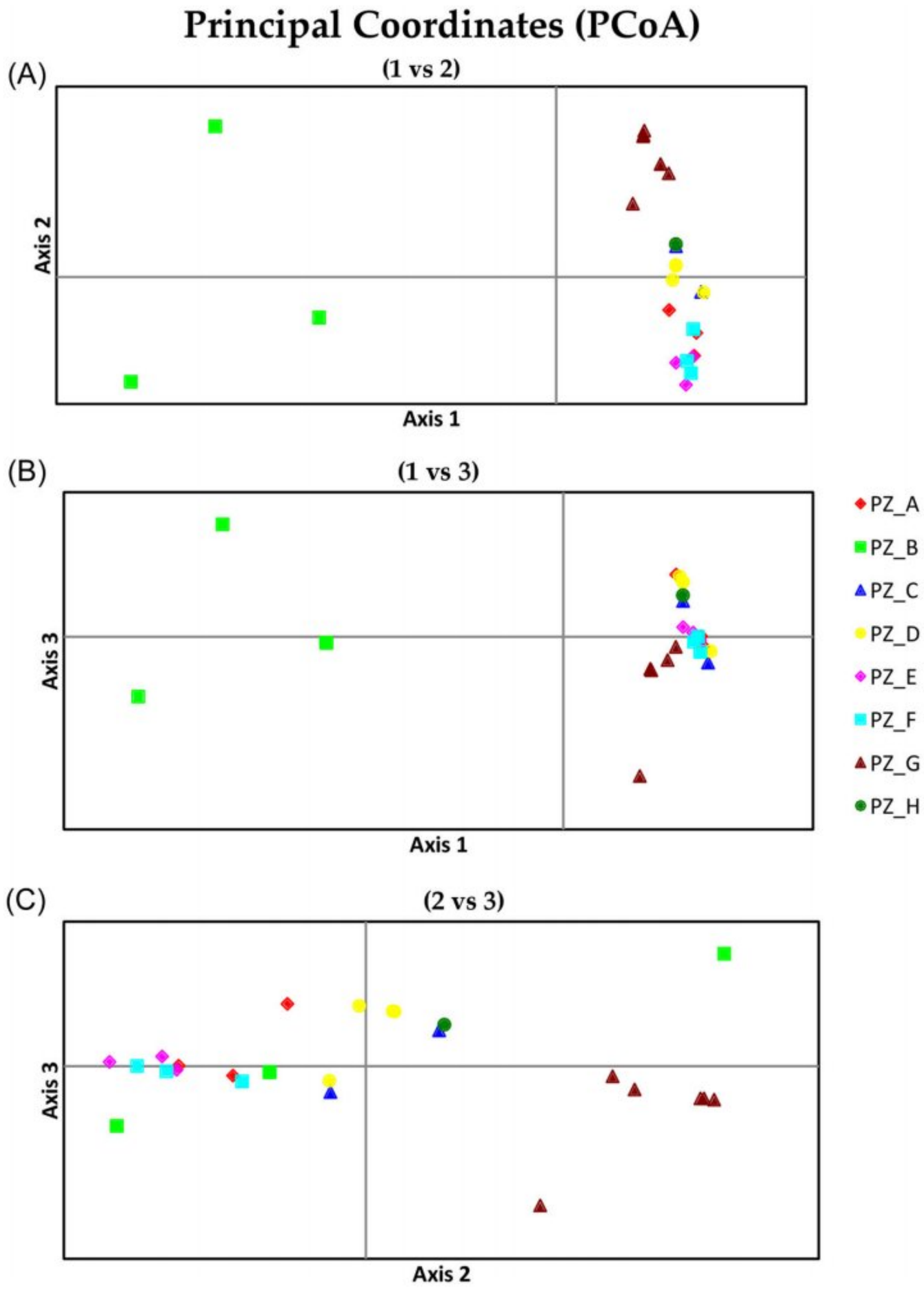


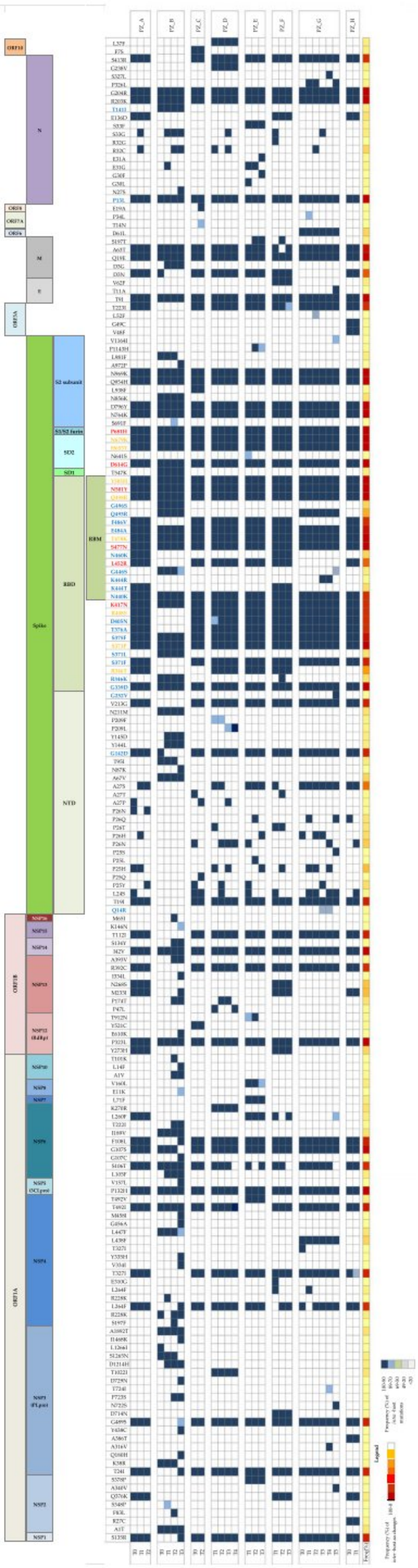
FIGURE 3 Analysis of principal coordinates (PCoA). (A) PCo1 versus PCo2, (B) PCo1 versus PCo3 and (C) PCo2 versus PCo3.

within-host viral isolates (Figure 3). The first three axes (Axis 1: 55.96; Axis 2: 13.52; Axis 3: 7.48) collectively explain 76.96% of the total variability.

3.6 | Temporal dynamics of intra-host SARS-CoV-2 variants evolution

The *intra*-host mutational pattern on all 27 longitudinal isolates during persistent SARS-CoV-2 infection were further examined. Overall, the highest percentage of aa substitutions was observed in the S glycoprotein domains (72/186, 39%), followed by the Open-Reading Frame 1A (ORF1A) with 61/186 (33%) aa changes, particularly in the nonstructural protein 3 (NSP3), encoding for a papain-like protease (PLpro), NSP4 and NSP6, the Nucleocapsid (N) protein with 18/186 (10%) and the ORF1b with 17/186 (9%) aa substitutions. The complete pattern of aa changes observed in the entire genome of longitudinal viral isolates is reported in Figure 4.

In the S1 subunit of SARS-CoV-2 S glycoprotein, several MOIs and aa substitutions involved in viral immune escape and high transmissibility have been documented (Figure 4). Various aa changes were identified in the "supersite" of the NTD (including residues 14–20, 140–158, and 245–264), which is targeted by several nAbs and plays a role in binding to glycan groups in the cellular glyco-environment.[45] Among them are Q14R, T19I, Y144L/D, Y145D, V213G, and G252V. The Q14R was acquired during long-term persistent infection of PZ_G patient (in T3 and T4); while G142D was present in all longitudinal respiratory samples. The mutation in residue 19 (T19I) was observed in all isolates at different time points, except for BA.1 patient (PZ_B). The Y144L/D and Y145D, unusual mutations with a global prevalence below 0.01%, were reported in unvaccinated PZ_B patient at 18, 41 and 56 days after infection. The aa substitution G252V was acquired only by patient PZ_G in the last



SARS-CoV-2 isolate (after 75 days of viral persistence), while V213G was observed in all isolates except for PZ_B patient. The L24S, P25Y/Q/H/L/S, P26N/T/Q/N, A27P/T/S, A67V, N87K, T95I, P209L/F, N211I noncanonical NTD-aa substitutions were also observed (Figure 4).

Several MOIs leading to higher infectivity, transmissibility, and antibody escape were also identified in the RBD (residues 331–528), which contains receptor-binding motif (RBM) at residues 436–506.[45,46] In particular, K417N, S477N and N501Y were observed in all divergent strains at different time points, while L452R MOI only in BA.5 patients (Figure 4). In addition to MOIs, S371L, S375F and T376A, aa changes which confer escape to nAbs, were shared by all divergent viral isolates.

Two additional MOIs (D614G and P681H) were also identified in all longitudinal isolates in the C-terminal domains, consisting of SD1 and SD2 domains (residues 528-686) and in the S1/S2 furin cleavage site.

In addition to MOIs, other aa substitutions associated with mAbs evasion were observed in S glycoprotein domains. In particular, G339D, R346K/T, S371F/L, S375F, T376A, D405N, N440K, K444T/R, G446S, N460K, E484A, F486V, Q493R, G496S and Q498R (Figure 4).[47–51]

The persistent SARS-CoV-2 patients PZ_E and PZ_D, who were treated with a combination of RDM *plus* TIX/CIL mAbs and TCZ, TIX/CIL mAbs *plus* RDM, respectively, exhibited aa changes R346T, S371F, L452R, E484A, F486V already in their first viral isolates. These conferred resistance to treatment with TIX/CIL.[47,49–52] Also, the onco-hematology patient PZ_G, treated with RMD *plus* CAS/IMD mAbs, carried S371F, N440K, G446S, E484A aa substitutions. These substitutions are associated with a reduced susceptibility to CAS/IMD mAbs.[47,49–52] In particular, the G446S aa substitution was observed only in the last viral isolate (T5, after 78 days of infection); while E484A was acquired twenty-2 days (T2) after the administration of antiviral therapy (Figure 4). The R346T, S373P, R408S, T478K, Q498R, Y505H, H655Y and N679K RBD-aa changes were also observed (Figure 4).[47,48,52,53]

FIGURE 4 Mutational pattern of longitudinal SARS-CoV-2 isolates from persistent infected patients. In bold were reported the MOIs (red), the aa changes involved in escaping to immune response or treatments (blue), the aa substitutions which lead to enhancing of ACE2 binding affinity (orange), noncanonical aa changes (black) occurred during persistent infection. The *inter*-host frequency (%) of mutations within 27 viral isolates is illustrated with scale of colors (from 100% to 0%). The *intra*-host frequency (%) of aa substitutions within 27 viral isolates is illustrated as follows: 100%–90% (blue), 89%–70% (light blue), 69%–50% (green), 49%–20% (dark gray) and 20%–0% (light gray). The details of all *intra*-host variants frequency are showed in Supporting Information S2: Supplementary Table 1.

In the N protein, two aa changes, P13L and T141I, were identified. These changes are known to impact T cell responsiveness during SARS-CoV-2 infection.[54,55] The P13L was observed in all isolates during persistent infection, while T141I only in PZ_B patient (Figure 4).

Notably, in other 12 viral proteins, different to S glycoprotein, were observed several aa changes shared by all patients at different time points of their persistent SARS-CoV-2 infection: S135R in NSP1; T24I and G489S in PLpro (NS3); L264F and T327I in NSP4; P132H in NSP5; S106T, G107S and F108L in NSP6; P323L in viral RNA-polymerase-RNA-dependent (RdRp, NSP12); R392C in NSP13; I42V in NSP14; T112I in NSP15; T223I in ORF3; T9I in envelope (E) protein and Q19E and A63T in Membrane (M) glycoprotein (Figure 4). Their impact is not yet known.

Interestingly, comparing first and last isolates among all chronic SARS-CoV-2 patients, we observed an evident accumulation of aa changes in the entire SARS-CoV-2 genome. In particular, the BA.1 unvaccinated patient (PZ_B), 56 days after first diagnosis (at T3), we observed 82 aa substitutions in the entire genome versus 50 aa changes in the first viral isolate (T0), collected 7 days after the initial diagnosis (Figure 4). The patient PZ_D treated with anti-CD20 mAbs, 5 months before the admission, showed a rate of evolution comparable to other persistent patients. Conversely, the onco-hemayological PZ_G patient treated with anti-CD20 mAbs up to 2 months before the SARS-CoV-2 infection, developed a high number of aa changes in the last divergent viral isolate (T5, 90 days after the onset of infection). In particular, were observed aa changes, such as Q340V in NSP2, N722S in the PLPro protease, L260F in NSP6, P25S in NTD, G252V and G446S in RBD, V1164I in S2 domain of S protein and T11A in E protein, never showed by other divergent isolates of persistent patients. After screening of minority mutations in our 27 viral isolates from persistent SARS-CoV-2 infected patients, no significant *intra*-host low-frequency mutations were identified across the entire set of reads, neither event of recombination from co-infection (for further details on the identified percentage values see Supporting Information S2: Supplementary Table 1).

4 | DISCUSSION

In this retrospective longitudinal study, including eight patients with persistent SARS-CoV-2 infection, we have delineated the *intra*- and *inter*-host genetic diversity, with focus on the evolution of the virus during the persistence of the infection and its mutational patterns. To our knowledge, this study represents the first instance of reporting such a substantial number of persistent SARS-CoV-2 patients infecting with Omicron sub-lineages. In addition, unlike to other similar studies,[22-26] we reported for each patient a complete clinical picture, such as COVID-19 form, treatments, comorbidities etc., to assign clinical relevance to molecular virological features reported in this retrospective longitudinal study. The most of previous studies have linked the evolution of SARS-CoV-2 genetic variants to persistent infections in a single case report or case series (around two individuals) with immunocompromising conditions. In our study we included a relatively small population of persistent SARS-CoV-2 infected patients, as defined by a nasopharyngeal viral shedding >17 days. This temporal criterion was established by two meta-analyses,[28,29] which reported mean duration of SARS-CoV-2 RNA shedding in upper respiratory tract of 17 days, with a 95% confidence interval ranging from 15.5 days to 18 days. Noteworthy, individuals with persistent SARS-CoV-2 infection exhibit notable characteristics, primarily marked by the presence of comorbidities and/or compromised immune system function, such as primary immunodeficiency or immunosuppressive therapies.[30,31] The majority of included patients required ICU admission for IMV support, delineating a population with critical form of COVID-19. One of them (PZ_D) was so critical that even required ECMO treatment and off-label use of RDM. It could be criticized the off-label administration of RDM, since recommended only in patients requiring conventional oxygen therapy, [56] or severe disease;[57] however, little data support potential benefits of RDM also in more severe patients.[58–60] Of note, 75% of patients were also vaccinated against SARS-COV-2% and 63% died. While anti-SARS-CoV-2 vaccines are crucial for preventing severe cases of COVID-19,[61] it's important to note that, as observed in our scenario, individuals with advanced age or underlying health conditions may still experience hospitalization and critical illness despite recent vaccination.[62,63] Moreover, the effectiveness of vaccination tends to diminish over time.[64,65] Additionally, our prior findings indicate a rise in the incidence of septic shock caused by multidrug resistant bacteria, such as *Acinetobacter baumannii*, during the pandemic, impacting the survival rates of affected patients.[66]

One of the key findings in our study revolves around the high evolutionary rate observed in divergent isolates within chronic SARS-CoV-2 persistent patients. In acute SARS-CoV-2 infection, RNA shedding is swiftly cleared within a few days to weeks, owing to the robust immune system response.[67] The large number of genetic variants arising during viral replication cycles is periodically purged through tight transmission bottlenecks.[5,6,26] In contrast, when SARS-CoV-2 persists for weeks to months in chronic infected patients with impaired immune response, severe clinical conditions or undergoing immunosuppressive treatments, there is a more rapid accumulation of mutations and a heightened acceleration of viral evolution, compared to the usual course of respiratory infection.[3,17-19,68] Indeed, the extended evolution of SARS-CoV-2 within a host, coupled with a sustained suboptimal immune response, vaccination, comorbidities, immunosuppressive therapies, and other clinical conditions, could elucidate the antigenic diversity observed among various intra-host lineages. This phenomenon aligns with previous observations, particularly notable in the case of Alpha and Omicron.[3,17–19,22–24]

The phylogenetic analysis, utilizing SARS-CoV-2 isolates from twenty-seven longitudinal samples also exhibited a different branch length in phylogenetic tree (as shown in Figure 2). This observation indicates a rapid *intra*-host diversification and emergence of new genetically divergent isolates during the persistence of the virus; in addition, it lends support to the notion that multiple novel variants

could emerge in chronic SARS-CoV-2 patients during persistent infection, potentially disseminating to others and undergoing further evolution. A parallel accelerated evolution has been previously observed in the Alpha and Omicron VOCs during their evolutionary trajectories.[3,18,19,25]

Notably, the unvaccinated patients with the BA.1 variant (PZ_B) showed a higher evolutionary rate compared to isolates from other persistent patients. This heightened rate of evolution was evident not only between isolates from different patients but also between the first (T0) and last (T3) viral variants within the same host. The absence of COVID-19 vaccination likely subjects the virus to selective pressure, prompting a higher mutation rate as an adaptive response to evade host innate immune responses. While it has been previously reported that vaccination may induce SARS-CoV-2 evolution, particularly in regions binding neutralizing antibodies located in the S protein, there is a selective pressure driving the accumulation of aa changes in T-cell epitopes and antibody-binding S domains even in the absence of vaccination, aiding the virus in evading host immune responses.[69]

Our phylogenomic reconstruction shows that the predominant variant among our persistently SARS-CoV-2 infected patients was Omicron BA.5, followed by Omicron BA.1 and BA.2 sub-lineages. These findings align with the data from Calabrian Surveillance.[8,9] Notably, there was an early decline of the BA.1 sub-lineage, evident as early as April 2022, contrasting with the broader Italian epidemiological context. This observation was exemplified by patient PZ_B, who contracted the BA.1 infection in March 2022. In April 2022, patient PZ_G was infected with the BA.2 Omicron sub-lineage, coinciding with the second peak of Omicron BA.2 in the Calabrian area in late March 2022. From July 2022 to March 2023, our patient cohort predominantly exhibited the Omicron BA.5 lineage, first identified in Calabria in June 2022.[8,9]

The Bayesian inference and PCoA analysis showed a host-based structuring among longitudinal viral isolates from chronic SARS-CoV-2 persistent patients. This structuring was observed regardless of the patients' primary disease, comorbidities, or the type of immunosuppressive therapy they underwent. As shown in Figure 2, the longitudinal divergent isolates from each persistent patient formed a distinct and unique cluster in the evolutionary tree. These clusters were closely related to, but clearly distinct from, other variants in the data set. The absence of recombination events or gene flows among ICU patients may reflect the positive effect of containment practices, such as the Personal Protective Equipment (PPE), limiting viral spread within the critical hospital area.

Another interesting finding is that a significant portion of the within-host divergent variants from included patients, frequently, exhibited the same combination of aa substitutions across the entire genome. Within the "supersite" of NTD, we identified various aa substitutions associated with reduced sensitivity to nAbs.[45,70] Specifically, Q14R and G142D, previously observed in the Delta and Omicron variants, have been shown to interfere with the neutralization of many NTD-binding Abs.[71,72] The D19I substitution, reported in all longitudinal samples except for patient PZ_B, appears to be associated with the removal of a N-glycosylation site at position 17, which could impact the antigenic and other properties of this strain.[73] Glycosylation of viral S protein is crucial for its folding, stability, and conformational dynamics, while glycans can also act as a shield for antibody epitopes.[73] Two additional NTD-aa changes at position 144 and 145 (Y144L/D, Y145D), conferring the ability to escape immune recognition, were also identified in the unvaccinated PZ_B patient at 18, 41 and 56 days after infection.

The RBD of S glycoprotein displayed a diverse array of MOIs, known to confer enhanced transmissibility of SARS-CoV-2 and reduced susceptibility to mAbs.[45,70] The S477N and N501Y MOIs shared by all divergent isolates, are common RBD mutations observed in Omicron variants. These mutations increase ACE2 binding strength and confer resistance to neutralization by the human convalescent sera.[47,51,74] Additionally, in vitro studies suggest that the combination of N501Y plus Q498R, present in each persistent patient, further enhances SARS-CoV-2 ACE2 binding affinity.[51]

Within the same RBD domain, all divergent isolates shared MOIs known to interfere with mAbs therapies. The K417N MOI, previously reported in the Beta (K417N), Gamma (K417T), and Omicron (K417N) variants, reduces susceptibility by >100-fold to Etesevimab (ETE) mAbs and ~30-fold to Casirivimab (CAS) mAbs.[47] The L452R MOI, reported in Delta and Omicron BA.4/5 variants, is associated with high-level resistance to Bamlanivimab (BAM) mAbs and a sixfold reduction in CIL mAbs susceptibility.[71,75–77]

In the SD2 and S1/S2 furin cleavage site, two additional MOIs (D614G and P681H) were found in all longitudinal isolates, which enhance viral replication,[78] and infectivity.[79,80] The D614G substitution is reported to raise viral replication in primary human airway tissues by increasing virion infectivity and stability.[78] Other key mutations in residues L452 or E484, present in all longitudinal isolates, enhanced fusion and infectivity in synthetic mutants, such as D614G + L452R, D614G + L452R + E484Q, compared to the original Wuhan-1 strain.[81] These findings suggest that the combination of different aa substitutions R346K/T, evolved during persistent SARS-CoV-2 infection, may collectively contribute to the acquisition of novel properties by the divergent strains.

In addition to MOIs, specific RBD aa changes (S371L, S375F, and T376A) shared by all divergent viral isolates, were found to confer the ability to escape several nAbs elicited by vaccination or SARS-CoV-2 infection, as demonstrated in in vitro studies.[48] The BA.1 and BA.2 patients shared the Q493R aa substitution at different time points in divergent isolates; this substitution was previously associated with prolonged SARS-CoV-2 infections of the B.1.1.7 (Alpha VOC) variant[82,83] and was linked to increased (≥25-fold) resistance to BAM/ETE and CAS.[84]

The P681H mutation at the S1/S2 furin cleavage site of the SARS-CoV-2 Spike protein has been reported in the Alpha, Delta, Kappa, Theta, and all sub-lineages of the Omicron variant. The positive charge associated with this mutation appears to influence virus tropism by increasing S1/S2 cleavage in human airway epithelial cells. Moreover, the double mutant D614G + P681R results in enhanced infectivity.[79,80]