SARS-CoV-2 Sequence Characteristics of COVID-19 Persistence and Reinfection

Manish C. Choudhary[1*], Charles R. Crain[1,2*], Xueting Qiu[3], William Hanage[3], Jonathan Z. Li,[1]

[1]Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

[2]Northeastern University, Boston, MA, USA

[3]Harvard T.H. Chan School of Public Health, Boston, MA, USA

*Authors contributed equally

Corresponding Author:

Jonathan Li, MD

Brigham and Women's Hospital

Harvard Medical School

jli@bwh.harvard.edu

***Key words*:** SARS-CoV-2, persistent COVID-19, reinfection, sequence analysis, immunosuppression

Funding: This study was funded in part by the NIH grant 106701.

Disclosures:

Dr. Li has consulted for Abbvie.

ABSTRACT

Background. Both SARS-CoV-2 reinfection and persistent infection have been described, but a systematic assessment of mutations is needed. We assessed sequences from published cases of COVID-19 reinfection and persistence, characterizing the hallmarks of reinfecting sequences and the rate of viral evolution in persistent infection.

Methods. A systematic review of PubMed was conducted to identify cases of SARS-CoV-2 reinfection and persistent infection with available sequences. Amino acid changes in the reinfecting sequence were compared to both the initial and contemporaneous community variants. Time-measured phylogenetic reconstruction was performed to compare intra-host viral evolution in persistent COVID-19 to community-driven evolution.

Results. Fourteen reinfection and five persistent infection cases were identified. Reports of reinfection cases spanned a broad distribution of ages, baseline health status, reinfection severity, and occurred as early as 1.5 months or >8 months after the initial infection. The reinfecting viral sequences had a median of 9 amino acid changes with enrichment of changes in the S, ORF8 and N genes. The number of amino acid changes did not differ by the severity of reinfection and reinfecting variants were similar to the contemporaneous sequences circulating in the community. Patients with persistent COVID-19 demonstrated more rapid accumulation of mutations than seen with community-driven evolution with continued viral changes during convalescent plasma or monoclonal antibody treatment.

Conclusions. SARS-CoV-2 reinfection does not require an unusual set of circumstances in the host or virus, while persistent COVID-19 is largely described in immunosuppressed individuals and is associated with accelerated viral evolution as measured by clock rates.

BACKGROUND

After resolution of coronavirus disease 2019 (COVID-19) following SARS-CoV-2 infection, antibodies against SARS-CoV-2 persist in the majority of patients for 6 months or more [1]. Despite this, there have now been a number of reports of COVID-19 reinfection that span a broad range of age groups, time frame between infections and disease severity [2-8]. There remains a great deal of uncertainty over the viral characteristics of reinfection cases, including the degree of sequence heterogeneity and the location of new mutations between the initial and reinfecting variants. In addition, the diagnosis of COVID-19 reinfection has been complicated by the increasing reports of persistent COVID-19 infection, especially in immunosuppressed individuals. Like reinfection cases, persistent COVID-19 can also span the range of disease severity, from asymptomatic to severe disease, and recurrent symptoms can last for months [9-12]. Differentiating between persistent and reinfection can be challenging, and little is known about differences in the location of SARS-CoV-2 mutations in these scenarios. We performed an analysis of SARS-CoV-2 sequences from published cases of COVID-19 reinfection and persistence, characterizing the hallmarks of reinfecting sequences and the rate of viral evolution in persistent infection.

METHODS

Data search and selection criteria

We conducted a systematic literature review in Pubmed through February 5, 2021 for cases of persistent COVID-19 using the search term "((covid or sars-CoV-2) AND (persistent or persistence or prolonged)) AND (sequence or evolution)". A search for COVID-19 reinfection reports was made using the terms "(covid or sars-CoV-2) AND (reinfection)". Both peer-reviewed and preprint results were evaluated. We used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) for reviewing literature and for reporting search results. For cases of reinfection, papers were included if the authors described it as a case of reinfection diagnosed >30 days after the initial infection and if whole genome SARS-CoV-2 sequences or sites of mutations relative to a reference sequence (e.g., Wuhan-Hu-1) from both infection time-points were available. Of the 249 results from the search, 10 articles met the inclusion criteria and were included in the present report along with 2 additional preprints that were identified (Supplemental Figure 1A).

Persistent cases were included if the authors described it as a case of persistent COVID-19 infection and if longitudinal whole genome SARS-CoV-2 sequences were available. The search returned 116 results, 4 of which met the inclusion criteria and were included in the present report along with one other preprint (Supplemental Figure 1B). Only sequences from direct patient nasopharyngeal or anterior nasal swabs were included in our analysis.

Sequences were downloaded and analyzed for mutations using NextClade (https://clades.nextstrain.org/) and snp-sites (https://github.com/sanger-pathogens/snp-sites). The degree of re-infection severity, either more or less severe compared to the first infection, was classified based on the reported symptoms and reported severity.

Sequencing dataset compilation and phylogenetic tree construction

The sequencing dataset contained a total of 266 globally representative SARS-CoV-2 genomes selected from GISAID and sequences from the reinfection and persistence cases (Supplemental Methods). The sampled sequences were chosen to be representative of global sequence diversity throughout the time course of the pandemic. Sequences of variants of concern B.1.1.7 and B.1.351 were also included. Nucleotide sequence alignment was performed using MAFFT (Multiple Alignment using Fast Fourier Transform) [13]. Best-fit nucleotide substitution was calculated using model selection followed by maximum likelihood (ML) phylogenetic tree construction using IQ-Tree with 1000-bootstrap replicates [13].

Mutation analysis

For reinfection cases, mutations were determined in two ways. First, amino acid changes were identified for the reinfection sequences relative to the first infection sequence. The frequency of amino acid changes within each gene was compared to the frequency of changes in the remainder of the genome by X^2 test with Yates correction. The relationship between disease severity and number of amino acid changes in the genome was assessed using a Mann Whitney test. Second, to identify unique characteristics of reinfecting viruses, each of the first and reinfection sequences were compared to circulating sequences in the community as defined by the same NextStrain clade sampled within one month from the same geographic location uploaded to GISAID (https://www.gisaid.org/; Supplemental Table 1, Supplemental Methods). Rare mutations were determined as polymorphisms that were present only in the reinfecting sequence (not the initial variant) and found in less than 1% of contemporaneous community sequences. Mutation locations are graphically represented in Circos plots [14].

For persistent infections, sequence changes were assessed at two time intervals: sequences obtained before or after convalescent plasma or monoclonal antibody treatment. Sequences sampled before convalescent plasma or antibody treatment were compared to the first sequence sampled. For sequences sampled after convalescent plasma or antibody

118 treatment, amino acid changes were determined relative to the last pre-treatment sequence.
119 Linear regression was used to estimate the rate of viral changes within each of the two intervals.
120 To compare the rate of intra-host viral evolution in persistent COVID-19 to the rate of
121 community-driven evolution, we performed time-measured phylogenetic reconstruction as noted
122 below.
123
124 **Time-measured phylogenetic analysis**
125 The temporal signal of the ML tree was examined in TempEst [15] regressing on root-to-tip
126 divergence, and outliers were inspected in the distribution of residuals. A high degree of
127 clock-like behavior in the whole dataset was observed (R^2 = 0.726) in root-to-tip
128 regression analysis with the slope rate as 7.33E-4 and the rough ancestral time of the
129 sample was calculated as 2019.88. This suggests that the most recent common
130 ancestor of the data set composed of only sequences from the persistent cases
131 provides a realistic temporal signal and it is appropriate for an estimation of temporal
132 parameters. No outliers were found in this sample. To compare the evolutionary rates
133 between the reported persistent infections and the general population infections, time-measured
134 phylogenetic reconstruction was conducted in Bayesian Evolutionary Analysis Sampling Trees
135 (BEAST) v1.10.4 [16]. Five partitions, including four persistent patients and the global
136 sequences, were used as separate groups of taxa, to estimate separate evolutionary rates. Due
137 to large uncertainties with small samples, patients with only two viral sequences were excluded
138 from this analysis. A general time reversible (GTR) model was applied with gamma-distributed
139 rate variations among sites. A lognormal relaxed molecular clock was used with an initial mean
140 of 0.0008 and a uniform prior ranging from 0.0 to 1.0. A logistic growth tree prior was applied.
141 Three independent Bayesian Markov Chain Monte Carlo (MCMC) chains of 100 million
142 generations were performed with a sampling step every 10,000 generations to yield 10,000

143 trees per run. To ensure a sufficient effective sample size ESS > 200, the convergence of three
144 runs was diagnosed in Tracer v 1.7.1 (http://tree.bio.ed.ac.uk/software/tracer/) for all
145 parameters. LogCombiner v1.10.4 as part of the BEAST software package was used to
146 combine the multiple runs to generate log and tree files after appropriate removal of the burn-in
147 from each MCMC chain. The comparison of the evolutionary rates from the combined log file is
148 analyzed and visualized in R v4.0.2 (https://www.r-project.org/).
149
150 **Statistical analysis**
151 Nonparametric Wilcoxon rank sum or matched pairs signed rank tests were used to compare
152 the number of amino acid changes between sequences. Statistical analyses were performed
153 using GraphPad Prism 9 (GraphPad Software, San Diego, CA).

RESULTS

Sequence analysis of reinfection cases

A total of fourteen cases from twelve reports were included in this analysis (Table 1) [2-5, 7, 8, 17-22]. A broad range of age groups were represented and 79% were under the age of 65 years. Most (71%) of the cases had no reported comorbidities and none of the patients were immunocompromised. The interval between diagnosis of the first infection and the second infection ranged from 46 days to 250 days with a median of 110 days. Four patients had more severe illness during the second infection, while five had less severe symptoms on reinfection, including two who were asymptomatic on reinfection. Two cases were asymptomatic in both infections, one case reported the same severity for both infections and no information on infection severity was available for two cases (Table 1). Four cases reported reinfection with a virus from the same clade.

For the reinfection cases, phylogenetic analysis demonstrated distinct branches for the two sequences. We compared amino acid changes in the reinfecting viral sequence compared to the initial sequence and found a median of 9 amino-acid changes (range 6-20) compared to the original sequence (Figure 2A). The amino acid changes were distributed across the SARS-CoV-2 genome, with significantly lower frequencies of changes in ORF1a ($P=0.008$) and ORF3a ($P=0.03$), and higher frequencies of changes in S ($P=0.02$), ORF8 ($P<0.001$), and N ($P=0.003$) (Figure 2B). Each reinfection case had at least one substitution or deletion in the S gene (Supplemental Table 2). Next, we assessed whether reinfection with a more divergent second virus resulted in more severe disease. We found no significant differences in the number of amino acid changes in the reinfecting virus compared to the original viral variant when categorized by the severity of the reinfection (Figure 2C). Both the initial and reinfecting SARS-CoV-2 variants were similar to the sequences circulating in the community at the time of reinfection. The reinfecting viruses harbored fewer rare mutations compared to the initial

179 infecting variant, with only a median of 1 rare amino acid compared to circulating variants in the
180 community (Figure 2D-E).
181
182 **Reclassification of one case**
183 Mulder*, et al.* described as a case of reinfection in an 89-year-old female with Waldenström
184 macroglobulinemia treated with B cell-depleting therapy [6]. The patient had two symptomatic
185 episodes separated by 59 days with no RT-PCR testing between the two episodes and
186 recrudescent symptoms shortly after receiving chemotherapy. The virus from the second
187 timepoint clustered with the initial sequence (P6 in Figure 1) and both had the same two
188 nucleotide substitutions and the same deletion in ORF1a relative to contemporaneous
189 sequences. Neither substitution nor deletion was observed in other community sequences
190 sampled at the time of the second episode. Given these features, we have classified this case
191 as a persistent infection for this analysis.
192
193 **Sequence analysis of persistent COVID-19 cases**
194 A total of six reports describing persistent infection were retrieved from our literature search. Of
195 these six cases, all but one had B cell immunodeficiency [6, 9-11, 23]. Four of these five were
196 treated with B cell-depleting therapy for lymphoma or autoimmune disorders, while one had
197 chronic lymphocytic lymphoma with acquired hypogammaglobulinemia (Table 2). The patient
198 without immunodeficiency was an outlier: he was a young patient without a known
199 immunosuppressing condition and with more than 180 days between symptomatic episodes
200 [24]. Phylogenetic analysis showed the two sequences arising from the same root (P3, Figure
201 1), but uncertainty about whether this case represented reinfection or persistent infection led us
202 to exclude it from this analysis. For the 5 participants included in this analysis as persistent
203 infection, the median length of infection was 154 days and most cases (4/5) ended in death.
204 One patient had asymptomatic disease throughout [10]. Three patients were treated with

convalescent plasma at least once during their illness [10, 11, 23], and one patient was treated with the monoclonal antibodies casirivimab and imdevimab [9].

Phylogenetic analysis revealed that, for each of the five patients, sequences formed a distinct cluster (Figure 1). New mutations emerging over time were detected in all of the persistent COVID-19 patients with further changes identified after treatment with convalescent plasma or monoclonal antibodies (Figure 3). The rate of viral evolution was plotted for each patient both for the interval before and after convalescent plasma/antibody treatment. Overall, the rate of amino acid changes over time appeared faster before treatment (Figure 4a), but treatment with convalescent plasma or antibody cocktail treatment were insufficient to halt intra-host viral evolution (Figure 4b).

We also performed time-measured phylogenetic reconstruction with the pre-treatment persistent sequences to compare the rate of intra-host viral evolution in persistent COVID-19 to the rate of community-driven evolution. SARS-CoV-2 evolution was faster in these persistent infection individuals compared to the rate in the general public population, though substantial uncertainties are shown in these estimates given the limited sequence sampling for each case (Figure 4C).

DISCUSSION

We conducted a systematic review and pooled analysis of sequences from reports of COVID-19 reinfection and persistent infection. Reports of reinfection cases demonstrate a wide range of situations: spanning a broad distribution of ages (from individuals in their 20s to >70 years), baseline health status, reinfection severity compared to the initial infection, and occurring as early as 1.5 months or >8 months after the initial infection. Common explanations for the presence of reinfection involves either waning SARS-CoV-2 antibodies or the presence of viral escape mutations [25, 26]. While most cases of SARS-CoV-2 reinfection did involve infection with a different clade (including the variant B.1.1.7), it is noteworthy that mutations were identified throughout the genomes and the frequency of mutations within the S gene was only modestly higher than the rate across the entire genome. In addition, individuals with more severe reinfections did not have significantly greater frequency of S gene mutations. Finally, the presence of rare mutations was uncommon in the reinfecting virus, which largely mirrored the contemporaneously circulating variants in the region of infection. The interpretation of this analysis is limited by the lack of immune profiling, but the results suggest that reinfection does not require an unusual set of circumstances with respect to the reinfecting virus.

While the number of immunosuppressed individuals with available sequences remains limited, the results suggest that the rate of viral evolution, meaning the rate at which non-synonymous mutations lead to changes in protein sequences, is accelerated within immunosuppressed individuals. In addition, treatment with convalescent plasma or monoclonal antibody cocktails was insufficient to fully halt of viral evolution and the emergence of viral escape has been documented [23, 27]. The results raise the possibility that novel variants, including those harboring escape mutations against current treatments, could arise from immunosuppressed individuals and suggest that immunosuppressed individuals should be a focus of public health efforts. Amongst the current reports of persistent COVID-19, B-cell dysfunction appears to be a common thread, including in reports that were not included in this

analysis due to a lack of available full-length sequences [28-32]. It is important to note, though, that T cell function may also play a role in protection against SARS-CoV-2 [33] and a subset of these patients also included concurrent suppression of other aspects of the immune response. Additional studies are needed to fully define the type and intensity of immunosuppression that would place patients at greatest risk of persistent COVID-19.

Two factors generally differentiated between reinfection and persistent infection scenarios: first, reinfections have so far been largely described in immunocompetent individuals while the majority of persistent COVID cases have been in immunosuppressed patients. Secondly, phylogenetic analysis can generally differentiate between reinfection and persistent infection, especially in cases where persistent infection allowed the longitudinal collection of >2 sequences. However, given the slow rate of SARS-CoV-2 evolution and limited viral diversity [34], it can be challenging to differentiate between reinfection and persistent infection, especially in situations with limited sampling and/or duration between samples. Overall, our results demonstrate the need to further explore factors that increase the risk of breakthrough reinfections and persistent COVID-19. This line of investigation will have important implications for preventing the rise of novel variants and the durability of current available vaccines.

Acknowledgements

We thank Jeremy Luban and Ronald Bosch for their feedback and discussion.

REFERENCES

1. Dan JM, Mateus J, Kato Y, et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. Science **2021**.
2. Van Elslande J, Vermeersch P, Vandervoort K, et al. Symptomatic SARS-CoV-2 reinfection by a phylogenetically distinct strain. Clin Infect Dis **2020**.
3. To KK, Hung IF, Ip JD, et al. COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing. Clin Infect Dis **2020**.
4. Tillett RL, Sevinsky JR, Hartley PD, et al. Genomic evidence for reinfection with SARS-CoV-2: a case study. Lancet Infect Dis **2021**; 21(1): 52-8.
5. Selhorst P, Van Ierssel S, Michiels J, et al. Symptomatic SARS-CoV-2 reinfection of a health care worker in a Belgian nosocomial outbreak despite primary neutralizing antibody response. Clin Infect Dis **2020**.
6. Mulder M, van der Vegt D, Oude Munnink BB, et al. Reinfection of SARS-CoV-2 in an immunocompromised patient: a case report. Clin Infect Dis **2020**.
7. Harrington D, Kele B, Pereira S, et al. Confirmed Reinfection with SARS-CoV-2 Variant VOC-202012/01. Clin Infect Dis **2021**.
8. Goldman JD, Wang K, Roltgen K, et al. Reinfection with SARS-CoV-2 and Failure of Humoral Immunity: a case report. medRxiv **2020**.
9. Choi B, Choudhary MC, Regan J, et al. Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. N Engl J Med **2020**; 383(23): 2291-3.
10. Avanzato VA, Matson MJ, Seifert SN, et al. Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Individual with Cancer. Cell **2020**; 183(7): 1901-12 e9.
11. Baang JH, Smith C, Mirabelli C, et al. Prolonged Severe Acute Respiratory Syndrome Coronavirus 2 Replication in an Immunocompromised Patient. J Infect Dis **2021**; 223(1): 23-7.
12. Kemp SA, Collier DA, Datir R, et al. Neutralising antibodies in Spike mediated SARS-CoV-2 adaptation. medRxiv **2020**.
13. 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.
14. Krzywinski MI, Schein JE, Birol I, et al. Circos: An information aesthetic for comparative genomics. Genome Research **2009**.
15. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol **2016**; 2(1): vew007.
16. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol **2012**; 29(8): 1969-73.
17. Prado-Vivar B, Becerra-Wong M, Guadalupe JJ, et al. A case of SARS-CoV-2 reinfection in Ecuador. Lancet Infect Dis **2020**.
18. Resende PC, de Vasconcelos RHT, Arantes I, et al. Spike E484K mutation in the first SARS-CoV-2 reinfection case confirmed in Brazil. Virologica **2021**.
19. Colson P, Finaud M, Levy N, Lagier JC, Raoult D. Evidence of SARS-CoV-2 re-infection with a different genotype. J Infect **2020**.
20. Nonaka CKV, Franco MM, Graf T, et al. Genomic Evidence of a Sars-Cov-2 Reinfection Case With E484K Spike Mutation in Brazil. Preprintsorg **2021**.
21. Gupta V, Bhoyar RC, Jain A, et al. Asymptomatic reinfection in two healthcare workers from India with genetically distinct SARS-CoV-2. Clin Infect Dis **2020**.
22. Abu-Raddad LJ, Chemaitelly H, Malek JA, et al. Assessment of the risk of SARS-CoV-2 reinfection in an intense re-exposure setting. Clin Infect Dis **2020**.

23. Kemp SA, Collier DA, Datir RP, et al. SARS-CoV-2 evolution during treatment of chronic infection. Nature **2021**.
24. Molina LP, Chow SK, Nickel A, Love JE. Prolonged Detection of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) RNA in an Obstetric Patient With Antibody Seroconversion. Obstet Gynecol **2020**.
25. Wibmer CK, Ayres F, Hermanus T, et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. bioRxiv **2021**.
26. Weisblum Y, Schmidt F, Zhang F, et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. Elife **2020**; 9.
27. Starr TN, Greaney AJ, Addetia A, et al. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. Science **2021**; 371(6531): 850-4.
28. Camprubi-Ferrer D, Gaya A, Marcos MA, et al. Persistent replication of SARS-CoV-2 in a severely immunocompromised treated with several courses of remdesivir. Int J Infect Dis **2020**.
29. Hensley MK, Bain WG, Jacobs J, et al. Intractable COVID-19 and Prolonged SARS-CoV-2 Replication in a CAR-T-cell Therapy Recipient: A Case Study. Clin Infect Dis **2021**.
30. Martinot M, Jary A, Fafi-Kremer S, et al. Remdesivir failure with SARS-CoV-2 RNA-dependent RNA-polymerase mutation in a B-cell immunodeficient patient with protracted Covid-19. Clin Infect Dis **2020**.
31. Helleberg M, Niemann CU, Moestrup KS, et al. Persistent COVID-19 in an Immunocompromised Patient Temporarily Responsive to Two Courses of Remdesivir Therapy. J Infect Dis **2020**; 222(7): 1103-7.
32. Sepulcri C, Dentone C, Mikulska M, et al. The longest persistence of viable SARS-CoV-2 with recurrence of viremia and relapsing symptomatic COVID-19 in an immunocompromised patient – a case study. medRxiv **2021**: 2021.01.23.21249554.
33. McMahan K, Yu J, Mercado NB, et al. Correlates of protection against SARS-CoV-2 in rhesus macaques. Nature **2020**.
34. Rausch JW, Capoferri AA, Katusiime MG, Patro SC, Kearney MF. Low genetic diversity may be an Achilles heel of SARS-CoV-2. Proc Natl Acad Sci U S A **2020**; 117(40): 24614-6.

346 **Figure 1.** Maximum-likelihood phylogenetic tree of sequences from persistent COVID-19 cases
347 (P1-P6), COVID-19 reinfection cases (R1-R12), the variants of concern B. 1.1.7 and B.1.351,
348 and globally sampled sequences from GISAID.

349 **Figure 2**. Comparison of viral sequences from reinfection cases. (A) Circos plot showing
350 location of nucleotide changes in the re-infecting sequence for each of the 14 cases. Inner ring
351 indicates nucleotide position. Synonymous changes are in green, nonsynonymous changes in
352 orange, deletions in black. (B) Amino acid (AA) substitution frequency pooled across all
353 reinfection cases for each SARS-CoV-2 gene. Dashed line indicates global substitution
354 frequency across the whole genome. Substitution frequency for each gene was compared to the
355 substitution frequency in the rest of the genome using a X^2 test with Yates correction. * <0.05, **
356 <0.01 and ***<0.001 (C) Amino acid changes in the second infection relative to the first infection
357 by clinical disease severity. Mutations shown for the whole genome, S gene, and receptor
358 binding domain (RBD) P=0.8, Mann Whitney test. (D) Circos plot showing location of nucleotide
359 mutations from the second infection relative to other viruses circulating at the same time in the
360 same geographic region. Only rare mutations present in <1% of contemporaneous community
361 sequences are shown. (E) Number of rare amino acids at each time point relative to circulating
362 sequences in the community. P=0.01, Wilcoxon matched pairs signed rank test. ORF: open
363 reading frame, S: Spike, E: Envelope, M: Membrane, N: Nucleocapsid.

364 **Figure 3.** Circos plots mapping mutations for the persistent COVID-19 cases. Mutations are
365 marked relative to the first timepoint. P1 * shows N501Y mutation (days 128-152). ** shows
366 E484K mutation (days 75 and 81 only). P5 *** shows Δ69/Δ70 mutation. Synonymous
367 mutations in green, nonsynonymous mutations in orange, deletions in black. Red text indicates
368 timepoint was sampled after first convalescent plasma or antibody cocktail treatment. Inner ticks
369 indicate nucleotide position. ORF: open reading frame, S: Spike, E: Envelope, M: Membrane, N:
370 Nucleocapsid.

371 **Figure 4.** SARS-CoV-2 evolutionary rate in the persistent COVID-19 cases. (a) Amino acid
372 changes in samples taken prior to convalescent plasma or monoclonal antibody treatment
373 relative to first sampled sequence in each persistently infected patient. Regression line and 95%
374 confidence interval of the slopes are shown. (b) Amino acid changes in samples taken after
375 convalescent plasma or monoclonal antibody treatment relative to last sample taken prior to
376 treatment in each persistently infected patient. Regression line and 95% confidence interval of
377 the slope are shown. (c) Substitution rate (substitutions per site per year) of sampled global
378 SARS-CoV-2 sequences relative to four persistent patients based on Markov chain Monte Carlo
379 time-measured phylogenetic reconstruction.
380

Table 1. Reinfection cases

Patient	Authors	Publication/Pre-print server (year)	Age	Sex	Comorbidities	Time between infections (days)	Second infection severity	First infection clade	Second infection clade
R1	Selhorst, *et al.*	*Clin. Infect. Dis.* (2020)	39	F	None	185	Less	19A	20A
R2	To, *et al.*	*Clin. Infect. Dis.* (2020)	33	M	None	144	Less	19A	20E
R3	Prado-Vivar, *et al.*	*Lancet Infect. Dis.* (2021)	46	M	None	63	More	20A	19B
R4	Tillett, *et al.*	*Lancet Infect. Dis.* (2020)	25	M	None	48	More	20C	20C
R5	Goldman, *et al.*	medRxiv	60-69	N/A	Emphysema, hypertension	139	Less	19B	20A
R6	Resende, *et al.*	Virological	37	F	None	116	Similar	20B	20B
R7	Harrington, *et al.*	*Clin. Infect. Dis.* (2021)	78	M	Diabetic nephropathy with hemodialysis, COPD, sleep apnea, ischemic heart disease	250	More	19A	B.1.1.7
R8	Van Elslande, *et al.*	*Clin. Infect. Dis.* (2020)	51	F	Asthma	93	Less	20B	19B
R9	Colson, *et al.*	*J. Infect.* (2020)	70	M	None	105	Less	20A	20A.EU2
R10	Nonaka, *et al.*	Preprints	45	F	None	147	More	20B	20B
R11-1	Gupta, *et al.*	*Clin. Infect. Dis* (2020)	25	M	None	108	Both asymptomatic	19A	20A
R11-2			28	F	None	111	Both asymptomatic	20A	20A
R12-1	Abu-Raddad, *et al.*	*Clin. Infect. Dis* (2020)	25-29	M	N/A	46	N/A	19A	20A
R12-2			40-44	M	N/A	71	N/A	19A	20A

Table 2. Persistent cases

Patient	Case	Publication (year)	Age	Sex	Underlying conditions	Immunosupressants	Antiviral treatment	Infection length	Fatal?
P1	Choi, *et al.*	*New Engl. J. Med.* (2020)	45	M	Antiphospholipid antibody syndrome	Rituximab, eculizumab, cyclophosphamide, corticosteroids	Remdesivir, casirivimab and imdevimab	154	Yes
P2	Baang, *et al.*	*J. Infect. Dis.* (2021)	60	M	Refractory mantle cell lymphoma	CD20 bispecific Ab, B-cell directed Ab, cyclophosphamide, doxorubicin, prednisone	Remdesivir	156	Yes
P4	Avanzato, *et al.*	*Cell* (2020)	71	F	CLL, acquired hypogammaglobulinemia	-	Convalescent plasma	156	No
P3	Molina, *et al.*	Research Square (2020)	35	M	-	-	Convalescent plasma	181	No
P5	Kemp, *et al.*	*Nature* (2020)	-	-	Marginal B cell lymphoma, hypogammaglobulinemia	B-cell depleting therapy	Remdesivir, convalescent plasma	102	Yes
P6	Mulder, *et al.*	*Clin. Infect. Dis.* (2020)	89	F	Waldenström macroglobulinemia	B-cell depleting therapy	-	59	Yes

CLL: chronic lymphocytic leukemia, RDV: remdesivir, CP: convalescent plasma, Ab: antibody

Figure 1

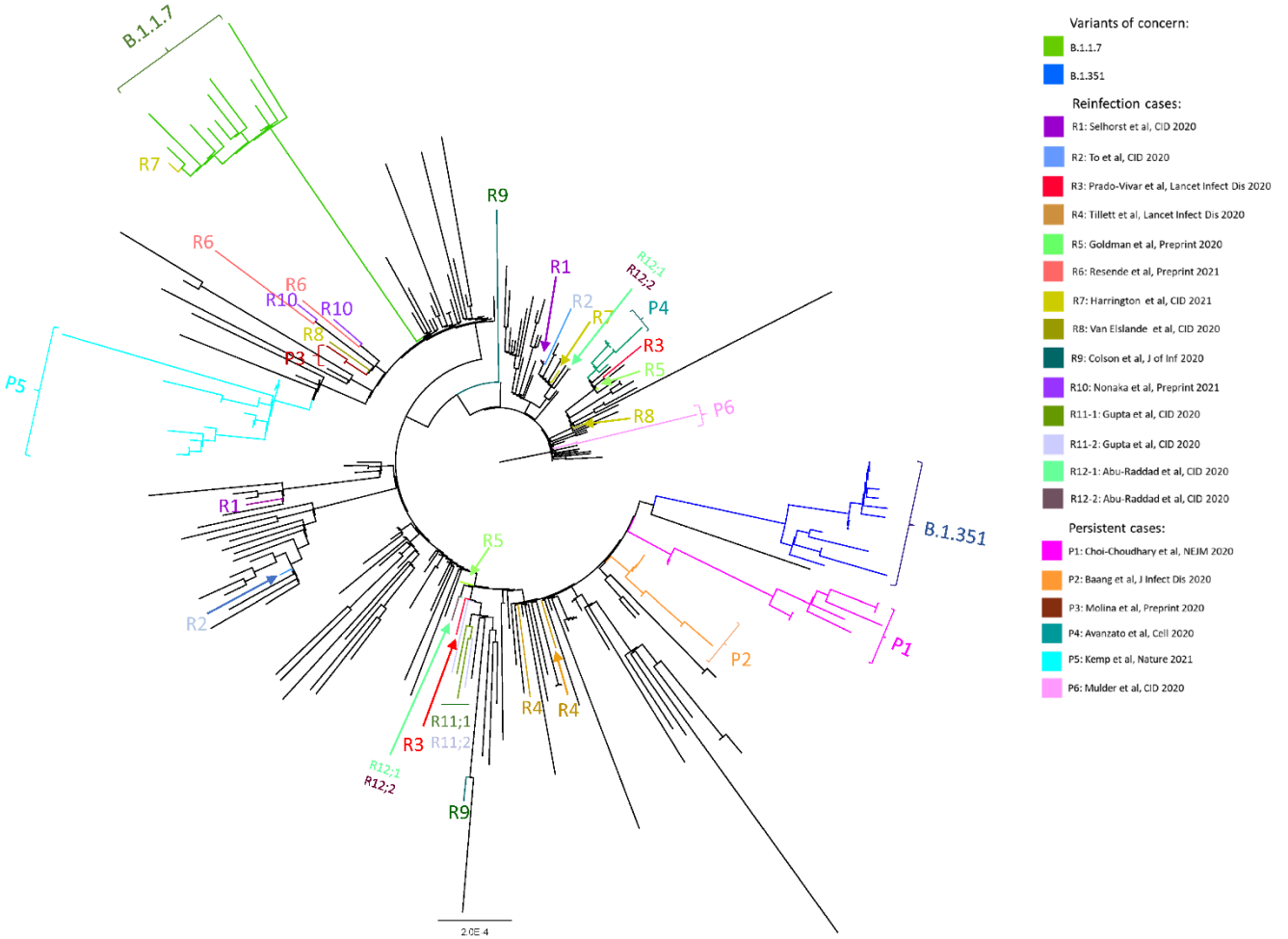

Figure 2

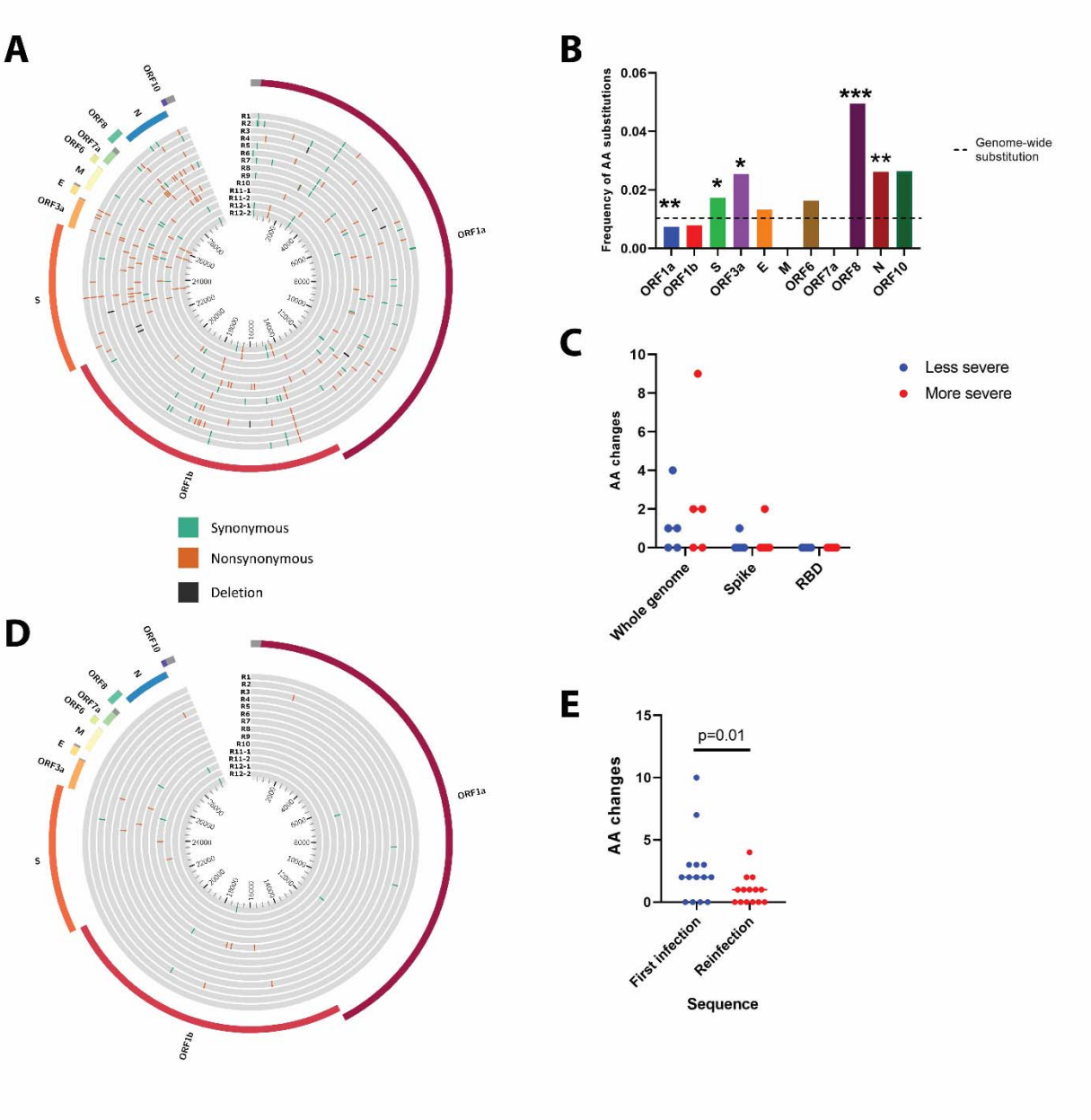

Figure 3

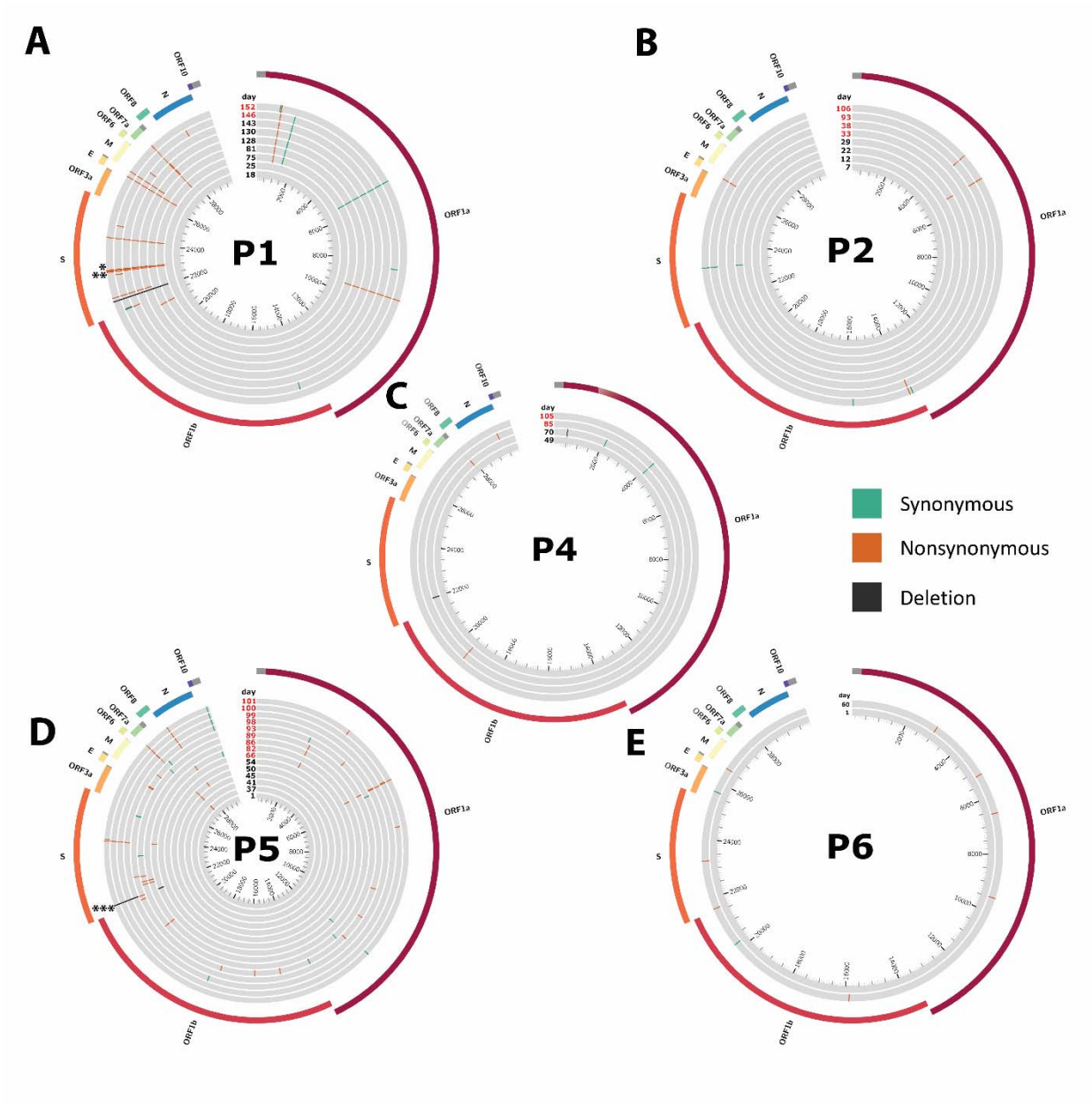

Figure 4