

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

2
3 Manish C. Choudhary^{1*}, Charles R. Crain^{1,2*}, Xueting Qiu³, William Hanage³, Jonathan Z. Li,¹

4
5 ¹Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

6 ²Northeastern University, Boston, MA, USA

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

8
9 *Authors contributed equally

10
11 **Corresponding Author:**

12 Jonathan Li, MD

13 Brigham and Women's Hospital

14 Harvard Medical School

15 jlj@bwh.harvard.edu

16
17
18 **Key words:** SARS-CoV-2, persistent COVID-19, reinfection, sequence analysis,
19 immunosuppression

20
21 **Funding:** This study was funded in part by the NIH grant 106701.

22
23 **Disclosures:**

24 Dr. Li has consulted for Abbvie.

26 **ABSTRACT**

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

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

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

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

48

49 **BACKGROUND**

50 After resolution of coronavirus disease 2019 (COVID-19) following SARS-CoV-2 infection,
51 antibodies against SARS-CoV-2 persist in the majority of patients for 6 months or more [1].
52 Despite this, there have now been a number of reports of COVID-19 reinfection that span a
53 broad range of age groups, time frame between infections and disease severity [2-8]. There
54 remains a great deal of uncertainty over the viral characteristics of reinfection cases, including
55 the degree of sequence heterogeneity and the location of new mutations between the initial and
56 reinfected variants. In addition, the diagnosis of COVID-19 reinfection has been complicated by
57 the increasing reports of persistent COVID-19 infection, especially in immunosuppressed
58 individuals. Like reinfection cases, persistent COVID-19 can also span the range of disease
59 severity, from asymptomatic to severe disease, and recurrent symptoms can last for months [9-
60 12]. Differentiating between persistent and reinfection can be challenging, and little is known
61 about differences in the location of SARS-CoV-2 mutations in these scenarios. We performed
62 an analysis of SARS-CoV-2 sequences from published cases of COVID-19 reinfection and
63 persistence, characterizing the hallmarks of reinfected sequences and the rate of viral evolution
64 in persistent infection.

65

66

67 **METHODS**

68 **Data search and selection criteria**

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

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

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

91 **Sequencing dataset compilation and phylogenetic tree construction**

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

100

101 **Mutation analysis**

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

114 For persistent infections, sequence changes were assessed at two time intervals:
115 sequences obtained before or after convalescent plasma or monoclonal antibody treatment.
116 Sequences sampled before convalescent plasma or antibody treatment were compared to the
117 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).

154 **RESULTS**

155 **Sequence analysis of reinfection cases**

156 A total of fourteen cases from twelve reports were included in this analysis (Table 1) [2-5, 7, 8,
157 17-22]. A broad range of age groups were represented and 79% were under the age of 65
158 years. Most (71%) of the cases had no reported comorbidities and none of the patients were
159 immunocompromised. The interval between diagnosis of the first infection and the second
160 infection ranged from 46 days to 250 days with a median of 110 days. Four patients had more
161 severe illness during the second infection, while five had less severe symptoms on reinfection,
162 including two who were asymptomatic on reinfection. Two cases were asymptomatic in both
163 infections, one case reported the same severity for both infections and no information on
164 infection severity was available for two cases (Table 1). Four cases reported reinfection with a
165 virus from the same clade.

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

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

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

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

221

222

223 **DISCUSSION**

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

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

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

254 Two factors generally differentiated between reinfection and persistent infection
255 scenarios: first, reinfections have so far been largely described in immunocompetent individuals
256 while the majority of persistent COVID cases have been in immunosuppressed patients.
257 Secondly, phylogenetic analysis can generally differentiate between reinfection and persistent
258 infection, especially in cases where persistent infection allowed the longitudinal collection of >2
259 sequences. However, given the slow rate of SARS-CoV-2 evolution and limited viral diversity
260 [34], it can be challenging to differentiate between reinfection and persistent infection, especially
261 in situations with limited sampling and/or duration between samples. Overall, our results
262 demonstrate the need to further explore factors that increase the risk of breakthrough
263 reinfections and persistent COVID-19. This line of investigation will have important implications
264 for preventing the rise of novel variants and the durability of current available vaccines.

265

266

267 **Acknowledgements**

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

269 **REFERENCES**

- 270 1. Dan JM, Mateus J, Kato Y, et al. Immunological memory to SARS-CoV-2 assessed for up to 8
271 months after infection. *Science* **2021**.
- 272 2. Van Elslande J, Vermeersch P, Vandervoort K, et al. Symptomatic SARS-CoV-2 reinfection by a
273 phylogenetically distinct strain. *Clin Infect Dis* **2020**.
- 274 3. To KK, Hung IF, Ip JD, et al. COVID-19 re-infection by a phylogenetically distinct SARS-
275 coronavirus-2 strain confirmed by whole genome sequencing. *Clin Infect Dis* **2020**.
- 276 4. Tillett RL, Sevinsky JR, Hartley PD, et al. Genomic evidence for reinfection with SARS-CoV-2: a
277 case study. *Lancet Infect Dis* **2021**; 21(1): 52-8.
- 278 5. Selhorst P, Van Ierssel S, Michiels J, et al. Symptomatic SARS-CoV-2 reinfection of a health care
279 worker in a Belgian nosocomial outbreak despite primary neutralizing antibody response. *Clin*
280 *Infect Dis* **2020**.
- 281 6. Mulder M, van der Vegt D, Oude Munnink BB, et al. Reinfection of SARS-CoV-2 in an
282 immunocompromised patient: a case report. *Clin Infect Dis* **2020**.
- 283 7. Harrington D, Kele B, Pereira S, et al. Confirmed Reinfection with SARS-CoV-2 Variant VOC-
284 202012/01. *Clin Infect Dis* **2021**.
- 285 8. Goldman JD, Wang K, Roltgen K, et al. Reinfection with SARS-CoV-2 and Failure of Humoral
286 Immunity: a case report. *medRxiv* **2020**.
- 287 9. Choi B, Choudhary MC, Regan J, et al. Persistence and Evolution of SARS-CoV-2 in an
288 Immunocompromised Host. *N Engl J Med* **2020**; 383(23): 2291-3.
- 289 10. Avanzato VA, Matson MJ, Seifert SN, et al. Case Study: Prolonged Infectious SARS-CoV-2
290 Shedding from an Asymptomatic Immunocompromised Individual with Cancer. *Cell* **2020**;
291 183(7): 1901-12 e9.
- 292 11. Baang JH, Smith C, Mirabelli C, et al. Prolonged Severe Acute Respiratory Syndrome Coronavirus
293 2 Replication in an Immunocompromised Patient. *J Infect Dis* **2021**; 223(1): 23-7.
- 294 12. Kemp SA, Collier DA, Datir R, et al. Neutralising antibodies in Spike mediated SARS-CoV-2
295 adaptation. *medRxiv* **2020**.
- 296 13. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool
297 for maximum likelihood analysis. *Nucleic Acids Res* **2016**; 44(W1): W232-5.
- 298 14. Krzywinski MI, Schein JE, Birol I, et al. Circos: An information aesthetic for comparative
299 genomics. *Genome Research* **2009**.
- 300 15. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of
301 heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* **2016**; 2(1): vew007.
- 302 16. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the
303 BEAST 1.7. *Mol Biol Evol* **2012**; 29(8): 1969-73.
- 304 17. Prado-Vivar B, Becerra-Wong M, Guadalupe JJ, et al. A case of SARS-CoV-2 reinfection in
305 Ecuador. *Lancet Infect Dis* **2020**.
- 306 18. Resende PC, de Vasconcelos RHT, Arantes I, et al. Spike E484K mutation in the first SARS-CoV-2
307 reinfection case confirmed in Brazil. *Virologica* **2021**.
- 308 19. Colson P, Finaud M, Levy N, Lagier JC, Raoult D. Evidence of SARS-CoV-2 re-infection with a
309 different genotype. *J Infect* **2020**.
- 310 20. Nonaka CKV, Franco MM, Graf T, et al. Genomic Evidence of a Sars-Cov-2 Reinfection Case With
311 E484K Spike Mutation in Brazil. *Preprintsorg* **2021**.
- 312 21. Gupta V, Bhojar RC, Jain A, et al. Asymptomatic reinfection in two healthcare workers from
313 India with genetically distinct SARS-CoV-2. *Clin Infect Dis* **2020**.
- 314 22. Abu-Raddad LJ, Chemaitelly H, Malek JA, et al. Assessment of the risk of SARS-CoV-2 reinfection
315 in an intense re-exposure setting. *Clin Infect Dis* **2020**.

- 316 23. Kemp SA, Collier DA, Datir RP, et al. SARS-CoV-2 evolution during treatment of chronic infection.
317 Nature **2021**.
- 318 24. Molina LP, Chow SK, Nickel A, Love JE. Prolonged Detection of Severe Acute Respiratory
319 Syndrome Coronavirus 2 (SARS-CoV-2) RNA in an Obstetric Patient With Antibody
320 Seroconversion. *Obstet Gynecol* **2020**.
- 321 25. Wibmer CK, Ayres F, Hermanus T, et al. SARS-CoV-2 501Y.V2 escapes neutralization by South
322 African COVID-19 donor plasma. *bioRxiv* **2021**.
- 323 26. Weisblum Y, Schmidt F, Zhang F, et al. Escape from neutralizing antibodies by SARS-CoV-2 spike
324 protein variants. *Elife* **2020**; 9.
- 325 27. Starr TN, Greaney AJ, Addetia A, et al. Prospective mapping of viral mutations that escape
326 antibodies used to treat COVID-19. *Science* **2021**; 371(6531): 850-4.
- 327 28. Camprubi-Ferrer D, Gaya A, Marcos MA, et al. Persistent replication of SARS-CoV-2 in a severely
328 immunocompromised treated with several courses of remdesivir. *Int J Infect Dis* **2020**.
- 329 29. Hensley MK, Bain WG, Jacobs J, et al. Intractable COVID-19 and Prolonged SARS-CoV-2
330 Replication in a CAR-T-cell Therapy Recipient: A Case Study. *Clin Infect Dis* **2021**.
- 331 30. Martinot M, Jary A, Fafi-Kremer S, et al. Remdesivir failure with SARS-CoV-2 RNA-dependent
332 RNA-polymerase mutation in a B-cell immunodeficient patient with protracted Covid-19. *Clin*
333 *Infect Dis* **2020**.
- 334 31. Helleberg M, Niemann CU, Moestrup KS, et al. Persistent COVID-19 in an Immunocompromised
335 Patient Temporarily Responsive to Two Courses of Remdesivir Therapy. *J Infect Dis* **2020**; 222(7):
336 1103-7.
- 337 32. Sepulcri C, Dentone C, Mikulska M, et al. The longest persistence of viable SARS-CoV-2 with
338 recurrence of viremia and relapsing symptomatic COVID-19 in an immunocompromised patient
339 – a case study. *medRxiv* **2021**: 2021.01.23.21249554.
- 340 33. McMahan K, Yu J, Mercado NB, et al. Correlates of protection against SARS-CoV-2 in rhesus
341 macaques. *Nature* **2020**.
- 342 34. Rausch JW, Capoferri AA, Katusiime MG, Patro SC, Kearney MF. Low genetic diversity may be an
343 Achilles heel of SARS-CoV-2. *Proc Natl Acad Sci U S A* **2020**; 117(40): 24614-6.

344

345

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 $\Delta 69/\Delta 70$ 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, <i>et al.</i>	<i>Clin. Infect. Dis.</i> (2020)	39	F	None	185	Less	19A	20A
R2	To, <i>et al.</i>	<i>Clin. Infect. Dis.</i> (2020)	33	M	None	144	Less	19A	20E
R3	Prado-Vivar, <i>et al.</i>	<i>Lancet Infect. Dis.</i> (2021)	46	M	None	63	More	20A	19B
R4	Tillett, <i>et al.</i>	<i>Lancet Infect. Dis.</i> (2020)	25	M	None	48	More	20C	20C
R5	Goldman, <i>et al.</i>	medRxiv	60-69	N/A	Emphysema, hypertension	139	Less	19B	20A
R6	Resende, <i>et al.</i>	Virological	37	F	None	116	Similar	20B	20B
R7	Harrington, <i>et al.</i>	<i>Clin. Infect. Dis.</i> (2021)	78	M	Diabetic nephropathy with hemodialysis, COPD, sleep apnea, ischemic heart disease	250	More	19A	B.1.1.7
R8	Van Elslande, <i>et al.</i>	<i>Clin. Infect. Dis.</i> (2020)	51	F	Asthma	93	Less	20B	19B
R9	Colson, <i>et al.</i>	<i>J. Infect.</i> (2020)	70	M	None	105	Less	20A	20A.EU2
R10	Nonaka, <i>et al.</i>	Preprints	45	F	None	147	More	20B	20B
R11-1	Gupta, <i>et al.</i>	<i>Clin. Infect. Dis.</i> (2020)	25	M	None	108	Both asymptomatic	19A	20A
R11-2			28	F	None	111	Both asymptomatic	20A	20A
R12-1	Abu-Raddad, <i>et al.</i>	<i>Clin. Infect. Dis.</i> (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	Immunosuppressants	Antiviral treatment	Infection length	Fatal?
P1	Choi, <i>et al.</i>	<i>New Engl. J. Med.</i> (2020)	45	M	Antiphospholipid antibody syndrome	Rituximab, eculizumab, cyclophosphamide, corticosteroids	Remdesivir, casirivimab and imdevimab	154	Yes
P2	Baang, <i>et al.</i>	<i>J. Infect. Dis.</i> (2021)	60	M	Refractory mantle cell lymphoma	CD20 bispecific Ab, B-cell directed Ab, cyclophosphamide, doxorubicin, prednisone	Remdesivir	156	Yes
P4	Avanzato, <i>et al.</i>	<i>Cell</i> (2020)	71	F	CLL, acquired hypogammaglobulinemia	-	Convalescent plasma	156	No
P3	Molina, <i>et al.</i>	Research Square (2020)	35	M	-	-	Convalescent plasma	181	No
P5	Kemp, <i>et al.</i>	<i>Nature</i> (2020)	-	-	Marginal B cell lymphoma, hypogammaglobulinemia	B-cell depleting therapy	Remdesivir, convalescent plasma	102	Yes
P6	Mulder, <i>et al.</i>	<i>Clin. Infect. Dis.</i> (2020)	89	F	Waldenström macroglobulinemia	B-cell depleting therapy	-	59	Yes

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

Figure 1

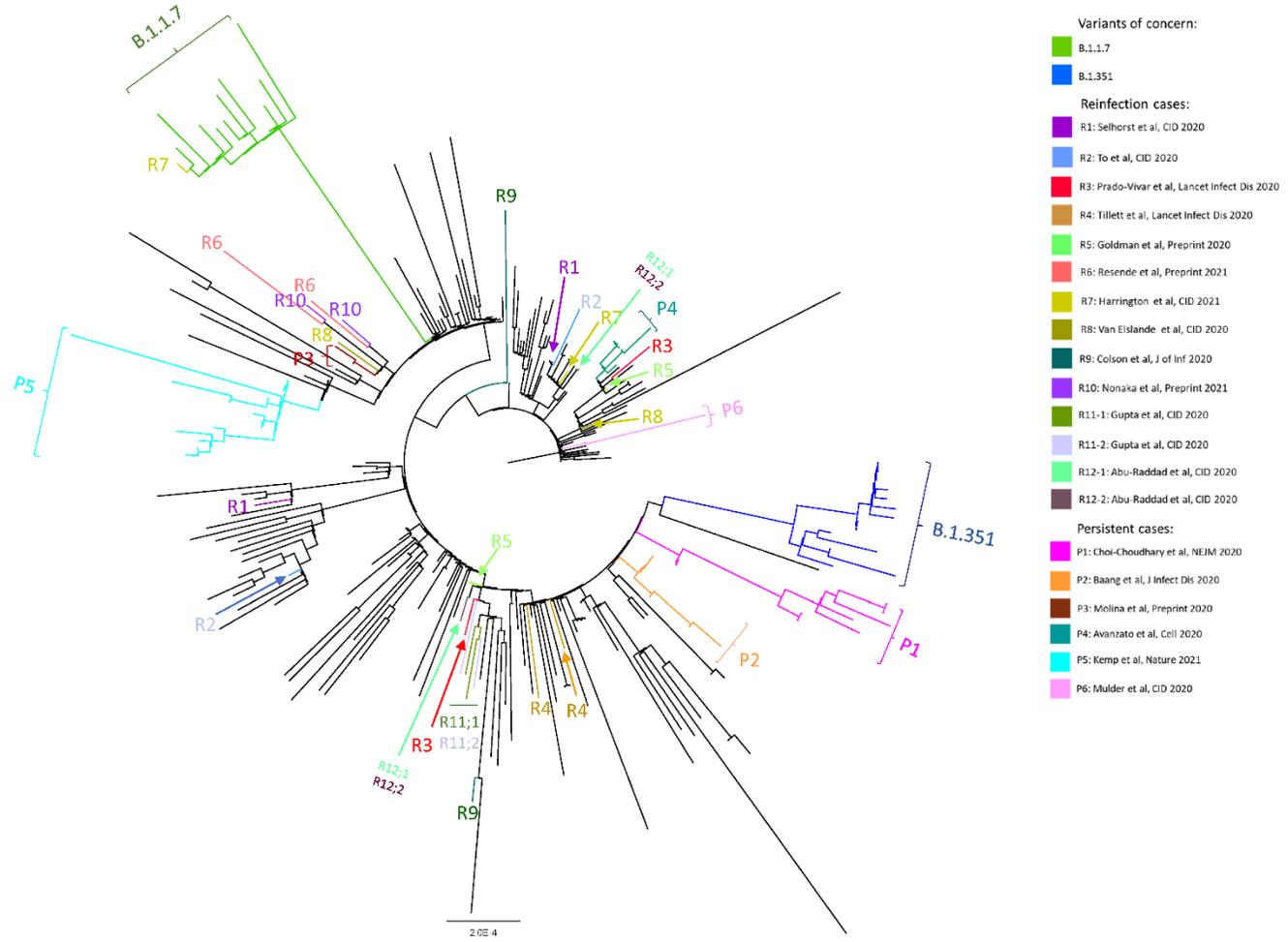


Figure 2

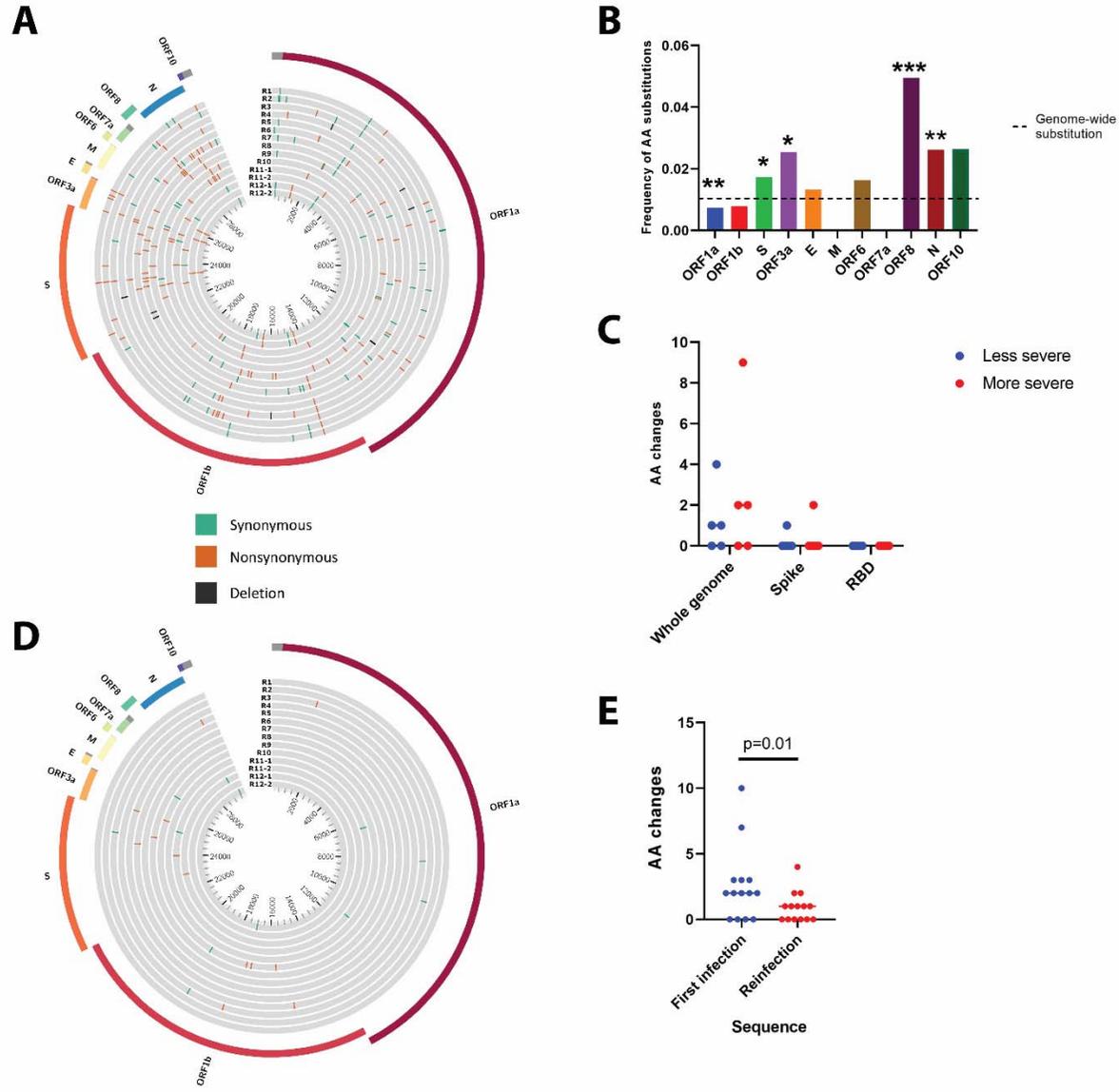


Figure 3

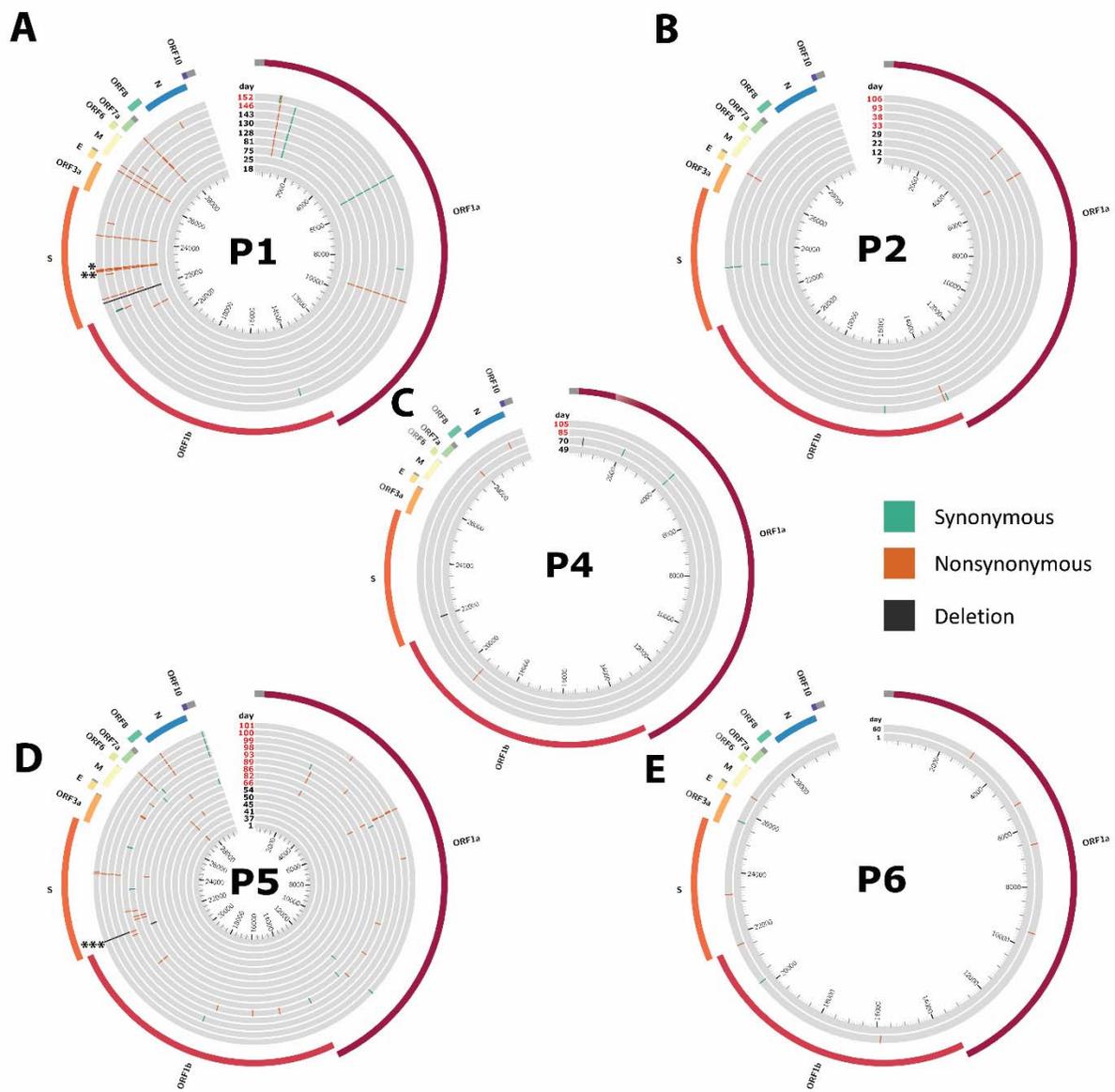


Figure 4

