

1 High number of SARS-CoV-2 persistent infections 2 uncovered through genetic analysis of samples from a 3 large community-based surveillance study

4
5 Mahan Ghafari^{*1,2}, Matthew Hall¹, Tanya Golubchik^{1,3}, Daniel Ayoubkhani^{4,5}, Thomas
6 House⁶, George MacIntyre-Cockett^{1,7}, Helen Fryer¹, Laura Thomson¹, Anel Nurtay¹, David
7 Buck⁷, Angie Green⁷, Amy Trebes⁷, Paolo Piazza⁷, Lorne J Lonie⁷, Ruth Studley⁴, Emma
8 Rourke⁴, Darren Smith^{8,9}, Matthew Bashton^{8,9}, Andrew Nelson⁹, Matthew Crown^{8,9}, Clare
9 McCann⁹, Gregory R Young^{8,9}, Rui Andre Nunes dos Santos⁹, Zack Richards⁹, Adnan
10 Tariq⁹, Roberto Cahuantzi⁴, Wellcome Sanger Institute COVID-19 Surveillance Team^{**},
11 COVID-19 Infection Survey Group, The COVID-19 Genomics UK (COG-UK) Consortium⁺⁺¹⁰,
12 Jeff Barrett¹¹, Christophe Fraser^{1,7,11}, David Bonsall^{1,7,16}, Ann Sarah Walker^{12,13,14,15}, Katrina
13 Lythgoe^{1,2}

14

15 *Corresponding author: mahan.ghafari@ndm.ox.ac.uk

16

17 ¹Big Data Institute, Nuffield Department of Medicine, University of Oxford, Old Road Campus, Oxford
18 OX3 7LF, UK.

19 ²Department of Biology, University of Oxford, Oxford OX1 3SZ, UK

20 ³Sydney Infectious Diseases Institute (Sydney ID), School of Medical Sciences, Faculty of

21 ⁴Office for National Statistics, Newport, UK

22 ⁵Medicine and Health, University of Sydney

23 ⁶Leicester Real World Evidence Unit, Diabetes Research Centre, University of Leicester, Leicester,
24 UK

25 ⁷Department of Mathematics, University of Manchester, Manchester, M13 9PL, UK

26 ⁸Wellcome Centre for Human Genetics, Nuffield Department of Medicine, NIHR Biomedical Research
27 Centre, University of Oxford, Old Road Campus, Oxford OX3 7BN, UK.

28 ⁹The Hub for Biotechnology in the Built Environment, Department of Applied Sciences, Faculty of
29 Health and Life Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK

30 ¹⁰Department of Applied Sciences, Faculty of Health and Life Sciences, Northumbria University,
31 Newcastle upon Tyne, NE1 8ST, UK

32 ¹¹<https://www.cogconsortium.uk>

33 ¹²Wellcome Sanger Institute, Cambridge CB10 1SA, UK

34 ¹³Nuffield Department of Medicine, University of Oxford, Oxford, UK

35 ¹⁴The National Institute for Health Research Health Protection Research Unit in Healthcare
36 Associated Infections and Antimicrobial Resistance at the University of Oxford, Oxford, UK.

37 ¹⁵The National Institute for Health Research Oxford Biomedical Research Centre, University of
38 Oxford, Oxford, UK

39 ¹⁶MRC Clinical Trials Unit at UCL, UCL, London, UK

40 ¹⁷Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Headington, Oxford
41 OX3 9DU, UK

42

43 ^{**}The full list of names is available at <https://www.sanger.ac.uk/project/wellcome-sanger-institute-covid-19-surveillance-team/>

44 ⁺⁺The full list of names and affiliations of COG-UK members is provided in the appendix

45

46 **Abstract**

47

48 Persistent severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
49 infections may act as viral reservoirs that could seed future outbreaks ¹⁻⁵, give rise to
50 highly divergent lineages ⁶⁻⁸, and contribute to cases with post-acute Coronavirus
51 disease 2019 (COVID-19) sequelae (Long Covid) ^{9,10}. However, the population
52 prevalence of persistent infections, their viral load kinetics, and evolutionary
53 dynamics over the course of infections remain largely unknown. We identified 381
54 infections lasting at least 30 days, of which 54 lasted at least 60 days. These
55 persistently infected individuals had more than 50% higher odds of self-reporting
56 Long Covid compared to the infected controls, and we estimate that 0.09-0.5% of
57 SARS-CoV-2 infections can become persistent and last for at least 60 days. In
58 nearly 70% of the persistent infections we identified, there were long periods during
59 which there were no consensus changes in virus sequences, consistent with
60 prolonged presence of non-replicating virus. Our findings also suggest reinfections
61 with the same major lineage are rare and that many persistent infections are
62 characterised by relapsing viral load dynamics. Furthermore, we found a strong
63 signal for positive selection during persistent infections, with multiple amino acid
64 substitutions in the Spike and ORF1ab genes emerging independently in different
65 individuals, including mutations that are lineage-defining for SARS-CoV-2 variants, at
66 target sites for several monoclonal antibodies, and commonly found in
67 immunocompromised patients ¹¹⁻¹⁴. This work has significant implications for
68 understanding and characterising SARS-CoV-2 infection, epidemiology, and
69 evolution.

70

71 **Main**

72

73 The emergence of highly divergent variants of SARS-CoV-2 has been a defining
74 feature of the COVID-19 pandemic. While the evolutionary origins of these variants
75 are still a matter of speculation, multiple pieces of evidence point to chronic
76 persistent infections as their most likely source ^{5,7,15}. In particular, infections in
77 immunocompromised patients who cannot clear the virus may lead to persistence for
78 months ^{6,7,16,17} or even years ^{8,18} before potentially seeding new outbreaks in the
79 community ³. Persistence of SARS-CoV-2 during chronic infections exposes the viral
80 population to host-immune responses and other selective pressures as a result of
81 treatments over prolonged periods of time. They also release the virus from
82 undergoing the tight population bottlenecks that are characteristic of SARS-CoV-2
83 transmission ^{19,20}, making the viral population less vulnerable to stochastic genetic
84 drift. These adaptive intrahost changes can lead to elevated evolutionary rates,
85 particularly in key regions of the Spike protein that are often associated with immune
86 escape and elevated rates of transmission ^{13,14}.

87

88 Despite the significant public health implications of persistent infections, uncertainty
89 still surrounds how common these infections are among the general population, how

90 long they last, their potential for adaptive evolution, and their contribution to Long
91 Covid.

92

93 In this work, we leveraged genetic, symptom, and epidemiological data from the
94 Office for National Statistics Covid Infection Survey (ONS-CIS) ²¹, an ongoing large
95 scale community-based surveillance study carried out in the UK. We identified
96 individuals with persistent SARS-CoV-2 infection and characterised various aspects
97 of their infection including evolutionary changes in the virus, viral load kinetics,
98 number of reported symptoms, and prevalence of Long Covid.

99

100 Identifying persistent infections

101

102 We considered more than 100,000 high-quality sequenced samples from the ONS-
103 CIS collected between 2nd November 2020 to 15th August 2022, and representing
104 ~95,000 people living in ~75,000 households across the UK (see **Methods**).

105 Individuals in the survey were typically sampled once a week for the first four weeks
106 of their enrolment, and then monthly thereafter. To identify persistent infections we
107 first limited the dataset to individuals with two or more RT-PCR positive samples with
108 cycle threshold (Ct) values ≤ 30 (in which sequencing was attempted; a proxy for viral
109 load), taken at least 26 days apart, and where the consensus sequences were of the
110 same major lineages of Alpha, Delta, BA.1 or BA.2 (BA.4 and BA.5 not considered).
111 If those sequences also shared the same rare single nucleotide polymorphisms
112 (SNPs) at one or more sites relative to the major-lineage population-level consensus,
113 we classified them as having a persistent infection.

114

115 We defined a rare mutation as one observed in 400 or fewer samples within the
116 entire ONS-CIS dataset, giving a false positive rate of identifying persistent infections
117 of 0-3% depending on the major lineage (see **Methods** and **Supplementary Figure**
118 **1**). We note that the rare SNP method provides a conservative estimate for the true
119 number of persistent infections since some persistent infections may not have rare
120 mutations. To further evaluate the robustness of our method in identifying persistent
121 infections, we considered the phylogenetic relationship between the sequences from
122 persistent infections relative to other sequences of the same major lineage that
123 belonged to individuals with only a single sequence within the ONS-CIS dataset. The
124 great majority of sets of sequences identified as belonging to the same persistent
125 infection formed monophyletic groups with strong bootstrap support (**Figure 1a**; see
126 also **Supplementary Figure 2**). However, seven sequences did not group with the
127 other sequence(s) from the same persistent infection. All of these had high Ct values
128 (Ct~30) and poor genome coverage which may explain their lack of clustering on the
129 phylogeny (**Supplementary Figure 2**).

130

131 We found 381 persistent infections with sequences spanning at least 26 days, of
132 which 54 spanned at least 56 days, representing nearly 0.07% (54/77,561) of all
133 individuals with one or more sequences (with $Ct \leq 30$) of the four major lineages we

134 investigated in this study (**Figure 1b**; see also **Table 1**). Notably, 9% (2/23), 9%
135 (19/219), and 8% (8/208) of sequences from persistent infections with Alpha, Delta,
136 and BA.1, respectively, were sampled weeks after when the corresponding major
137 lineage has dropped to $\leq 1\%$ frequency of all the ONS-CIS sequences (**Figure 1c**);
138 the longest infection was with BA.1 and lasted for at least 193 days (see **Figure 1b**).

139
140 The actual duration of persistent infections are likely to be at least 3-4 days longer
141 than the time between when the first and last sequenced samples were collected,
142 since it typically takes 3-4 days since the start of infection for viral loads to be
143 sufficiently high to be sequenced ($Ct \leq 30$)^{22,23}, and similarly viral loads will be too low
144 (Ct values too high) to sequence at the tail end of infection. Since individuals were
145 typically sampled weekly during the first four weeks of enrollment, followed by
146 monthly sampling thereafter, it is unsurprising that most persistent infections had
147 observable durations clustering around 30 or 60 days (see **Supplementary Figure**
148 **3**).

149
150 We found evidence suggestive of transmission from 11 persistently infected
151 individuals: we identified two households each with two members having concurrent
152 persistent infections, and in nine other households, one member tested positive for
153 SARS-CoV-2 by RT-PCR within 10 days of a positive sequenced sample being
154 collected from a persistently infected individual from the same household. As well as
155 being clustered in households, the sequences from the suspected persistently
156 infected source and the recipient had either no consensus nucleotide differences
157 (nine cases) or one consensus nucleotide difference (two cases), consistent with
158 transmission. If these do represent transmission, it is worth noting that none of these
159 persistent infections involved highly divergent sequences relative to their first
160 sequence, and hence these do not provide an example of the emergence and
161 spread of highly divergent lineages.

162 163 Identifying reinfections with the same major lineage

164
165 We considered a pair of sequences from the same individual to indicate a reinfection
166 if they were sampled at least 26 days apart, had at least one consensus nucleotide
167 difference between the sequenced sampling timepoints, and shared no rare SNPs
168 (see **Methods**). This criterion may overestimate the true number of reinfections as
169 some persistent infections may not have a rare SNP, and within-host evolution can
170 lead to the loss of a rare SNP and/or the gain of other mutations leading to
171 differences in the consensus sequence between the samples. We identified three
172 individuals for which pairs of sequences from different sampling timepoints had no
173 identical rare SNPs and at least one consensus difference, but whose viral load
174 trajectories were consistent with a persistent chronic infection. We therefore
175 excluded these individuals from the reinfection group (see **Supplementary Figure**
176 **4**).

177

178 Overall, we identified 60 reinfections with the same major lineage (**Table 1**). Of all
179 the cases classed as either persistent infections or reinfections with the same major
180 lineage, 10-15% were classed as reinfections (see **Table 1**), rising to 20-40% if only
181 samples collected at least 56 days apart were included (see **Figure 1b**). This
182 suggests the number of individuals reinfected with the same major lineage is low
183 compared to the number of individuals with persistent infection. Sequences from
184 individuals identified as reinfected, collected at the point of primary infection and
185 reinfection, did not form monophyletic groups and mostly belonged to distantly
186 related subclades, and hence supports our method for identifying reinfections
187 (**Figure 1a; Supplementary Figure 2**).

188

189 Evidence of non-replicating virus during infection

190

191 Of the 381 persistently infected individuals that we identified, nearly 70% (267/381)
192 had a pair of sequenced samples taken at least 26 days apart with no nucleotide
193 differences at the consensus level. This is striking given the between-host within-
194 lineage evolutionary rate of SARS-CoV-2 of ~2 single nucleotide variants (SNVs) per
195 month²⁴⁻²⁶. We contrasted the number of consensus nucleotide differences between
196 pairs of samples from persistent infections with 16,000 random pairs of sequences
197 sampled from the complete set of sequenced samples from the ONS-CIS, and with
198 each pair from the same major lineage (**Supplementary Figure 5**). Of these, only 6
199 pairs had no SNVs (i.e., less than 0.04% of pairs).

200

201 Emergence of notable mutations

202

203 For all pairs of sequences from each of the persistent infections, we identified
204 mutations for which there was a change in consensus between the two sampling
205 time points. Among the 381 persistently infected individuals, we observed 317
206 changes in the consensus nucleotide representing 277 unique mutations, and 31
207 deletions representing 18 unique deletions. Many of these mutations have previously
208 been identified as either lineage-defining mutations for variants of concern or
209 variants of interest²⁷ (eight mutations and two deletions), recurrent mutations in
210 immunocompromised individuals¹²⁻¹⁴ (15 mutations and four deletions), or key
211 mutations with antibody escape properties and target sites for various different
212 monoclonal antibodies^{11,28} (seven mutations) (**Supplementary Table 1**).

213

214 We observed several mutations at the same genomic positions in multiple individuals
215 over the course of their persistent infections. For example, three BA.2 infected
216 individuals from different households acquired a mutation at codon position 547 in
217 Spike (**Figure 2**), two of which were the T547K mutation which is a lineage-defining
218 mutation for BA.1, and one the K547T mutation (**Figure 2c**; also see
219 **Supplementary Table 1**). Strikingly, twelve individuals acquired a deletion
220 (ORF1ab: Δ 81-87) in the NSP1 coding region. A similar deletion has previously been
221 observed during the chronic infection of an immunocompromised individual with

222 cancer¹⁶, and was associated with lower type I interferon response in infected cells
223²⁹. We also identified a persistent infection with BA.1 lasting for at least 133 days
224 during which 33 unique mutations (23 mutations in ORF1ab, 6 in Spike, 1 in ORF3a,
225 1 in M, and 2 in ORF7) were observed (see **Supplementary Figure 2**); eleven of the
226 ORF1ab mutations and all of the mutations in Spike, ORF3a, and ORF7 were
227 nonsynonymous.

228
229 Overall, we observed a strong signal for positive selection in Spike, with nearly nine-
230 fold more nonsynonymous compared to synonymous mutations (**Figure 2b**). With a
231 total of seven nonsynonymous mutations, ORF8 had the highest per base number of
232 nonsynonymous mutations followed by Spike with 61 nonsynonymous mutations.

233 234 Persistence with relapsing viral load

235
236 Of the 381 persistent infections, 65 had three or more RT-PCR tests taken over the
237 course of their infection. We classified these infections as persistent-relapsing if they
238 had a negative RT-PCR test during the infection (n=20), and the rest as persistent-
239 chronic (n=45) (**Figure 3a,b**). Given the weekly or monthly sampling of individuals
240 enrolled in the ONS-CIS, infections classed as persistent-chronic may have
241 unsampled periods of very low viral burden, meaning the persistent-relapsing
242 category is likely to be an underestimate. Nonetheless, the observation of relapsing
243 viral load dynamics in over 30% of cases is striking given that, in the absence of
244 genetic information, they could have been misidentified as reinfections, depending
245 on the definition used. Of the 27 cases identified as reinfections with three or more
246 RT-PCR tests, all showed relapsing viral load dynamics (**Figure 3c**).

247
248 As the sampling strategy of ONS-CIS is based on testing representative individuals
249 across the UK regardless of symptoms, we can estimate the percentage of SARS-
250 CoV-2 infections that are persistent and last for longer than 60 days in the general
251 population. This requires making assumptions about how many persistent infections
252 are missed among ONS-CIS participants due to the monthly (and weekly) sampling.
253 More precisely, estimating the proportion of infections that are persistent depends on
254 the proportion of days the infection has sequenceable virus during the infection
255 (would have Ct \leq 30 if tested); the fewer the number of days the infection has
256 sequenceable virus, the more likely it is that a persistent infection is missed. By
257 taking two extreme scenarios for the proportion of days that the virus is
258 sequenceable during persistent infection (0.7 and 0.14; see **Methods**), we estimate
259 that approximately 0.1-0.5% of infections become persistent for 60 days or more.

260 261 Difference in viral load and symptoms

262
263 For the majority of persistent infections, Ct values (inversely proportional to viral titre
264³⁰) were higher at the last sequenced time point compared to the first sequenced
265 time point (**Figure 3d**). For reinfections with the same major lineage, the last

266 sequenced sample also had higher Ct values than the first, but the magnitude of the
267 difference was smaller compared to persistent infections (**Figure 2d**). In both cases,
268 the rise in Ct value (decrease in viral titre) during infections or between reinfections
269 could be a consequence of host immunity or within-host compartmentalisation.
270 Additionally, the rise in Ct for reinfections could be due to the disproportionate
271 sampling of individuals with older infections, which tend to have lower viral loads,
272 towards the end of an epidemic wave ^{31,32}.

273
274 Individuals with persistent infections remained largely asymptomatic during the later
275 stages of infection, with, on average, reporting two fewer symptoms in the last 7
276 days at the last time of sampling (at which a sequence was obtained) when
277 compared to the first. They also consistently reported very few or no symptoms after
278 the first positive sample (**Figure 3e**). In comparison, individuals reinfected with the
279 same major lineage reported on average only one fewer symptom at the reinfection
280 sampling time point compared to the primary sampling time point (**Figure 3e**).

281

282 Prevalence of Long Covid

283

284 From February 2021, as well as reporting symptoms, participants were asked if they
285 describe themselves as having Long Covid and still experiencing symptoms more
286 than four weeks after they first had COVID-19 (see **Methods**). We estimated the
287 prevalence of self-reported Long Covid in the persistently infected individuals
288 compared with a control group, accounting for several confounding variables (see
289 **Methods**). In the persistent infection group, 9.0% of respondents (32/354) self-
290 reported Long Covid at their first visit ≥ 12 weeks since the start of infection, and
291 7.1% (19/266) reported Long Covid at ≥ 26 weeks. However, in the control group,
292 only 5.1% (4,976/97,404) reported Long Covid at their first visit ≥ 12 weeks, and 4.5%
293 (3,261/72,407) reported Long Covid at ≥ 26 weeks.

294

295 Correcting for confounders, we found strong evidence for a 55% higher odds of
296 reporting Long Covid ≥ 12 weeks post-infection among persistently infected
297 individuals compared to controls (Chi square test with Yates correction ³³ $p=0.004$ for
298 the unadjusted model, $p=0.021$ for the adjusted model), but no evidence of a
299 difference for Long Covid ≥ 26 weeks post-infection ($p=0.127$ for the unadjusted
300 model, $p=0.367$ for the adjusted model) (**Table 2**). The lower rate of reporting Long
301 Covid 26 weeks post-infection could be because the majority of the persistent
302 infections we identified lasted for less than 26 weeks, and hence persistence of an
303 infection may no longer be a contributing factor to Long Covid.

304

305 **Discussion**

306

307 We developed a robust approach for identifying persistent SARS-CoV-2 infections in
308 individuals with sequenced samples spanning a month or longer. Of the 381
309 persistent infections we identified among participants of the ONS-CIS, 54 lasted at

310 least two months, two over six months, and in some cases the infecting lineage had
311 gone extinct in the general population. In contrast, we only identified 60 reinfections
312 by the same major lineage as the primary infection, suggesting immunity to the same
313 variant remains strong after infection, at least until the lineage has gone extinct.

314
315 The large number of persistent infections we uncovered is striking, given the leading
316 hypothesis that many of the variants of concern (VOCs) emerged wholly or partially
317 during long-term chronic infections in immunocompromised individuals¹. Since the
318 ONS-CIS is a community-based surveillance study, our observations suggest the
319 pool of people in which long-term infections could occur, and hence potential
320 sources of divergent variants, may be much larger than generally thought. We
321 estimate that more than nearly one in a thousand of all infections, and potentially as
322 many as one in 200, may become persistent for at least two months. The harbouring
323 of these persistent infections in the general community may also help explain the
324 early detection of cryptic lineages circulating in wastewaters^{34,35} long before they
325 spread in the population at large.

326
327 In support of the hypothesis that VOCs may emerge during prolonged infections, a
328 number of studies have shown elevated evolutionary rates driven by selection during
329 chronic infections of immunocompromised individuals⁶⁻⁸. Among the persistently
330 infected individuals we identified, we also found strong evidence for positive
331 selection and parallel evolution, particularly in Spike and ORF1ab. In the most
332 extreme case, we observed one persistent infection with 33 substitutions over a four-
333 month period, 20 of which were nonsynonymous.

334
335 Potentially more remarkable, however, was our discovery of viral infections exhibiting
336 latent evolutionary dynamics, with no consensus level genome changes for two
337 months or longer. As far as we are aware, this is the first study to demonstrate that
338 decelerated evolutionary rates may be a common outcome of persistent infection.
339 The factors causing these low rates of evolution are unknown, but one possible
340 explanation is infection of long-lived cells^{36,37} which then seed new outbreaks within
341 the infected individual weeks or months later. Although the relapsing viral load
342 dynamics we found in many of the persistently infected individuals supports this
343 hypothesis, we did not find any consistent pattern between the amount of viral
344 divergence during infection and the lower viral load activity in these individuals.

345
346 Intriguingly, individuals with persistent infections report fewer symptoms later in a
347 persistent infection compared to at their first positive sample, or remain
348 asymptomatic throughout infection, but have more than 50% higher odds of Long
349 Covid compared to a control group. Although the link between viral persistence and
350 Long Covid may not be causal, these results suggest persistent infections could be
351 contributing to the pathophysiology of Long Covid^{10,38}, as also evidenced by the
352 observation of circulating SARS-CoV-2 S1 spike protein in a subset of patients with
353 Long Covid months after first infection³⁹. The association between persistent

354 infection and Long Covid does not imply that every persistent infection can lead to
355 Long Covid, only 9% of persistently infected individuals reported having Long Covid,
356 nor does it mean that all cases of Long Covid are due to a persistent infection.
357 Indeed, many other possible mechanisms have been suggested to contribute to
358 Long Covid including autoimmunity/inflammation, organ damage, EBV reactivation,
359 and micro thrombosis (see ref ¹⁰ for a recent review).

360
361 Taken together, our observations highlight the continuing importance of community-
362 based genomic surveillance both to monitor the emergence and spread of new
363 variants, but also to gain a fundamental understanding of the natural history and
364 evolution of novel pathogens and their clinical implications for patients.

365 366 **Methods**

367 ONS COVID-19 Infection Survey

368
369 The ONS-CIS is a UK household-based surveillance study in which participant
370 households are approached at random from address lists across the country to
371 provide a representative sample of the population ⁴⁰. All individuals aged two years
372 and older from each household who provide written informed consent provide swab
373 samples (taken by the participant or parent/carer for those under 12 years),
374 regardless of symptoms, and complete a questionnaire at assessments, which occur
375 weekly for the first month in the survey and then monthly. From 26 April 2020 to 31
376 July 2022, assessments were conducted by study workers visiting each household;
377 from 14 July 2022 onwards assessments were remote, with swabs taken using kits
378 posted to participants and returned by post or courier, and questionnaires completed
379 online or by telephone. Positive swab samples with $Ct \leq 30$ were sent for sequencing
380 (see below). For this analysis, we included data from 2nd November 2020 to 15th
381 August 2022, spanning a period from the earliest Alpha to latest Omicron BA.2
382 sequences within the ONS-CIS dataset.
383
384

385
386 This work contains statistical data from ONS which is Crown Copyright. The use of
387 the ONS statistical data in this work does not imply the endorsement of the ONS in
388 relation to the interpretation or analysis of the statistical data. This work uses
389 research datasets which may not exactly reproduce National Statistics aggregates.

390 Sequencing

391
392 From December 2020 onwards sequencing was attempted on all positive samples
393 with $Ct \leq 30$; before this date, sequencing was attempted in real-time wherever
394 possible, with some additional retrospective sequencing of stored samples. The vast
395 majority of samples were sequenced on Illumina Novaseq, with a small number
396 using Oxford Nanopore GridION or MINION. One of two protocols were used: either
397

398 the ARTIC amplicon protocol ⁴¹ with consensus FASTA sequence files generated
399 using the ARTIC nextflow processing pipeline ⁴², or veSeq, an RNASeq protocol
400 based on a quantitative targeted enrichment strategy ^{19,43} with consensus sequences
401 produced using *shiver* ⁴⁴. During our study period, we identified 94,943 individuals
402 with a single sequence and 5,774 individuals with two or more sequences. Here, we
403 only included sequences with $\geq 50\%$ genome coverage.

404

405 Identifying rare SNPs

406

407 An important criterion for determining whether two sequences from the same
408 individual are from the same infection is whether they share a rare single nucleotide
409 polymorphism (SNP). These are defined as SNPs that are shared by fewer than 400
410 sequences corresponding to each major lineage within the full ONS-CIS dataset
411 (**Supplementary Figure 1**). The thresholds were chosen to maximise the number of
412 persistent infections identified whilst minimising the number of false positives (see
413 below). The major lineages we considered were Alpha (B.1.1.7), Delta (B.1.617.2),
414 Omicron BA.1 and Omicron BA.2, including their sublineages. Approximately 92-
415 98% of all sequences from the four major lineages had a rare SNP relative to the
416 major-lineage population-level consensus.

417

418 Identifying reinfections with the same major lineage

419

420 Any pair of sequences from the same individual, of the same major lineage, and at
421 least 26 days apart were considered as candidate reinfections. Of these, pairs that
422 had at least one nucleotide difference at the consensus level, and did not share any
423 rare SNPs, were classed as reinfections. Pairs that had no identical rare SNPs, nor
424 any nucleotide differences at the consensus level, were classed as undetermined.

425

426 Identifying individuals with persistent infection

427

428 We first identified individuals with two or more sequenced samples taken at least 26
429 days apart. We chose this cutoff because the majority of acutely infected individuals
430 shed the virus for <20 days and no longer than 30 days in the respiratory tract ^{23,45}.
431 Given the extreme heterogeneity in the shedding profiles during some acute
432 infections ²³, we also considered a more conservative 56-day cutoff for some
433 analyses.

434

435 For each identified individual we calculated the number of consensus nucleotide
436 differences per site for all within-individual pairwise combinations of samples
437 (**Supplementary Figure 6**). Only sites where a nucleotide difference could be called
438 were included.

439

440 Candidate persistent infections were defined in one of two ways: (1) pairs of
441 sequenced samples that belonged to the same major lineage, and (2) pairs of

442 sequenced samples where one or both had no defined phylogenetic lineage, but
443 where the genetic distance between them was lower than that required to
444 differentiate two major lineages (see **Supplementary Figure 6**). We assumed pairs
445 belonging to different major lineages were either coinfections or reinfections with two
446 different virus lineages. Only candidate persistent infections were considered in
447 further analysis.

448
449 Among the pool of candidate persistent infections, we defined persistent infections
450 as those with sequences sharing one or more rare SNPs at two or more consecutive
451 time points relative to the population-level consensus. A rare SNP is one that is
452 observed in 400 or fewer samples within the entire ONS-CIS dataset.

453 454 Determining the false positive rate for persistent infections

455
456 For each major lineage we generated a data set of 1,000 randomly paired
457 sequences from different individuals in the ONS-CIS, each sampled at least 26 days
458 apart. We determined the number of these pairs that would have been incorrectly
459 identified as persistent infections as a function of the threshold for determining if a
460 SNP is rare (**Supplementary Figure 1**). Although the total number of persistent
461 infections identified grew as the threshold for determining if a SNP is rare increased,
462 at very high thresholds the rate of false positives was also high. In our study, we
463 chose a threshold of 400 sequences (corresponding to all sequences of the same
464 major lineage within the full ONS-CIS dataset) for all of the major lineages, giving a
465 false positive rate (identifying an infection as persistent when it was not) of 0-3%.

466 467 Estimating the prevalence of persistent infections

468
469 Within the ONS-CIS we identified 54 infections that lasted 60 days or more.
470 Comparing this to the number of individuals that had sequenced samples belonging
471 to Alpha, Delta, BA.1 or BA.2, which to a good approximation will be the number of
472 infections of these variants, we identified approximately 54/77,561 (0.07%) infections
473 as persistent for ≥ 60 days. Since the ONS-CIS is a representative sample of
474 individuals from the general population, we can estimate the percentage of all SARS-
475 CoV-2 infections that became persistent for two months or longer. At one extreme, if
476 all persistent infections have sequenceable virus for only four days per month
477 (assuming viral dynamics similar to one acute infection each month), only 14% of
478 persistent infections would be detected through monthly sampling. Correcting for
479 this, we would estimate the percentage of persistent infections in the general
480 population to be 0.5% ($0.07\% / 0.14$). At the other extreme, if we assume all
481 persistent-chronic infections (70% of persistent infections; see Main text) are
482 detectable through monthly sampling, and the rest have detectable virus for 4 days
483 per month, then we estimate that 74% ($70\% + 0.14 \cdot 30\%$) of persistent infections
484 were identified, giving an estimate of 0.09% ($0.07\% / 0.74$) infections being
485 persistent in the general population.

486

487 Phylogenetic analysis

488

489 For each of the four major lineages, we chose 600 consensus sequences with at
490 least 95% coverage from the ONS-CIS dataset using weighted random sampling,
491 with each sample of major lineage i collected in week j given a weight $1/n_{ij}$, where n_{ij}
492 is the number of sequences of major lineage i collected during week j ²⁵. These
493 sequences were added as a background set to the collection of all consensus
494 sequences for samples from persistent infections and reinfections. Mapping of each
495 sequence to the Wuhan-Hu-1 reference sequence was already performed by *shiver*
496 and thus a full alignment for each of the four lineages could be constructed using
497 only this.

498

499 Maximum likelihood phylogenetic trees were constructed using IQ-TREE 1.6.12⁴⁶
500 using the GTR+gamma substitution model and the ultrafast bootstrap⁴⁷. Each tree
501 was rooted using the collection dates of the samples and the heuristic residual mean
502 square algorithm in TempEst⁴⁸. Visualisation used ggtree⁴⁹.

503

504 Comparing viral load activities and symptoms

505

506 To quantify the changes in viral load activities during persistent infections, we
507 compared Ct values at the last time point a sequence was obtained to when the first
508 sequence was collected. Similarly, for reinfections, we compared the changes in Ct
509 value between the primary infection and reinfection. We used a paired t-test to
510 calculate p-values in both cases as the distribution of differences in Ct values were
511 normally distributed for both persistent infections ($W=0.99$, $p=0.28$) and reinfections
512 ($W=0.99$, $p=0.78$) as determined by the Shapiro-Wilk test⁵⁰.

513

514 We also tracked 12 symptoms consistently solicited from all participants at every
515 assessment. Symptoms were fever, weakness/tiredness, diarrhoea, shortness of
516 breath, headache, nausea/vomiting, sore throat, muscle ache, abdominal pain,
517 cough, loss of smell, and loss of taste. At each follow-up assessment, participants
518 were asked whether these 12 symptoms had been present in the past seven days.
519 Symptom discontinuation was defined as the first occurrence of two successive
520 follow-up visits without reporting symptoms. To compare symptom counts during
521 persistent infections and reinfections, we used the paired Wilcoxon test as the
522 distribution of symptom differences is not normally distributed (see **Figure 3d**).

523

524 Long Covid analysis

525

526 From February 2021, at every assessment, participants were asked “would you
527 describe yourself as having Long Covid, that is, you are still experiencing symptoms
528 more than 4 weeks after you first had COVID-19, that are not explained by
529 something else?”. When estimating Long Covid prevalence in this analysis, we

530 considered the first assessment at least 12 weeks and at least 26 weeks after
531 infection. Our control group comprised all individuals with a positive PCR test and
532 $Ct \leq 30$, excluding the persistently infected individuals identified in this study, over the
533 same time span as persistent infections. We also ensured the follow-up from the
534 start of infection to first Long Covid response was similar between persistent
535 infections and controls (see **Table 2**).

536

537 In calculating the odds ratio of Long Covid in persistently infected individuals relative
538 to the control group, we accounted for confounding variables such as age at the last
539 birthday, sex, Ct value, calendar date, area deprivation quintile group, presence of
540 self-reported long-term health conditions (binary), vaccination status (unvaccinated
541 or single-vaccinated, fully-vaccinated or booster-vaccinated 14-89 days ago, fully-
542 vaccinated or booster-vaccinated 90-179 days ago, fully-vaccinated or booster-
543 vaccinated ≥ 180 days ago), and days from first positive test to Long Covid follow-up
544 response. All variables except the last one were defined at the time of the first
545 positive test. Continuous variables (age, Ct value, calendar date, days to follow-up
546 response) were modelled as restricted cubic splines with a single internal knot at the
547 median of the distribution and boundary knots at the 5th and 95th percentiles.
548 Vaccination status was derived from a combination of CIS and NIMS data for
549 participants in England, and CIS data alone for participants in Wales, Scotland and
550 Northern Ireland.

551

552 We were unable to do the Long Covid analysis for the reinfection group due to the
553 low number of participants in this cohort who reported new-onset Long Covid ≥ 12
554 weeks or ≥ 26 weeks after infections.

555

556

557 **Funding Statement**

558

559 The CIS was funded by the Department of Health and Social Care and the UK
560 Health Security Agency with in-kind support from the Welsh Government, the
561 Department of Health on behalf of the Northern Ireland Government and the Scottish
562 Government. COG-UK is supported by funding from the Medical Research Council
563 (MRC) part of UK Research & Innovation (UKRI), the National Institute of Health
564 Research (NIHR) (grant code: MC_PC_19027), and Genome Research Limited,
565 operating as the Wellcome Sanger Institute. The authors acknowledge use of data
566 generated through the COVID-19 Genomics Programme funded by the Department
567 of Health and Social Care. ASW is supported by the National Institute for Health
568 Research Health Protection Research Unit (NIHR HPRU) in Healthcare Associated
569 Infections and Antimicrobial Resistance at the University of Oxford in partnership
570 with the UK Health Security Agency (UK HSA) (NIHR200915) and the NIHR Oxford
571 Biomedical Research Centre, and is an NIHR Senior Investigator. TH is supported
572 by the Royal Society and Alan Turing Institute for Data Science and Artificial
573 Intelligence. KAL is supported by the Royal Society and the Wellcome Trust

574 (107652/Z/15/Z). The research was supported by the Wellcome Trust Core Award
575 Grant Number 203141/Z/16/Z with funding from the NIHR Oxford BRC. The views
576 expressed are those of the author(s) and not necessarily those of the NHS, the NIHR
577 or the Department of Health. The views expressed are those of the author and not
578 necessarily those of the Department of Health and Social Care or UKHSA.

579

580 **Data and materials availability**

581

582 All genomic data have been made publicly available as part of the COVID-19
583 Genomics UK (COG-UK) Consortium ⁵¹. All other data, excluding personal clinical
584 information on participants, are available in the main text, supplementary materials,
585 or our GitHub repository (https://github.com/mg878/ONS-CIS_analysis).

586

587 **References**

588

- 589 1. Dennehy, J. J., Gupta, R. K., Hanage, W. P., Johnson, M. C. & Peacock, T. P. Where is
590 the next SARS-CoV-2 variant of concern? *Lancet* **399**, 1938–1939 (2022).
- 591 2. Otto, S. P. *et al.* The origins and potential future of SARS-CoV-2 variants of concern in
592 the evolving COVID-19 pandemic. *Curr. Biol.* **31**, R918–R929 (2021).
- 593 3. Gonzalez-Reiche, A. S. *et al.* SARS-CoV-2 variants in the making: Sequential intrahost
594 evolution and forward transmissions in the context of persistent infections. *medRxiv*
595 2022.05.25.22275533 (2022) doi:10.1101/2022.05.25.22275533.
- 596 4. Hill, V. *et al.* The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the
597 UK. *Virus Evolution* **8**, veac080 (2022).
- 598 5. Ghafari, M., Liu, Q., Dhillon, A., Katzourakis, A. & Weissman, D. B. Investigating the
599 evolutionary origins of the first three SARS-CoV-2 variants of concern. *Frontiers in*
600 *Virology Section: Diversification and Evolution*, (2022).
- 601 6. Choi, B. *et al.* Persistence and Evolution of SARS-CoV-2 in an Immunocompromised
602 Host. *N. Engl. J. Med.* **383**, 2291–2293 (2020).
- 603 7. Kemp, S. A. *et al.* SARS-CoV-2 evolution during treatment of chronic infection. *Nature*
604 **592**, 277–282 (2021).
- 605 8. Chaguza, C. *et al.* Accelerated SARS-CoV-2 intrahost evolution leading to distinct
606 genotypes during chronic infection. *medRxiv* (2022) doi:10.1101/2022.06.29.22276868.
- 607 9. Swank, Z. *et al.* Persistent Circulating Severe Acute Respiratory Syndrome Coronavirus
608 2 Spike Is Associated With Post-acute Coronavirus Disease 2019 Sequelae. *Clin. Infect.*
609 *Dis.* ciac722 (2022) doi:10.1093/cid/ciac722.
- 610 10. Davis, H. E., McCorkell, L., Vogel, J. M. & Topol, E. J. Long COVID: major findings,
611 mechanisms and recommendations. *Nat. Rev. Microbiol.* (2023) doi:10.1038/s41579-
612 022-00846-2.
- 613 11. Starr, T. N. *et al.* Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain
614 Reveals Constraints on Folding and ACE2 Binding. *Cell* **182**, 1295–1310.e20 (2020).

- 615 12. McCarthy, K. R. *et al.* Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive
616 antibody escape. *Science* **371**, 1139–1142 (2021).
- 617 13. Harari, S. *et al.* Drivers of adaptive evolution during chronic SARS-CoV-2 infections.
618 *Nature Medicine* **28**, 1501–1508 (2022).
- 619 14. Wilkinson, S. A. J. *et al.* Recurrent SARS-CoV-2 mutations in immunodeficient patients.
620 *Virus Evol* **8**, veac050 (2022).
- 621 15. Rambaut, A. *et al.* Preliminary genomic characterisation of an emergent SARS-CoV-2
622 lineage in the UK defined by a novel set of spike mutations. *Virological*
623 [https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-](https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563)
624 [lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563](https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563) (2020).
- 625 16. Avanzato, V. A. *et al.* Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an
626 Asymptomatic Immunocompromised Individual with Cancer. *Cell* **183**, 1901–1912.e9
627 (2020).
- 628 17. Cele, S. *et al.* SARS-CoV-2 prolonged infection during advanced HIV disease evolves
629 extensive immune escape. *Cell Host Microbe* **30**, 154–162.e5 (2022).
- 630 18. Nussenblatt, V. *et al.* Yearlong COVID-19 Infection Reveals Within-Host Evolution of
631 SARS-CoV-2 in a Patient With B-Cell Depletion. *J. Infect. Dis.* **225**, 1118–1123 (2022).
- 632 19. Lythgoe, K. A. *et al.* SARS-CoV-2 within-host diversity and transmission. *Science* **372**,
633 (2021).
- 634 20. Braun, K. M. *et al.* Acute SARS-CoV-2 infections harbor limited within-host diversity and
635 transmit via tight transmission bottlenecks. *PLoS Pathog.* **17**, e1009849 (2021).
- 636 21. Pouwels, K. B. *et al.* Community prevalence of SARS-CoV-2 in England from April to
637 November, 2020: results from the ONS Coronavirus Infection Survey. *Lancet Public*
638 *Health* **6**, e30–e38 (2021).
- 639 22. Kissler, S. M. *et al.* Viral Dynamics of SARS-CoV-2 Variants in Vaccinated and
640 Unvaccinated Persons. *N. Engl. J. Med.* **385**, 2489–2491 (2021).
- 641 23. Jones, T. C. *et al.* Estimating infectiousness throughout SARS-CoV-2 infection course.
642 *Science* **373**, (2021).

- 643 24. Ghafari, M. *et al.* Purifying selection determines the short-term time dependency of
644 evolutionary rates in SARS-CoV-2 and pH1N1 influenza. Preprint at
645 <https://doi.org/10.1101/2021.07.27.21261148>.
- 646 25. Lythgoe, K. A. *et al.* Lineage replacement and evolution captured by the United
647 Kingdom Covid Infection Survey. *medRxiv* (2022) doi:10.1101/2022.01.05.21268323.
- 648 26. Neher, R. A. Contributions of adaptation and purifying selection to SARS-CoV-2
649 evolution. *Virus Evol* **8**, veac113 (2022).
- 650 27. Hodcroft, E. B. CoVariants: SARS-CoV-2 Mutations and Variants of Interest.
651 <https://covariants.org/> (2021).
- 652 28. Cox, M. *et al.* SARS-CoV-2 variant evasion of monoclonal antibodies based on in vitro
653 studies. *Nat. Rev. Microbiol.* **21**, 112–124 (2023).
- 654 29. Lin, J.-W. *et al.* Genomic monitoring of SARS-CoV-2 uncovers an Nsp1 deletion variant
655 that modulates type I interferon response. *Cell Host Microbe* **29**, 489–502.e8 (2021).
- 656 30. Tom, M. R. & Mina, M. J. To Interpret the SARS-CoV-2 Test, Consider the Cycle
657 Threshold Value. *Clinical infectious diseases: an official publication of the Infectious*
658 *Diseases Society of America* vol. 71 2252–2254 (2020).
- 659 31. Hay, J. A. *et al.* Estimating epidemiologic dynamics from cross-sectional viral load
660 distributions. *Science* **373**, (2021).
- 661 32. Fryer, H. R. *et al.* Viral burdens are associated with age and viral variant in a population-
662 representative study of SARS-CoV-2 that accounts for time-since-infection related
663 sampling bias. Preprint at <https://doi.org/10.1101/2022.12.02.518847> (2022).
- 664 33. Yates, F. Contingency tables involving small numbers and the χ^2 test. *Suppl. J. R. Stat.*
665 *Soc.* **1**, 217 (1934).
- 666 34. Karthikeyan, S. *et al.* Wastewater sequencing reveals early cryptic SARS-CoV-2 variant
667 transmission. *Nature* **609**, 101–108 (2022).
- 668 35. Shafer, M. M. *et al.* Tracing the origin of SARS-CoV-2 Omicron-like Spike sequences
669 detected in wastewater. *medRxiv* (2022) doi:10.1101/2022.10.28.22281553.
- 670 36. Lythgoe, K. A., Gardner, A., Pybus, O. G. & Grove, J. Short-Sighted Virus Evolution and

- 671 a Germline Hypothesis for Chronic Viral Infections. *Trends Microbiol.* **25**, 336–348
672 (2017).
- 673 37. Lythgoe, K. A., Lumley, S. F., Pellis, L., McKeating, J. A. & Matthews, P. C. Estimating
674 hepatitis B virus cccDNA persistence in chronic infection. *Virus Evol* **7**, veaa063 (2021).
- 675 38. Marx, V. Scientists set out to connect the dots on long COVID. *Nat. Methods* **18**, 449–
676 453 (2021).
- 677 39. Schultheiß, C. *et al.* Liquid biomarkers of macrophage dysregulation and circulating
678 spike protein illustrate the biological heterogeneity in patients with post-acute sequelae
679 of COVID-19. *J. Med. Virol.* **95**, e28364 (2023).
- 680 40. Pouwels, K. B. *et al.* Community prevalence of SARS-CoV-2 in England from April to
681 November, 2020: results from the ONS Coronavirus Infection Survey. *Lancet Public
682 Health* **6**, e30–e38 (2021).
- 683 41. COG-UK. COG-UK publication. <https://www.protocols.io/workspaces/coguk/publication>
684 (2020).
- 685 42. Connor Lab. *ncov2019-artic-nf*. (Github).
- 686 43. Bonsall, D. *et al.* A Comprehensive Genomics Solution for HIV Surveillance and Clinical
687 Monitoring in Low-Income Settings. *J. Clin. Microbiol.* **58**, (2020).
- 688 44. Wymant, C. *et al.* Easy and accurate reconstruction of whole HIV genomes from short-
689 read sequence data with shiver. *Virus Evol* **4**, vey007 (2018).
- 690 45. Cevik, M. *et al.* SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration
691 of viral shedding, and infectiousness: a systematic review and meta-analysis. *Lancet
692 Microbe* **2**, e13–e22 (2021).
- 693 46. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
694 Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
- 695 47. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2:
696 Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
- 697 48. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal
698 structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus*

- 699 *Evol* **2**, vew007 (2016).
- 700 49. Yu, G., Smith, D. K. & Tsan-Yuk, Z. H. L. ggtree: an R package for visualization and
701 annotation of phylogenetic trees with their covariates and other associated data.
702 *Methods in Ecology and Evolution* **8**, 28–36 (2017).
- 703 50. Shapiro, S. S. & Wilk, M. B. An Analysis of Variance Test for Normality (Complete
704 Samples). *Biometrika* **52**, 591–611 (1965).
- 705 51. COVID-19 Genomics UK (COG-UK) consortiumcontact@cogconsortium.uk. An
706 integrated national scale SARS-CoV-2 genomic surveillance network. *Lancet Microbe* **1**,
707 e99–e100 (2020).
- 708

709 **Tables**

710

711 **Table 1: Frequency of persistent infections and reinfections per major lineage.**

712

Major lineage	Reinfection >26 days	Reinfection >56 days	Persistent infection >26 days	Persistent infections >56 days	%Reinfection >26 days	%Reinfection >56 days
B.1.1.7	7	3	11	3	39%	50%
B.1.617.2	11	4	106	13	9%	24%
BA.1	14	2	97	15	13%	12%
BA.2	28	15	167	23	14%	40%

713

714

715 **Table 2: Prevalence of Long Covid in persistently infected individuals.**

716

717 **Long Covid ≥12 weeks post-infection**

Group	<i>n</i>	Long Covid	Median follow-up (IQR*)	Unadjusted OR† (95% CI**)	Adjusted OR (95% CI)
Exposed	356	32 (9.0%)	101 (91-113)	1.27 (1.19-2.47)	1.55 (1.07-2.25)
Control	78,902	4,291 (5.4%)	100 (91-115)	Reference	Reference

718

719 **Long Covid ≥26 weeks post-infection**

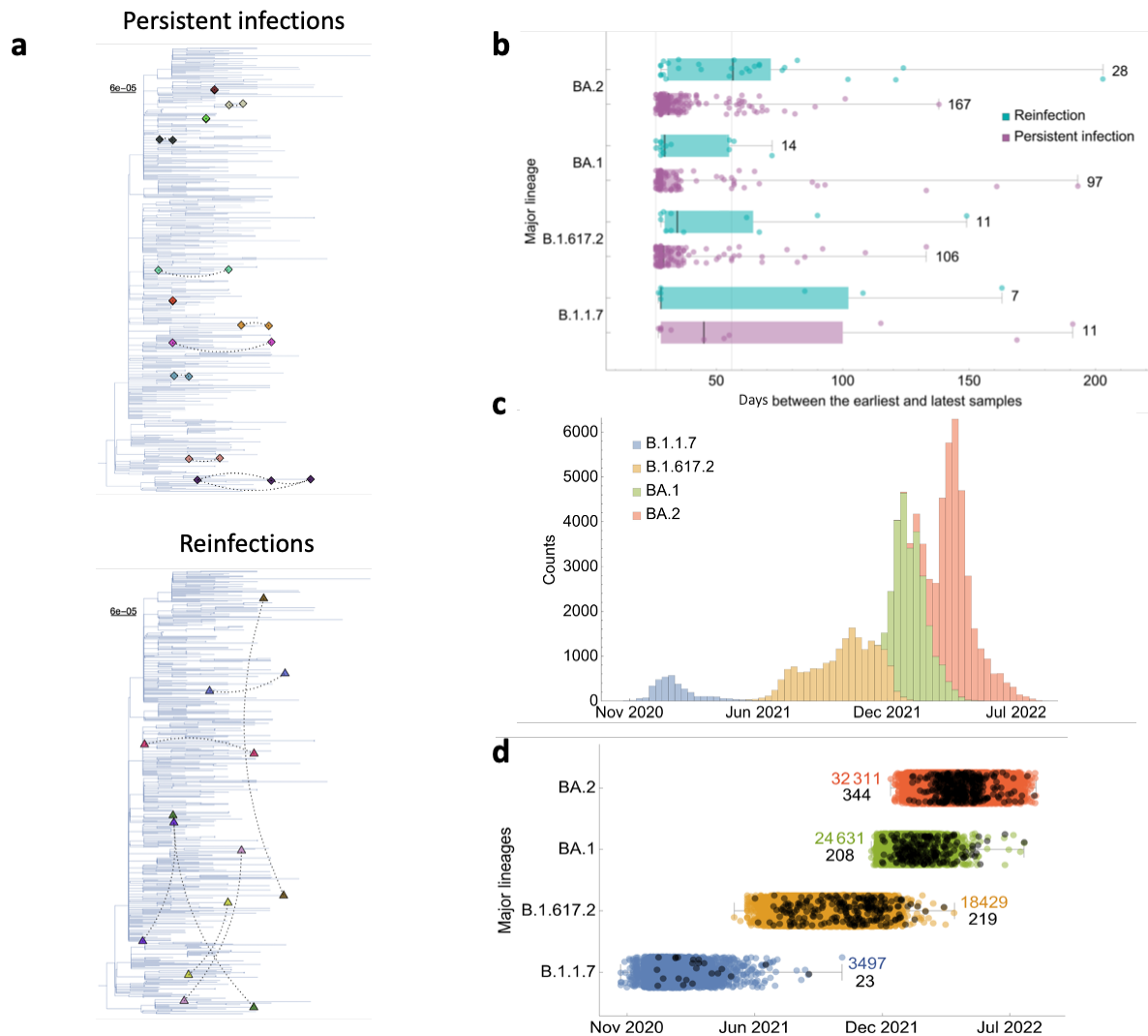
Group	<i>n</i>	Long Covid	Median follow-up (IQR)	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Exposed	326	19 (5.8%)	312 (271-390)	1.44 (0.90-2.29)	1.24 (0.77-2.00)
Control	72,608	3,000 (4.1%)	320 (272-384)	Reference	Reference

720 * IQR=interquartile range

721 †OR=odds ratio

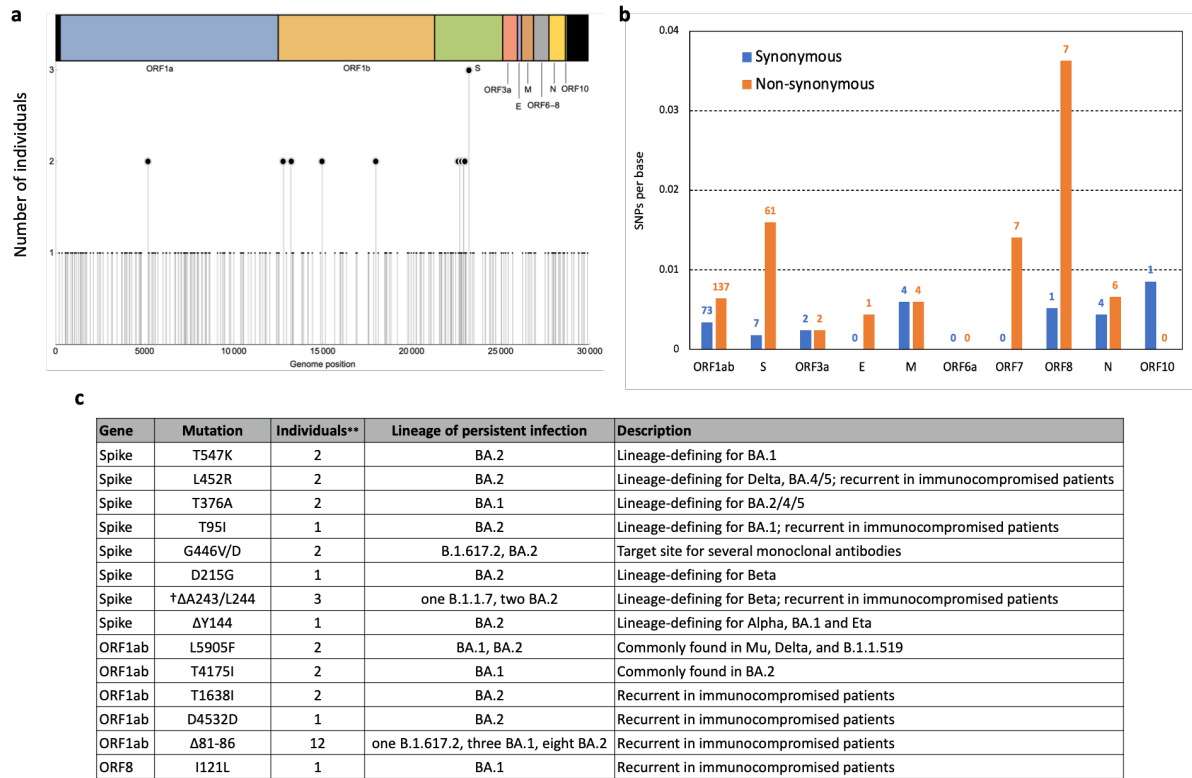
722 **CI=confidence interval

723 **Figures**
724



725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741

Figure 1: Individuals identified with persistent infections and reinfections within the ONS-CIS. (a) Phylogenetic relationship between samples from persistent infections and reinfections with a representative background population of Alpha (see **Supplementary Figure 2** for the analysis on the other three major lineages). Dashed lines connect every pair of sequences from the same individual. Pairs from persistent infections cluster closely together while reinfections do not. All sequences from the same individual are given the same colour. **(b)** Days between the earliest and latest genomic samples from persistent infections and reinfections. Each point represents a single individual. Solid vertical lines show the 26- and 56-day cutoffs. Numbers on the side of each box shows the total counts per category for each major lineage. **(c)** Total number of sequences in the ONS-CIS per major lineage over time. **(d)** Timing of persistent infections (black) during the UK epidemic. Some persistent infections can be identified up to weeks after the lineage has been replaced at the population level. The numbers on the side of each box shows the total sequence counts for each category.



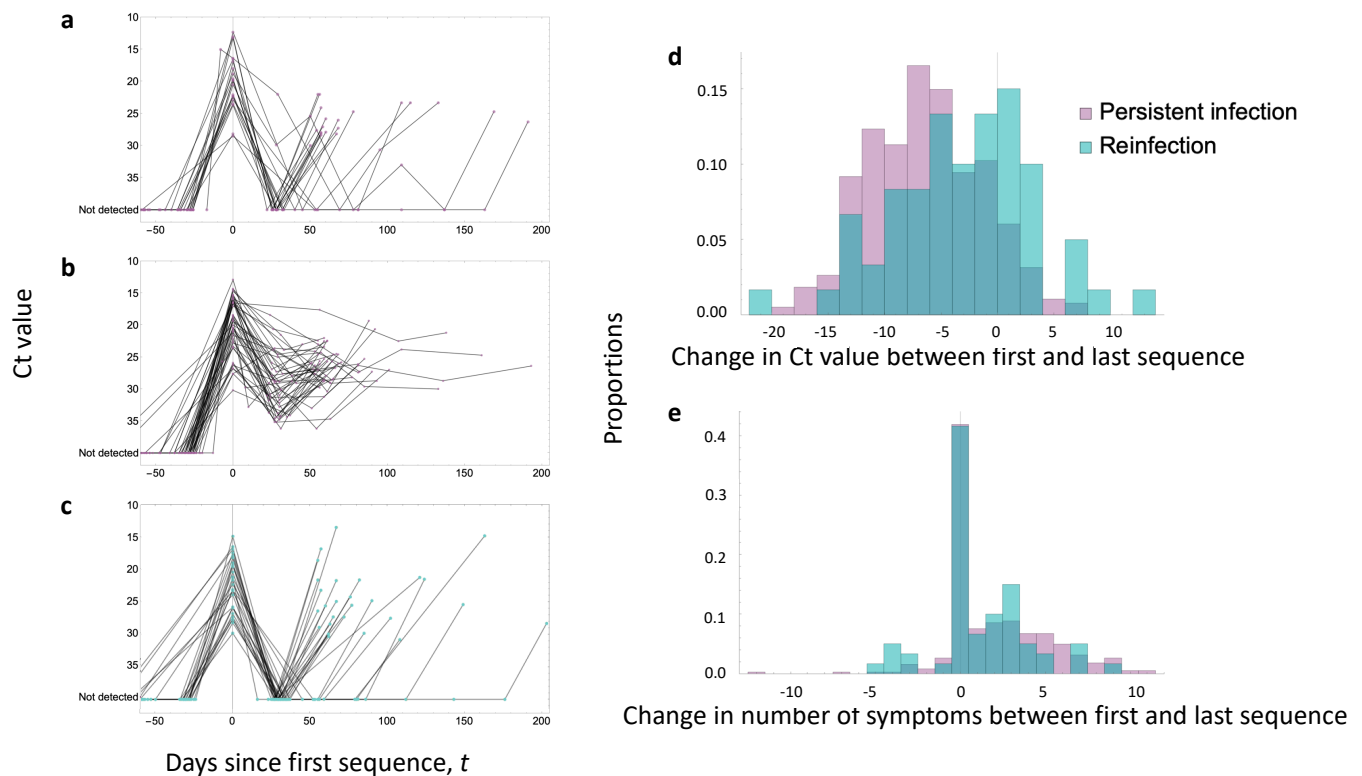
742
743

744 **Figure 2: Distribution of single nucleotide polymorphisms and non-synonymous vs.**
 745 **synonymous mutations detected during persistent infections. (a)** Frequency of
 746 mutations that resulted in a SNP change during one or more persistent infections. **(b)**
 747 Number of synonymous (blue) and non-synonymous (orange) mutations per gene during
 748 persistent infections. Numbers on each column show the total counts of SNPs in each
 749 category of mutations. **(c)** Description of recurrent mutations and deletions identified during
 750 persistent infections. See **Supplementary Table 1** for information about other mutations.

751 **None of the recurrent mutations were from households with other infections.

752 †Δ represents a deletion

753



754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

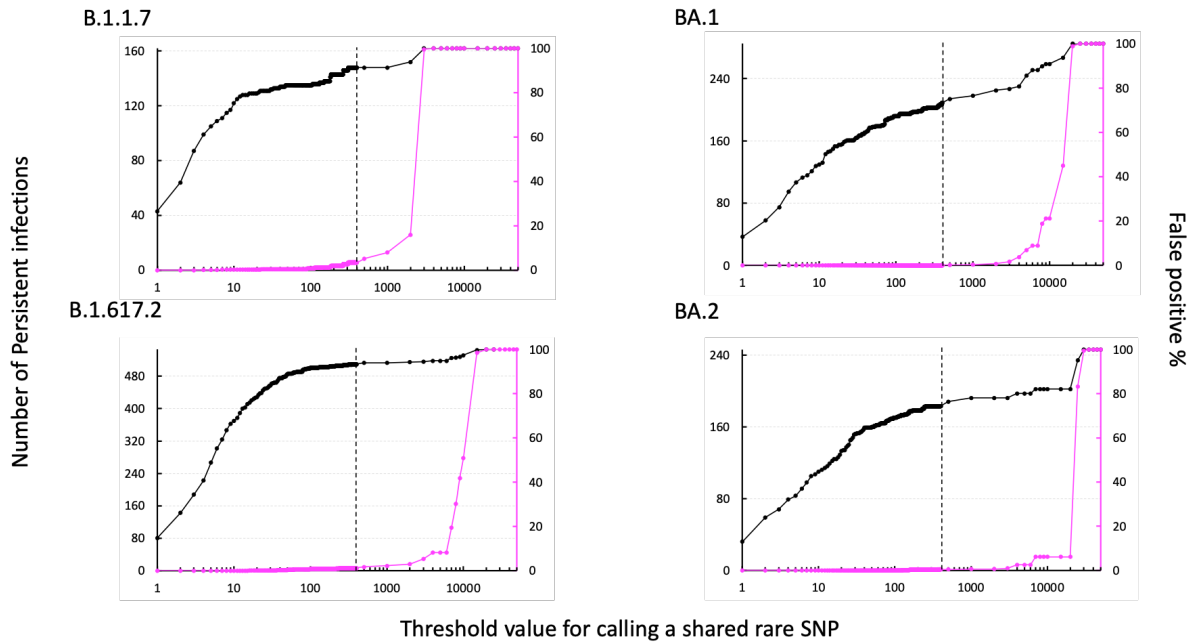
Figure 3: Comparison of viral load dynamics and number of reported symptoms

during persistent infections and reinfections.

Viral load trajectories of persistent infections with (a) relapsing and (b) chronic persistent infections and (c) reinfections with at least three PCR tests taken over the course of infection/until reinfection. (d) Change in cycle threshold (Ct) value and (e) total number of symptoms reported between the first and last time points with sequenced samples for all 381 persistent infections (purple) and 60 reinfections (cyan). Viral load at the last time point is significantly lower for persistent infections with Ct value being more than +6.7 (IQR: +3.2,+10.2) units higher at the last time point (paired t-test $p < 10^{-8}$). The difference is less pronounced for reinfections with +2.5 (IQR: -1.1, +7.4) units difference between primary infection and reinfections (paired t-test $p = 0.0003$). Similarly, persistently infected individuals tend to report fewer number of symptoms at the later stages of their infection compared to reinfections in the last 7 days at the last time point (at which a sequence was obtained) when compared to the first. On average, persistently infected individuals report two more symptoms at the time of their first sequenceable sample (virus samples with a positive $Ct \leq 30$) relative to their last sample, with a median of 1 (IQR: 0, 4) fewer reported symptoms (paired Wilcoxon $p < 10^{-29}$), while reinfected individuals report only one more symptom at the point of primary infection relative to reinfection, with a median of 0 (IQR: 0, 3) fewer reported symptoms (paired Wilcoxon $p = 0.005$).

774 **Supplementary figures**

775



776

777

778

779

780

781

782

783

784

785

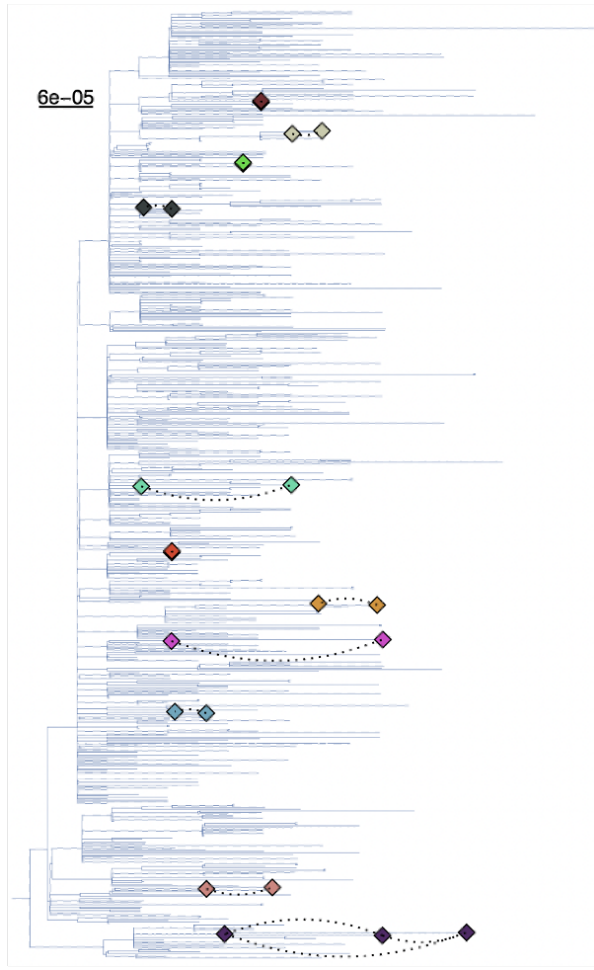
786

787

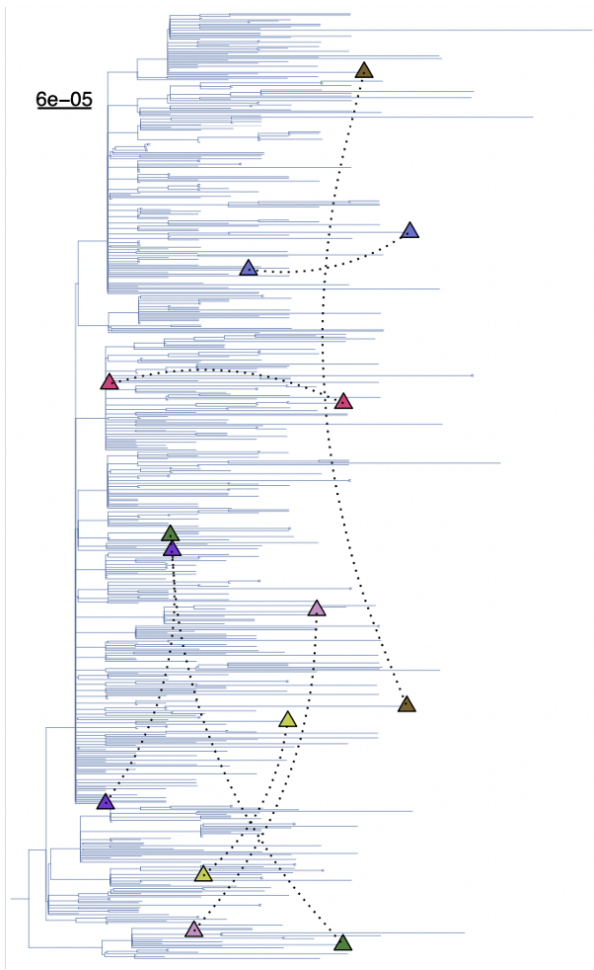
Supplementary Figure 1: Number of persistent infections identified with a shared rare SNP as a function of the threshold for calling a rare SNP. Threshold value of 1 for rare SNPs implies the rare SNP is only found in one sequence of that lineage in the ONS-CIS dataset, excluding sequences from any persistently infected individuals. As the threshold value for calling a rare SNP increases, the number of persistent infections (of any duration) identified (black) increases until at some point the threshold value becomes so high that any individual with two or more sequences of the same major lineage would be identified as a persistent infection based on the rare SNP criterion at which point the false positivity rate (magenta) reaches 100%. At threshold value 400 (vertical dashed line) chosen in this study for identifying persistent infections, the percentage of false positives are 0% for BA.1 and BA.2 and 3% for Alpha and Delta.

a

Persistent infections



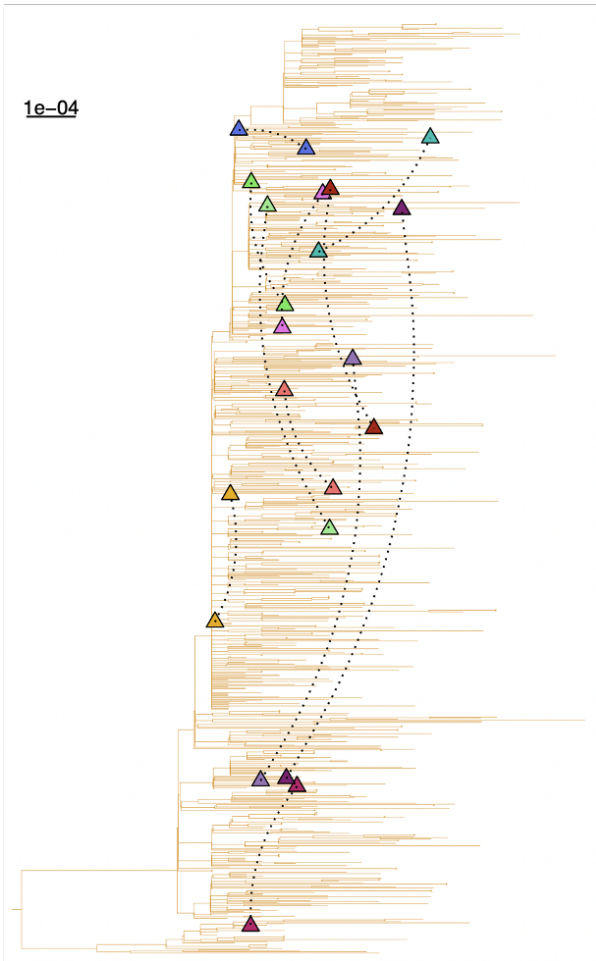
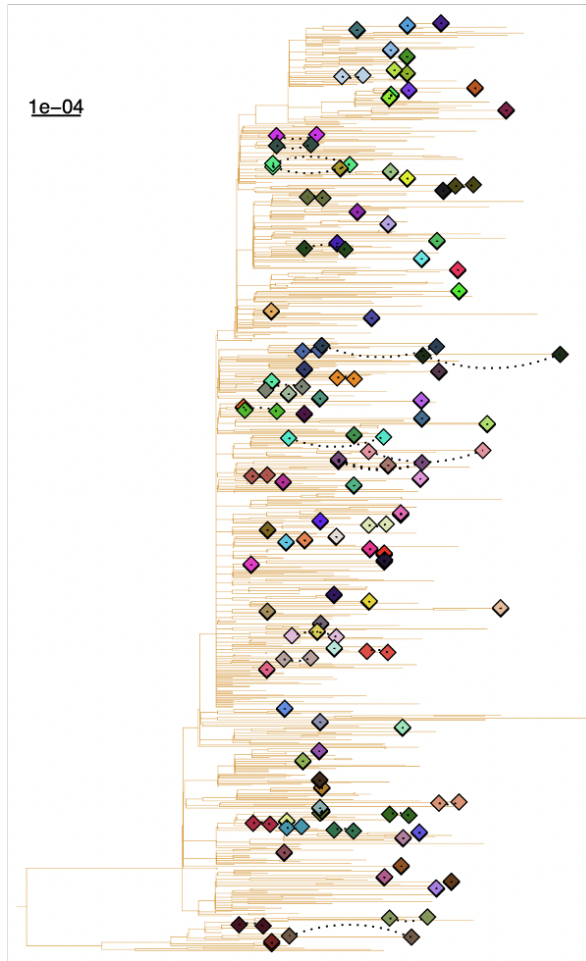
Reinfections



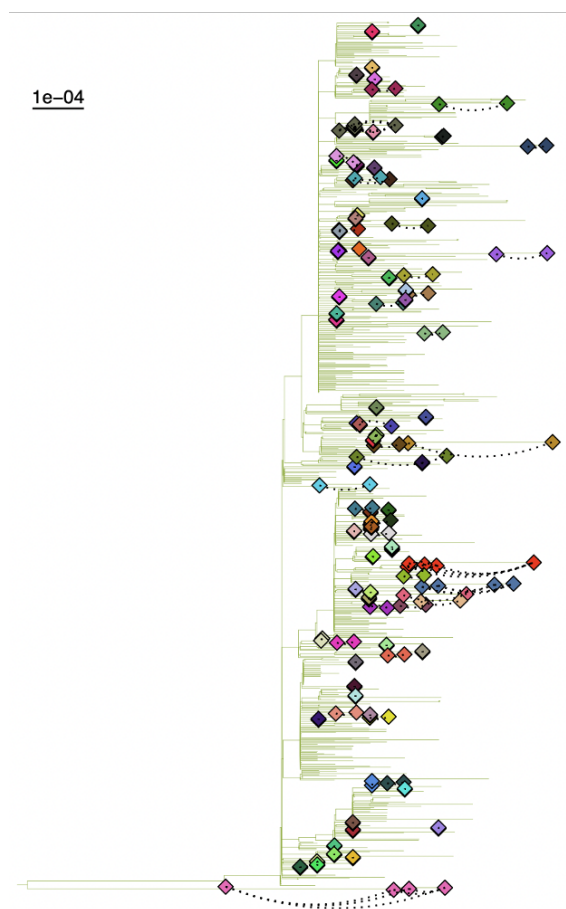
b

Persistent infections

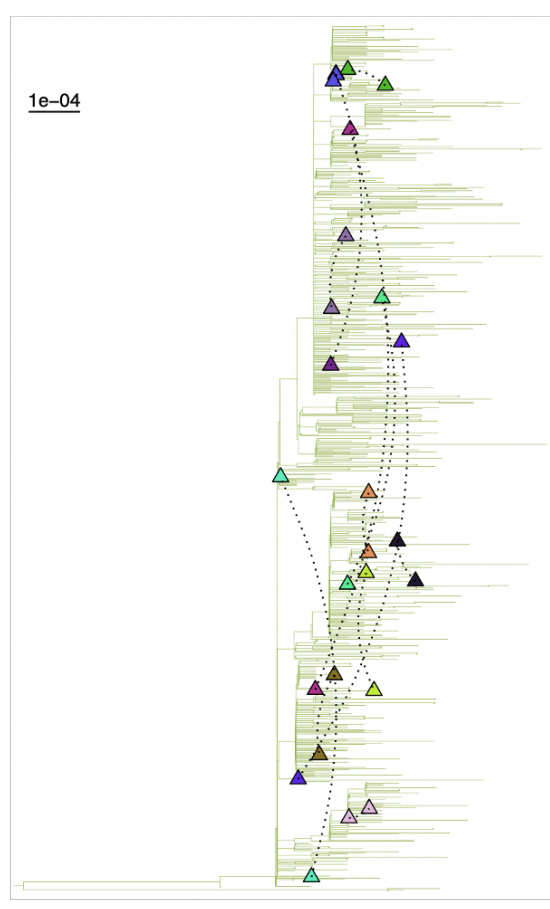
Reinfections



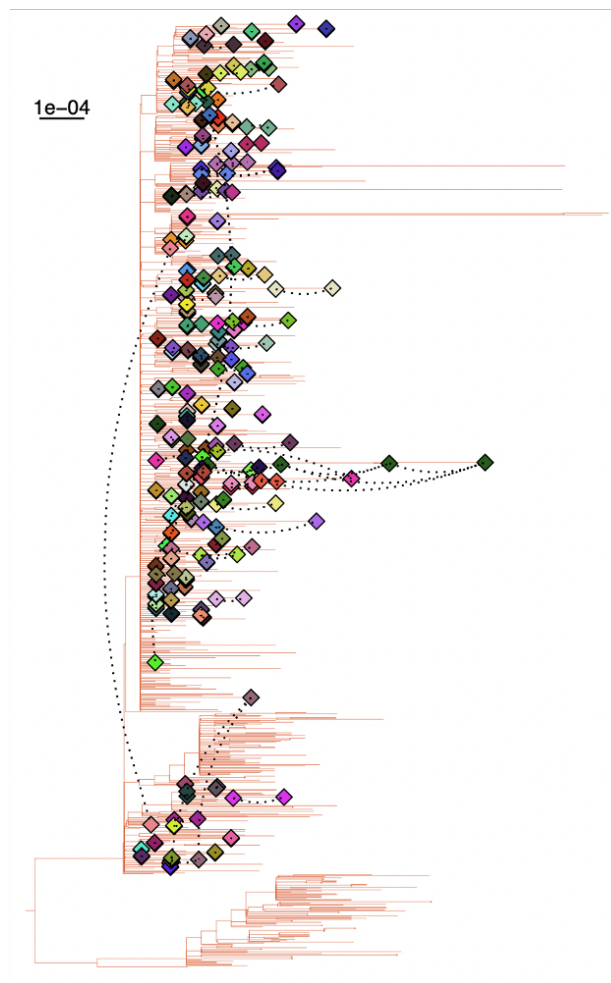
C Persistent infections



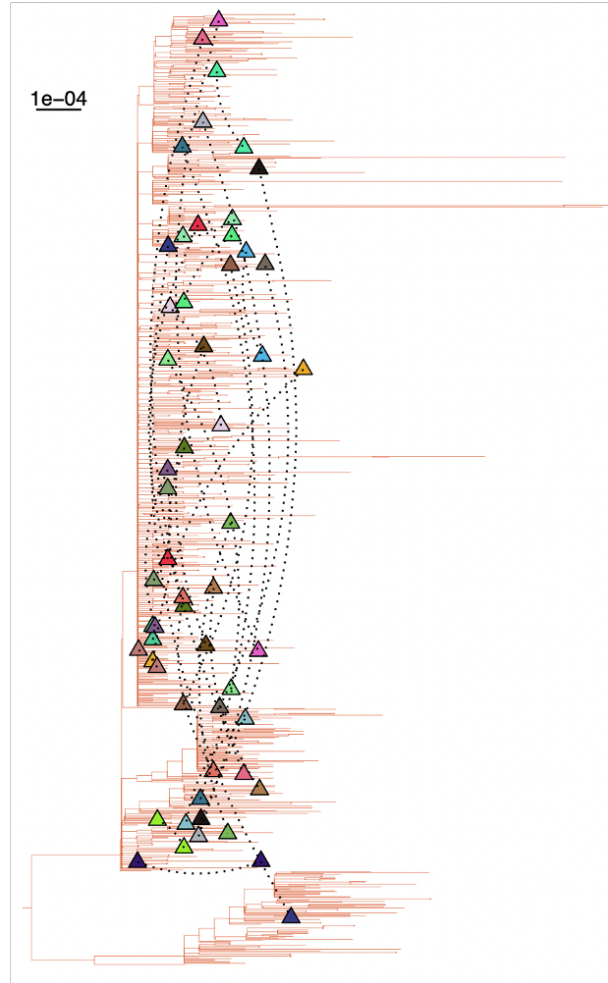
Reinfections



d Persistent infections

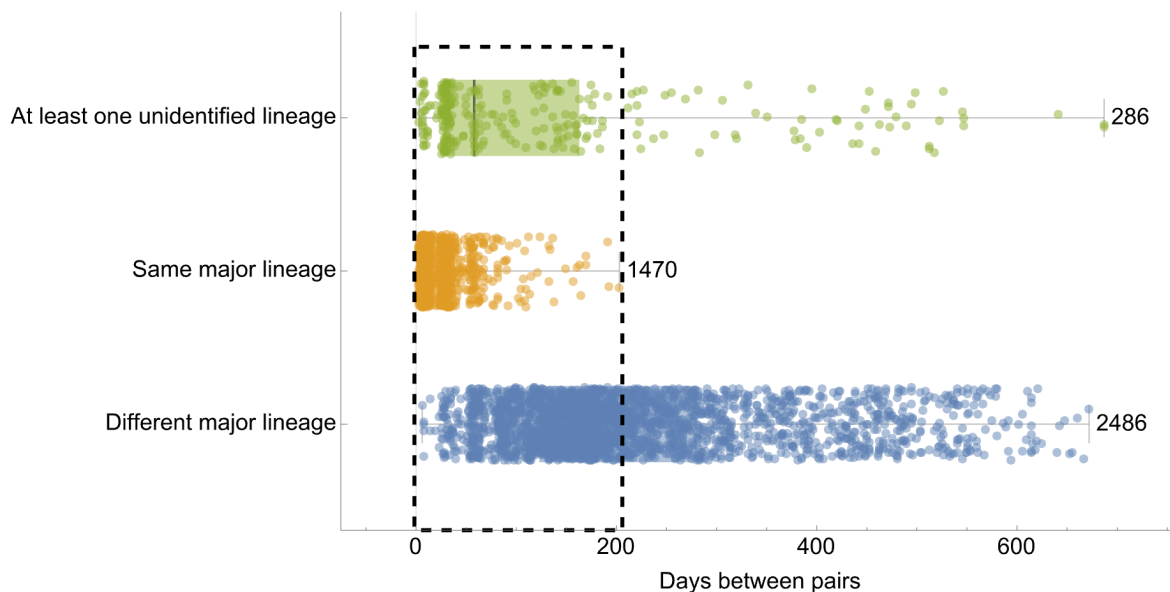


Reinfections

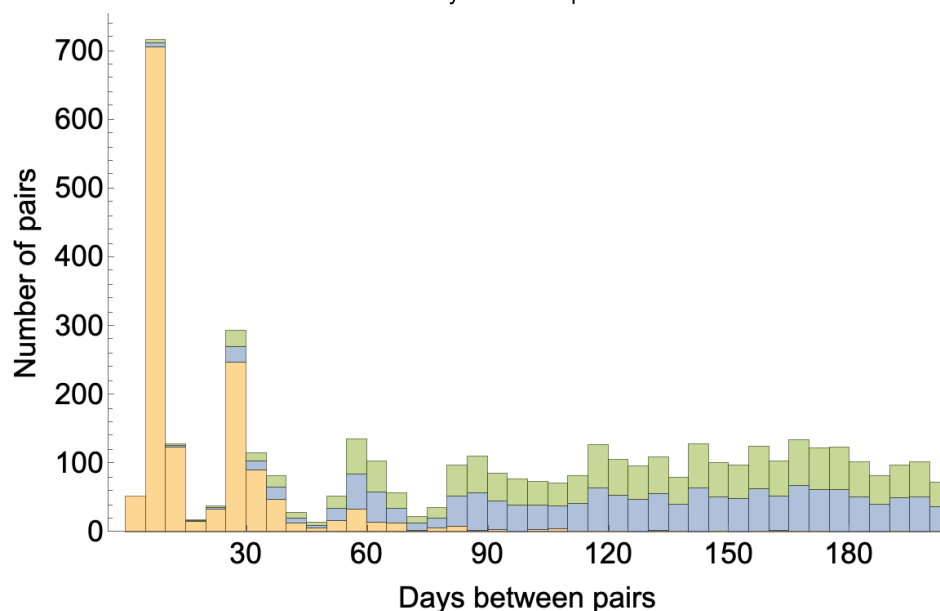


791
792

793 **Supplementary Figure 2: Phylogenetic relationship between samples from persistent**
794 **infections and a representative background population per major lineage.** Dashed lines
795 connect every pair of sequences from the same individual. All sequences from the same
796 individual are given the same colour. Pairs of sequences for (a) Alpha, (b) Delta, (c)
797 Omicron BA.1, and (d) Omicron BA.2 that belong to persistent infections cluster closely
798 together while reinfections do not. However, some of the sequences in 2 (out of 97)
799 persistent infections with BA.1 and 5 (out of 167) persistent infections with BA.2 have poor
800 bootstrap support (<80) and do not cluster together or cluster in a basal sister relationship. In
801 all of these 7 cases, at least one of the sequences from each individual has a Ct value close
802 to 30 with poor coverage. On the other hand, all sequences that belong to the same
803 individual and have strong bootstrap support (>80) cluster together.



804



805

806

807

808

809

810

811

812

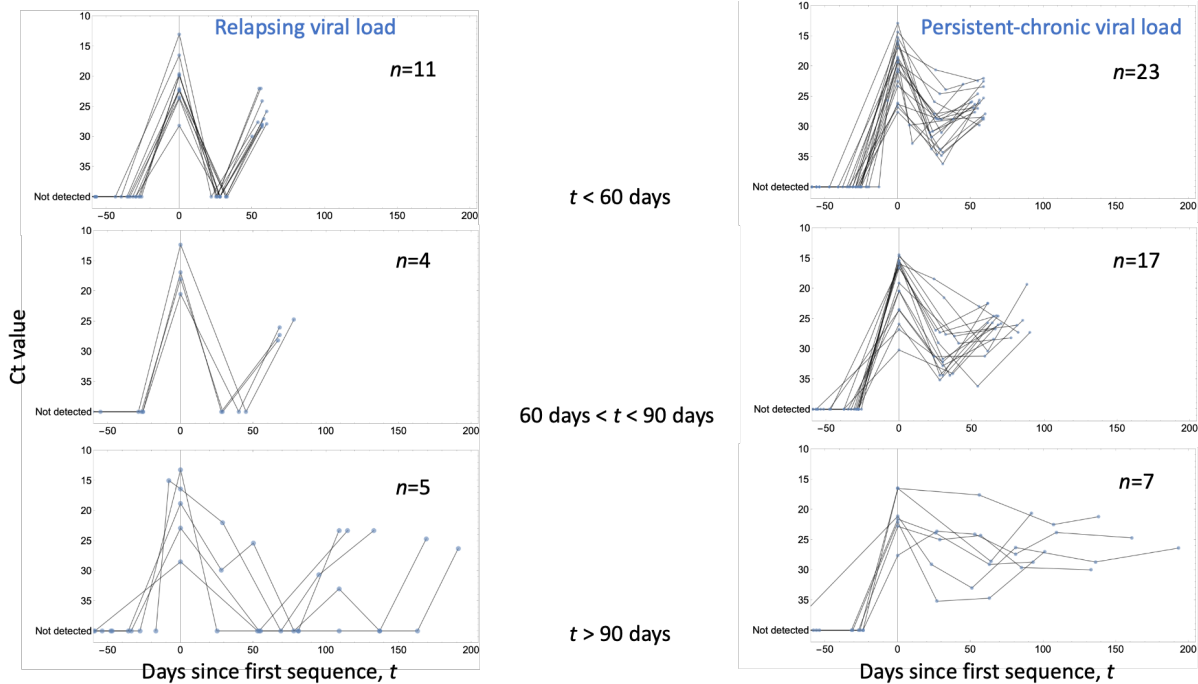
813

814

815

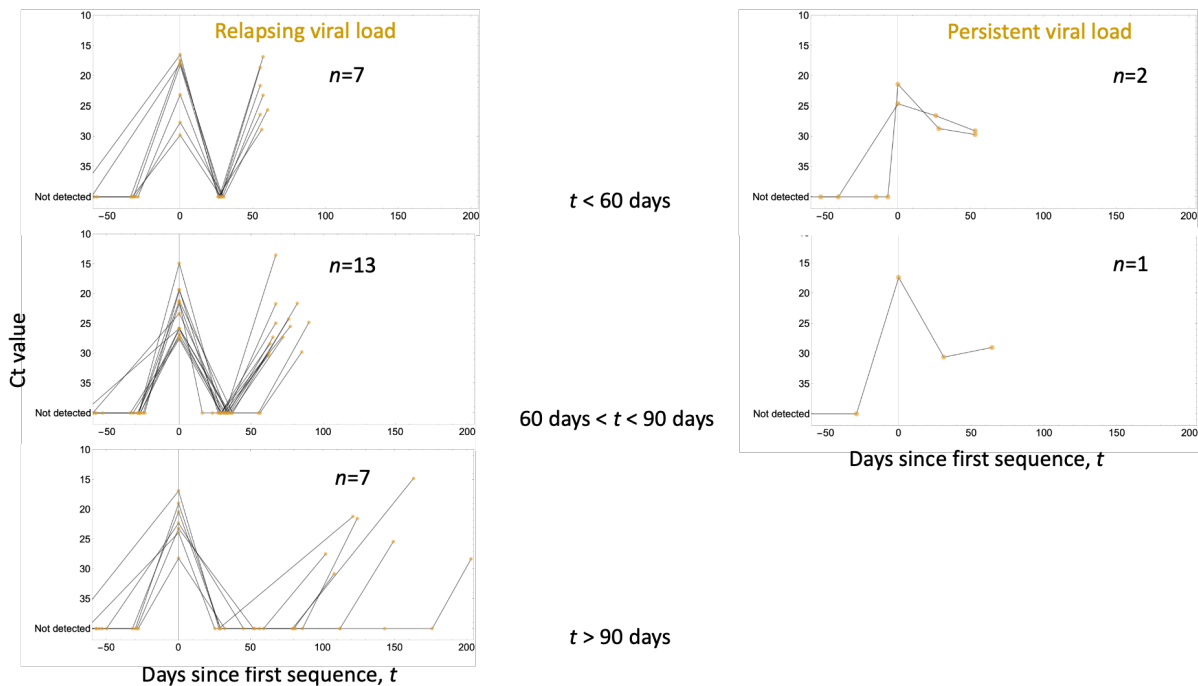
Supplementary Figure 3: Days between all pairs of sequences from the same individual with two or more sequences. Pairs of sequences are classified as (i) pairs with at least one unidentified Pango lineage (green), (ii) pairs with identical major lineage (orange), and (iii) pairs from different major lineages. Bottom panel shows the counts of pairs in each of these three categories for the first 200-day time span (highlighted in a dashed rectangle in the top panel). Pairs include all possible combinations of sequences from the same individual. The number of pairs peaks at the 7-, 30-, and 60-day periods due to the sampling frequency of ONS-CIS (see **Methods**). Note that pairs with identical major lineage may not necessarily have identical Pango lineages (see **Methods**).

816 a



817
818
819

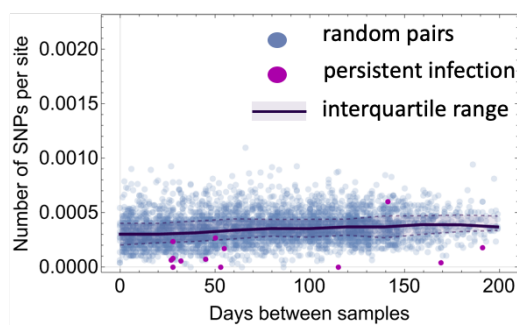
b



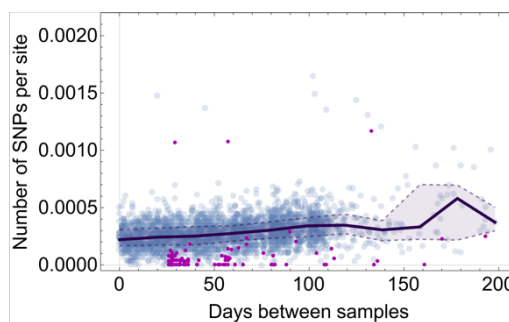
820

821 **Supplementary Figure 4: Viral load dynamics of individuals identified with persistent**
 822 **infections and reinfections stratified by duration and viral activity.** Viral load activities of
 823 individuals, with 3 or more PCR tests taken during infection/until reinfection, identified as
 824 having (a) persistent infections and (b) reinfections with relapsing (left column) and
 825 persistent chronic (right column) trajectories. Three reinfections (two occurring in < 60 days
 826 and one between 60 to 90 days since first sequence) with persistent chronic viral load
 827 dynamics are excluded from the reinfection group as they are deemed potential persistent
 828 infections which do not have rare SNPs.

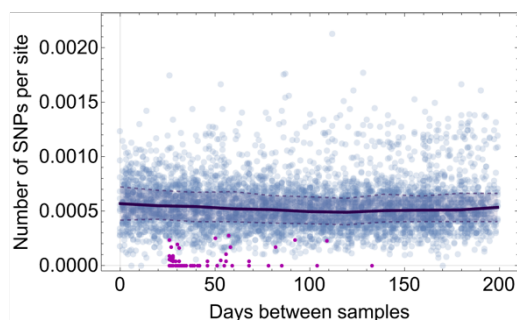
B.1.1.7



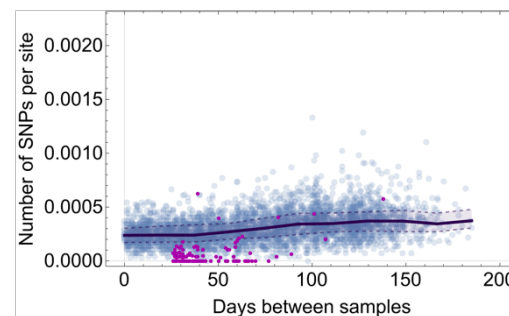
BA.1



B.1.617.2



BA.2



829

830

831

832

833

834

835

836

837

838

839

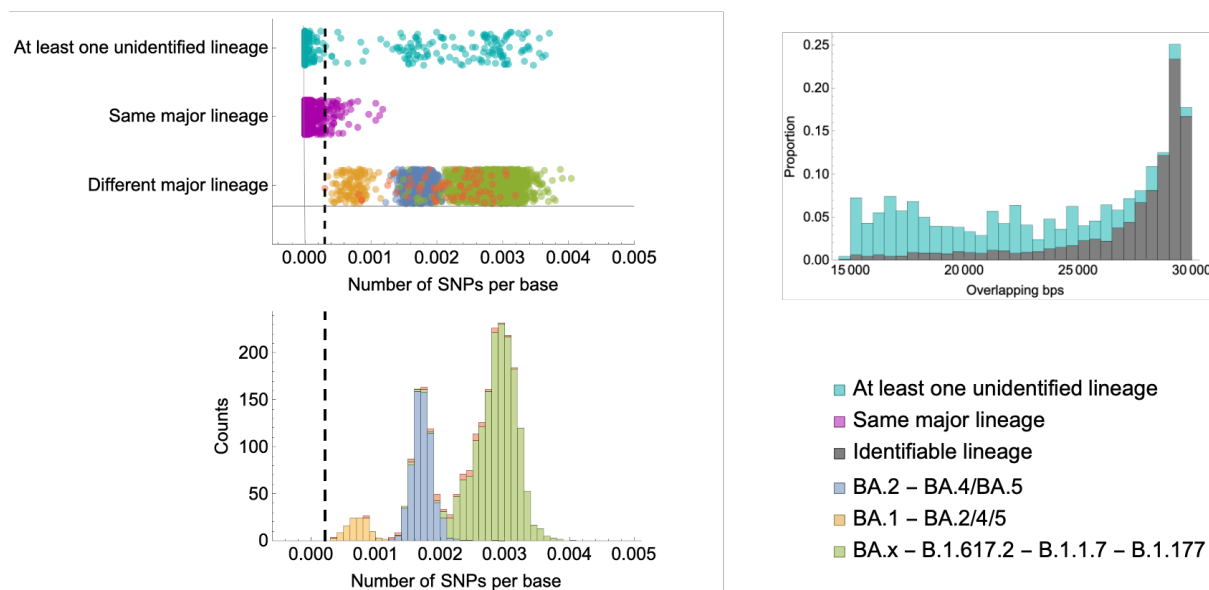
840

841

842

843

Supplementary Figure 5: Number of single nucleotide polymorphisms detected in pairs of sequences from persistent infections vs. random pairs from a representative background population. Number of Single Nucleotide Variants (SNVs) per site between all the sequences collected from persistent infections (purple) and random pairs from individuals with only a single sequence within the ONS-CIS (blue) as a function of the number of days between each pair. For each major lineage, a pool of sequences from individuals with only one sequence within the ONS-CIS was sub-sampled and 500 random pairs generated for every 20 additional days between samples. For some major lineages where there were fewer than 500 pairs available beyond a certain time point, all possible random pairs within that 20-day period are used. Solid line and shaded area show the median and interquartile range, respectively, for random pairs over time. Note that the line and shaded area in each graph does not represent the rate of evolution but can be deemed as a measure of lineage diversity as a function of time difference between samples.



844
845

846 **Supplementary Figure 6: Pairwise differences between sequences from individuals**
847 **with two or more sequences.** (Left column) Number of single nucleotide polymorphisms
848 (SNPs) between pairs of sequences from each individual with two or more sequences. Pairs
849 include all possible combinations of sequences from the same individual. Vertical dashed
850 line shows the lowest number of SNPs per base for pairs with different major lineages. Any
851 pair with at least one unidentified lineage with a SNP per base smaller than the dashed line
852 is selected as a candidate pair from a persistent infection. Pairs with different major lineages
853 are coloured based on their number of SNPs per base into three groups: (i) pairs with one
854 BA.1 and one BA.2 or BA.4 or BA.5 sequence (orange); (ii) pairs with one BA.2 and one
855 BA.4 or BA.5 sequence (blue); and (iii) pairs with one Omicron (including all BA.x lineages)
856 and one Delta (B.1.617.2), Alpha (B.1.1.7), or B.1.177 sequence (green). (Right column)
857 Proportion of sequences with different number overlapping base pairs. Those with at least
858 one unidentified lineage have a lower number of overlapping base pairs relative to pairs with
859 identifiable lineage (i.e. pairs with identical or different major lineage) mainly due to having
860 lower coverage.
861

862 **The COVID-19 Genomics UK (COG-UK) consortium**

863 **June 2021 V.1**

864

865 Funding acquisition, Leadership and supervision, Metadata curation, Project administration,
866 Samples and logistics, Sequencing and analysis, Software and analysis tools, and
867 Visualisation:

868 Dr Samuel C Robson PhD ^{13, 84}

869

870 Funding acquisition, Leadership and supervision, Metadata curation, Project administration,
871 Samples and logistics, Sequencing and analysis, and Software and analysis tools:

872 Dr Thomas R Connor PhD ^{11, 74} and Prof Nicholas J Loman PhD ⁴³

873

874 Leadership and supervision, Metadata curation, Project administration, Samples and
875 logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:

876 Dr Tanya Golubchik PhD ⁵

877

878 Funding acquisition, Leadership and supervision, Metadata curation, Samples and logistics,
879 Sequencing and analysis, and Visualisation:

880 Dr Rocio T Martinez Nunez PhD ⁴⁶

881

882 Funding acquisition, Leadership and supervision, Project administration, Samples and
883 logistics, Sequencing and analysis, and Software and analysis tools:

884 Dr David Bonsall PhD ⁵

885

886 Funding acquisition, Leadership and supervision, Project administration, Sequencing and
887 analysis, Software and analysis tools, and Visualisation:

888 Prof Andrew Rambaut DPhil ¹⁰⁴

889

890 Funding acquisition, Metadata curation, Project administration, Samples and logistics,
891 Sequencing and analysis, and Software and analysis tools:

892 Dr Luke B Snell MSc, MBBS ¹²

893

894 Leadership and supervision, Metadata curation, Project administration, Samples and
895 logistics, Software and analysis tools, and Visualisation:

896 Rich Livett MSc ¹¹⁶

897

898 Funding acquisition, Leadership and supervision, Metadata curation, Project administration,
899 and Samples and logistics:

900 Dr Catherine Ludden PhD ^{20, 70}

901

902 Funding acquisition, Leadership and supervision, Metadata curation, Samples and logistics,
903 and Sequencing and analysis:

904 Dr Sally Corden PhD ⁷⁴ and Dr Eleni Nastouli FRCPATH ^{96, 95, 30}

905

906 Funding acquisition, Leadership and supervision, Metadata curation, Sequencing and
907 analysis, and Software and analysis tools:

908 Dr Gaia Nebbia PhD, FRCPATH ¹²

909
910 Funding acquisition, Leadership and supervision, Project administration, Samples and
911 logistics, and Sequencing and analysis:
912 Ian Johnston BSc ¹¹⁶
913
914 Leadership and supervision, Metadata curation, Project administration, Samples and
915 logistics, and Sequencing and analysis:
916 Prof Katrina Lythgoe PhD ⁵, Dr M. Estee Torok FRCP ^{19, 20} and Prof Ian G Goodfellow PhD ²⁴
917 Leadership and supervision, Metadata curation, Project administration, Samples and
918 logistics, and Visualisation:
919 Dr Jacqui A Prieto PhD ^{97, 82} and Dr Kordo Saeed MD, FRCPATH ^{97, 83}
920
921 Leadership and supervision, Metadata curation, Project administration, Sequencing and
922 analysis, and Software and analysis tools:
923 Dr David K Jackson PhD ¹¹⁶
924
925 Leadership and supervision, Metadata curation, Samples and logistics, Sequencing and
926 analysis, and Visualisation:
927 Dr Catherine Houlihan PhD ^{96, 94}
928
929 Leadership and supervision, Metadata curation, Sequencing and analysis, Software and
930 analysis tools, and Visualisation:
931 Dr Dan Frampton PhD ^{94, 95}
932
933 Metadata curation, Project administration, Samples and logistics, Sequencing and analysis,
934 and Software and analysis tools:
935 Dr William L Hamilton PhD ¹⁹ and Dr Adam A Witney PhD ⁴¹
936
937 Funding acquisition, Samples and logistics, Sequencing and analysis, and Visualisation:
938 Dr Giselda Bucca PhD ¹⁰¹
939
940 Funding acquisition, Leadership and supervision, Metadata curation, and Project
941 administration:
942 Dr Cassie F Pope PhD^{40, 41}
943
944 Funding acquisition, Leadership and supervision, Metadata curation, and Samples and
945 logistics:
946 Dr Catherine Moore PhD ⁷⁴
947
948 Funding acquisition, Leadership and supervision, Metadata curation, and Sequencing and
949 analysis:
950 Prof Emma C Thomson PhD, FRCP ⁵³
951
952 Funding acquisition, Leadership and supervision, Project administration, and Samples and
953 logistics:
954 Dr Ewan M Harrison PhD ^{116, 102}
955
956 Funding acquisition, Leadership and supervision, Sequencing and analysis, and

957 Visualisation:
958 Prof Colin P Smith PhD ¹⁰¹
959
960 Leadership and supervision, Metadata curation, Project administration, and Sequencing and
961 analysis:
962 Fiona Rogan BSc ⁷⁷
963
964 Leadership and supervision, Metadata curation, Project administration, and Samples and
965 logistics:
966 Shaun M Beckwith MSc ⁶, Abigail Murray Degree ⁶, Dawn Singleton HNC ⁶, Dr Kirstine
967 Eastick PhD, FRCPath ³⁷, Dr Liz A Sheridan PhD ⁹⁸, Paul Randell MSc, PgD ⁹⁹, Dr Leigh M
968 Jackson PhD ¹⁰⁵, Dr Cristina V Ariani PhD ¹¹⁶ and Dr Sónia Gonçalves PhD ¹¹⁶
969
970 Leadership and supervision, Metadata curation, Samples and logistics, and Sequencing and
971 analysis:
972 Dr Derek J Fairley PhD ^{3, 77}, Prof Matthew W Loose PhD ¹⁸ and Joanne Watkins MSc ⁷⁴
973
974
975 Leadership and supervision, Metadata curation, Samples and logistics, and Visualisation:
976 Dr Samuel Moses MD ^{25, 106}
977
978 Leadership and supervision, Metadata curation, Sequencing and analysis, and Software and
979 analysis tools:
980 Dr Sam Nicholls PhD ⁴³, Dr Matthew Bull PhD ⁷⁴ and Dr Roberto Amato PhD ¹¹⁶
981
982 Leadership and supervision, Project administration, Samples and logistics, and Sequencing
983 and analysis:
984 Prof Darren L Smith PhD ^{36, 65, 66}
985
986 Leadership and supervision, Sequencing and analysis, Software and analysis tools, and
987 Visualisation:
988 Prof David M Aanensen PhD ^{14, 116} and Dr Jeffrey C Barrett PhD ¹¹⁶
989
990 Metadata curation, Project administration, Samples and logistics, and Sequencing and
991 analysis:
992 Dr Dinesh Aggarwal MRCP^{20, 116, 70}, Dr James G Shepherd MBCHB, MRCP ⁵³, Dr Martin D
993 Curran PhD ⁷¹ and Dr Surendra Parmar PhD ⁷¹
994
995 Metadata curation, Project administration, Sequencing and analysis, and Software and
996 analysis tools:
997 Dr Matthew D Parker PhD ¹⁰⁹
998
999 Metadata curation, Samples and logistics, Sequencing and analysis, and Software and
1000 analysis tools:
1001 Dr Catryn Williams PhD ⁷⁴
1002
1003 Metadata curation, Samples and logistics, Sequencing and analysis, and Visualisation:
1004 Dr Sharon Glaysher PhD ⁶⁸

- 1005
- 1006 Metadata curation, Sequencing and analysis, Software and analysis tools, and Visualisation:
- 1007 Dr Anthony P Underwood PhD ^{14, 116}, Dr Matthew Bashton PhD ^{36, 65}, Dr Nicole Pacchiarini
- 1008 PhD ⁷⁴, Dr Katie F Loveson PhD ⁸⁴ and Matthew Byott MSc ^{95, 96}
- 1009
- 1010 Project administration, Sequencing and analysis, Software and analysis tools, and
- 1011 Visualisation:
- 1012 Dr Alessandro M Carabelli PhD ²⁰
- 1013
- 1014 Funding acquisition, Leadership and supervision, and Metadata curation:
- 1015 Dr Kate E Templeton PhD ^{56, 104}
- 1016
- 1017 Funding acquisition, Leadership and supervision, and Project administration:
- 1018 Dr Thushan I de Silva PhD ¹⁰⁹, Dr Dennis Wang PhD ¹⁰⁹, Dr Cordelia F Langford PhD ¹¹⁶ and
- 1019 John Sillitoe BEng ¹¹⁶
- 1020
- 1021 Funding acquisition, Leadership and supervision, and Samples and logistics:
- 1022 Prof Rory N Gunson PhD, FRCPATH ⁵⁵
- 1023
- 1024 Funding acquisition, Leadership and supervision, and Sequencing and analysis:
- 1025 Dr Simon Cottrell PhD ⁷⁴, Dr Justin O'Grady PhD ^{75, 103} and Prof Dominic Kwiatkowski PhD
- 1026 ^{116, 108}
- 1027
- 1028 Leadership and supervision, Metadata curation, and Project administration:
- 1029 Dr Patrick J Lillie PhD, FRCP ³⁷
- 1030
- 1031 Leadership and supervision, Metadata curation, and Samples and logistics:
- 1032 Dr Nicholas Cortes MBChB ³³, Dr Nathan Moore MBChB ³³, Dr Claire Thomas DPhil ³³,
- 1033 Phillipa J Burns MSc, DipRCPATH ³⁷, Dr Tabitha W Mahungu FRCPATH ⁸⁰ and Steven Liggett
- 1034 BSc ⁸⁶
- 1035
- 1036 Leadership and supervision, Metadata curation, and Sequencing and analysis:
- 1037 Angela H Beckett MSc ^{13, 81} and Prof Matthew TG Holden PhD ⁷³
- 1038
- 1039 Leadership and supervision, Project administration, and Samples and logistics:
- 1040 Dr Lisa J Levett PhD ³⁴, Dr Husam Osman PhD ^{70, 35} and Dr Mohammed O Hassan-Ibrahim
- 1041 PhD, FRCPATH ⁹⁹
- 1042
- 1043 Leadership and supervision, Project administration, and Sequencing and analysis:
- 1044 Dr David A Simpson PhD ⁷⁷
- 1045
- 1046 Leadership and supervision, Samples and logistics, and Sequencing and analysis:
- 1047 Dr Meera Chand PhD ⁷², Prof Ravi K Gupta PhD ¹⁰², Prof Alistair C Darby PhD ¹⁰⁷ and Prof
- 1048 Steve Paterson PhD ¹⁰⁷
- 1049
- 1050 Leadership and supervision, Sequencing and analysis, and Software and analysis tools:
- 1051 Prof Oliver G Pybus DPhil ²³, Dr Erik M Volz PhD ³⁹, Prof Daniela de Angelis PhD ⁵², Prof
- 1052 David L Robertson PhD ⁵³, Dr Andrew J Page PhD ⁷⁵ and Dr Inigo Martincorena PhD ¹¹⁶

- 1053
- 1054 Leadership and supervision, Sequencing and analysis, and Visualisation:
1055 Dr Louise Aigrain PhD ¹¹⁶ and Dr Andrew R Bassett PhD ¹¹⁶
1056
- 1057 Metadata curation, Project administration, and Samples and logistics:
1058 Dr Nick Wong DPhil, MRCP, FRCPath ⁵⁰, Dr Yusri Taha MD, PhD ⁸⁹, Michelle J Erkiert BA ⁹⁹
1059 and Dr Michael H Spencer Chapman MBBS ^{116, 102}
1060
- 1061 Metadata curation, Project administration, and Sequencing and analysis:
1062 Dr Rebecca Dewar PhD ⁵⁶ and Martin P McHugh MSc ^{56, 111}
1063
- 1064 Metadata curation, Project administration, and Software and analysis tools:
1065 Siddharth Mookerjee MPH ^{38, 57}
1066
- 1067 Metadata curation, Project administration, and Visualisation:
1068 Stephen Aplin ⁹⁷, Matthew Harvey ⁹⁷, Thea Sass ⁹⁷, Dr Helen Umpleby FRCP ⁹⁷ and Helen
1069 Wheeler ⁹⁷
1070
- 1071 Metadata curation, Samples and logistics, and Sequencing and analysis:
1072 Dr James P McKenna PhD ³, Dr Ben Warne MRCP ⁹, Joshua F Taylor MSc ²², Yasmin
1073 Chaudhry BSc ²⁴, Rhys Izuagbe ²⁴, Dr Aminu S Jahun PhD ²⁴, Dr Gregory R Young PhD ^{36,}
1074 ⁶⁵, Dr Claire McMurray PhD ⁴³, Dr Clare M McCann PhD ^{65, 66}, Dr Andrew Nelson PhD ^{65, 66}
1075 and Scott Elliott ⁶⁸
1076
- 1077 Metadata curation, Samples and logistics, and Visualisation:
1078 Hannah Lowe MSc ²⁵
1079
- 1080 Metadata curation, Sequencing and analysis, and Software and analysis tools:
1081 Dr Anna Price PhD ¹¹, Matthew R Crown BSc ⁶⁵, Dr Sara Rey PhD ⁷⁴, Dr Sunando Roy PhD
1082 ⁹⁶ and Dr Ben Temperton PhD ¹⁰⁵
1083
- 1084
- 1085 Metadata curation, Sequencing and analysis, and Visualisation:
1086 Dr Sharif Shaaban PhD ⁷³ and Dr Andrew R Hesketh PhD ¹⁰¹
1087
- 1088 Project administration, Samples and logistics, and Sequencing and analysis:
1089 Dr Kenneth G Laing PhD⁴¹, Dr Irene M Monahan PhD ⁴¹ and Dr Judith Heaney PhD ^{95, 96, 34}
1090
- 1091 Project administration, Samples and logistics, and Visualisation:
1092 Dr Emanuela Pelosi FRCPath ⁹⁷, Siona Silveira MSc ⁹⁷ and Dr Eleri Wilson-Davies MD,
1093 FRCPath ⁹⁷
1094
- 1095 Samples and logistics, Software and analysis tools, and Visualisation:
1096 Dr Helen Fryer PhD ⁵
1097
- 1098 Sequencing and analysis, Software and analysis tools, and Visualization:
1099 Dr Helen Adams PhD ⁴, Dr Louis du Plessis PhD ²³, Dr Rob Johnson PhD ³⁹, Dr William T
1100 Harvey PhD ^{53, 42}, Dr Joseph Hughes PhD ⁵³, Dr Richard J Orton PhD ⁵³, Dr Lewis G Spurgin

1101 PhD ⁵⁹, Dr Yann Bourgeois PhD ⁸¹, Dr Chris Ruis PhD ¹⁰², Áine O'Toole MSc ¹⁰⁴, Marina
1102 Gourtovaia MSc ¹¹⁶ and Dr Theo Sanderson PhD ¹¹⁶
1103
1104 Funding acquisition, and Leadership and supervision:
1105 Dr Christophe Fraser PhD ⁵, Dr Jonathan Edgeworth PhD, FRCPath ¹², Prof Judith Breuer
1106 MD ^{96, 29}, Dr Stephen L Mitchell PhD ¹⁰⁵ and Prof John A Todd PhD ¹¹⁵
1107
1108 Funding acquisition, and Project administration:
1109 Michaela John BSc ¹⁰ and Dr David Buck PhD ¹¹⁵
1110
1111 Leadership and supervision, and Metadata curation:
1112 Dr Kavitha Gajee MBBS, FRCPath ³⁷ and Dr Gemma L Kay PhD ⁷⁵
1113
1114 Leadership and supervision, and Project administration:
1115 Prof Sharon J Peacock PhD ^{20, 70} and David Heyburn ⁷⁴
1116
1117 Leadership and supervision, and Samples and logistics:
1118 Katie Kitchman BSc ³⁷, Prof Alan McNally PhD ^{43, 93}, David T Pritchard MSc, CSci ⁵⁰, Dr
1119 Samir Dervisevic FRCPath ⁵⁸, Dr Peter Muir PhD ⁷⁰, Dr Esther Robinson PhD ^{70, 35}, Dr Barry
1120 B Vipond PhD ⁷⁰, Newara A Ramadan MSc, CSci, FIBMS ⁷⁸, Dr Christopher Jeanes MBBS
1121 ⁹⁰, Danni Weldon BSc ¹¹⁶, Jana Catalan MSc ¹¹⁸ and Neil Jones MSc ¹¹⁸
1122
1123 Leadership and supervision, and Sequencing and analysis:
1124 Dr Ana da Silva Filipe PhD ⁵³, Dr Chris Williams MBBS ⁷⁴, Marc Fuchs BSc ⁷⁷, Dr Julia
1125 Miskelly PhD ⁷⁷, Dr Aaron R Jeffries PhD ¹⁰⁵, Karen Oliver BSc ¹¹⁶ and Dr Naomi R Park PhD
1126 ¹¹⁶
1127
1128 Metadata curation, and Samples and logistics:
1129 Amy Ash BSc ¹, Cherian Koshy MSc, CSci, FIBMS ¹, Magdalena Barrow ⁷, Dr Sarah L
1130 Buchan PhD ⁷, Dr Anna Mantzouratou PhD ⁷, Dr Gemma Clark PhD ¹⁵, Dr Christopher W
1131 Holmes PhD ¹⁶, Sharon Campbell MSc ¹⁷, Thomas Davis MSc ²¹, Ngee Keong Tan MSc ²²,
1132 Dr Julianne R Brown PhD ²⁹, Dr Kathryn A Harris PhD ^{29, 2}, Stephen P Kidd MSc ³³, Dr Paul
1133 R Grant PhD ³⁴, Dr Li Xu-McCrae PhD ³⁵, Dr Alison Cox PhD ^{38, 63}, Pinglawathee Madona
1134 ^{38, 63}, Dr Marcus Pond PhD ^{38, 63}, Dr Paul A Randell MBCh ^{38, 63}, Karen T Withell FIBMS ⁴⁸,
1135 Cheryl Williams MSc ⁵¹, Dr Clive Graham MD ⁶⁰, Rebecca Denton-Smith BSc ⁶², Emma
1136 Swindells BSc ⁶², Robyn Turnbull BSc ⁶², Dr Tim J Sloan PhD ⁶⁷, Dr Andrew Bosworth PhD
1137 ^{70, 35}, Stephanie Hutchings ⁷⁰, Hannah M Pymont MSc ⁷⁰, Dr Anna Casey PhD ⁷⁶, Dr Liz
1138 Ratcliffe PhD ⁷⁶, Dr Christopher R Jones PhD ^{79, 105}, Dr Bridget A Knight PhD ^{79, 105}, Dr
1139 Tanzina Haque PhD, FRCPath ⁸⁰, Dr Jennifer Hart MRCP ⁸⁰, Dr Dianne Irish-Tavares
1140 FRCPath ⁸⁰, Eric Witele MSc ⁸⁰, Craig Mower BA ⁸⁶, Louisa K Watson DipHE ⁸⁶, Jennifer
1141 Collins BSc ⁸⁹, Gary Eltringham BSc ⁸⁹, Dorian Crudginton ⁹⁸, Ben Macklin ⁹⁸, Prof Miren
1142 Iturriza-Gomara PhD ¹⁰⁷, Dr Anita O Lucaci PhD ¹⁰⁷ and Dr Patrick C McClure PhD ¹¹³
1143
1144 Metadata curation, and Sequencing and analysis:
1145 Matthew Carlile BSc ¹⁸, Dr Nadine Holmes PhD ¹⁸, Dr Christopher Moore PhD ¹⁸, Dr
1146 Nathaniel Storey PhD ²⁹, Dr Stefan Rooke PhD ⁷³, Dr Gonzalo Yebra PhD ⁷³, Dr Noel Craine
1147 DPhil ⁷⁴, Malorie Perry MSc ⁷⁴, Dr Nabil-Fareed Alikhan PhD ⁷⁵, Dr Stephen Bridgett PhD ⁷⁷,
1148 Kate F Cook MScR ⁸⁴, Christopher Fearn MSc ⁸⁴, Dr Salman Goudarzi PhD ⁸⁴, Prof Ronan A

- 1149 Lyons MD ⁸⁸, Dr Thomas Williams MD ¹⁰⁴, Dr Sam T Haldenby PhD ¹⁰⁷, Jillian Durham BSc
1150 ¹¹⁶ and Dr Steven Leonard PhD ¹¹⁶
1151
1152 Metadata curation, and Software and analysis tools:
1153 Robert M Davies MA (Cantab) ¹¹⁶
1154
1155 Project administration, and Samples and logistics:
1156 Dr Rahul Batra MD ¹², Beth Blane BSc ²⁰, Dr Moira J Spyer PhD ^{30, 95, 96}, Perminder Smith
1157 MSc ^{32, 112}, Mehmet Yavus ^{85, 109}, Dr Rachel J Williams PhD ⁹⁶, Dr Adhyana IK Mahanama
1158 MD ⁹⁷, Dr Buddhini Samaraweera MD ⁹⁷, Sophia T Girgis MSc ¹⁰², Samantha E Hansford
1159 CSci ¹⁰⁹, Dr Angie Green PhD ¹¹⁵, Dr Charlotte Beaver PhD ¹¹⁶, Katherine L Bellis ^{116, 102},
1160 Matthew J Dorman ¹¹⁶, Sally Kay ¹¹⁶, Liam Prestwood ¹¹⁶ and Dr Shavanthi Rajatileka PhD
1161 ¹¹⁶
1162
1163 Project administration, and Sequencing and analysis:
1164 Dr Joshua Quick PhD ⁴³
1165
1166 Project administration, and Software and analysis tools:
1167 Radoslaw Poplawski BSc ⁴³
1168
1169 Samples and logistics, and Sequencing and analysis:
1170 Dr Nicola Reynolds PhD ⁸, Andrew Mack MPhil ¹¹, Dr Arthur Morriss PhD ¹¹, Thomas
1171 Whalley BSc ¹¹, Bindi Patel BSc ¹², Dr Iliana Georgana PhD ²⁴, Dr Myra Hosmillo PhD ²⁴,
1172 Malte L Pinckert MPhil ²⁴, Dr Joanne Stockton PhD ⁴³, Dr John H Henderson PhD ⁶⁵, Amy
1173 Hollis HND ⁶⁵, Dr William Stanley PhD ⁶⁵, Dr Wen C Yew PhD ⁶⁵, Dr Richard Myers PhD ⁷²,
1174 Dr Alicia Thornton PhD ⁷², Alexander Adams BSc ⁷⁴, Tara Annett BSc ⁷⁴, Dr Hibo Asad PhD
1175 ⁷⁴, Alec Birchley MSc ⁷⁴, Jason Coombes BSc ⁷⁴, Johnathan M Evans MSc ⁷⁴, Laia Fina ⁷⁴,
1176 Bree Gatica-Wilcox MPhil ⁷⁴, Lauren Gilbert ⁷⁴, Lee Graham BSc ⁷⁴, Jessica Hey BSc ⁷⁴,
1177 Ember Hilvers MPH ⁷⁴, Sophie Jones MSc ⁷⁴, Hannah Jones ⁷⁴, Sara Kumziene-
1178 Summerhayes MSc ⁷⁴, Dr Caoimhe McKerr PhD ⁷⁴, Jessica Powell BSc ⁷⁴, Georgia Pugh ⁷⁴,
1179 Sarah Taylor ⁷⁴, Alexander J Trotter MRes ⁷⁵, Charlotte A Williams BSc ⁹⁶, Leanne M
1180 Kermack MSc ¹⁰², Benjamin H Foulkes MSc ¹⁰⁹, Marta Gallis MSc ¹⁰⁹, Hailey R Hornsby MSc
1181 ¹⁰⁹, Stavroula F Louka MSc ¹⁰⁹, Dr Manoj Pohare PhD ¹⁰⁹, Paige Wolverson MSc ¹⁰⁹, Peijun
1182 Zhang MSc ¹⁰⁹, George MacIntyre-Cockett BSc ¹¹⁵, Amy Trebes MSc ¹¹⁵, Dr Robin J Moll
1183 PhD ¹¹⁶, Lynne Ferguson MSc ¹¹⁷, Dr Emily J Goldstein PhD ¹¹⁷, Dr Alasdair Maclean PhD ¹¹⁷
1184 and Dr Rachael Tomb PhD ¹¹⁷
1185
1186 Samples and logistics, and Software and analysis tools:
1187 Dr Igor Starinskij MSc, MRCP ⁵³
1188
1189 Sequencing and analysis, and Software and analysis tools:
1190 Laura Thomson BSc ⁵, Joel Southgate MSc ^{11, 74}, Dr Moritz UG Kraemer DPhil ²³, Dr Jayna
1191 Raghvani PhD ²³, Dr Alex E Zarebski PhD ²³, Olivia Boyd MSc ³⁹, Lily Geidelberg MSc ³⁹, Dr
1192 Chris J Illingworth PhD ⁵², Dr Chris Jackson PhD ⁵², Dr David Pascall PhD ⁵², Dr Sreenu
1193 Vattipally PhD ⁵³, Timothy M Freeman MPhil ¹⁰⁹, Dr Sharon N Hsu PhD ¹⁰⁹, Dr Benjamin B
1194 Lindsey MRCP ¹⁰⁹, Dr Keith James PhD ¹¹⁶, Kevin Lewis ¹¹⁶, Gerry Tonkin-Hill ¹¹⁶ and Dr
1195 Jaime M Tovar-Corona PhD ¹¹⁶
1196

1197 Sequencing and analysis, and Visualisation:

1198 MacGregor Cox MSci ²⁰

1199

1200 Software and analysis tools, and Visualisation:

1201 Dr Khalil Abudahab PhD ^{14, 116}, Mirko Menegazzo ¹⁴, Ben EW Taylor MEng ^{14, 116}, Dr Corin A

1202 Yeats PhD ¹⁴, Afrida Mukaddas BTech ⁵³, Derek W Wright MSc ⁵³, Dr Leonardo de Oliveira

1203 Martins PhD ⁷⁵, Dr Rachel Colquhoun DPhil ¹⁰⁴, Verity Hill ¹⁰⁴, Dr Ben Jackson PhD ¹⁰⁴, Dr JT

1204 McCrone PhD ¹⁰⁴, Dr Nathan Medd PhD ¹⁰⁴, Dr Emily Scher PhD ¹⁰⁴ and Jon-Paul Keatley ¹¹⁶

1205

1206 Leadership and supervision:

1207 Dr Tanya Curran PhD ³, Dr Sian Morgan FRCPATH ¹⁰, Prof Patrick Maxwell PhD ²⁰, Prof Ken

1208 Smith PhD ²⁰, Dr Sahar Eldirdiri MBBS, MSc, FRCPATH ²¹, Anita Kenyon MSc ²¹, Prof Alison

1209 H Holmes MD ^{38, 57}, Dr James R Price PhD ^{38, 57}, Dr Tim Wyatt PhD ⁶⁹, Dr Alison E Mather

1210 PhD ⁷⁵, Dr Timofey Skvortsov PhD ⁷⁷ and Prof John A Hartley PhD ⁹⁶

1211

1212 Metadata curation:

1213 Prof Martyn Guest PhD ¹¹, Dr Christine Kitchen PhD ¹¹, Dr Ian Merrick PhD ¹¹, Robert Munn

1214 BSc ¹¹, Dr Beatrice Bertolusso Degree ³³, Dr Jessica Lynch MBCHB ³³, Dr Gabrielle Vernet

1215 MBBS ³³, Stuart Kirk MSc ³⁴, Dr Elizabeth Wastnedge MD ⁵⁶, Dr Rachael Stanley PhD ⁵⁸,

1216 Giles Idle ⁶⁴, Dr Declan T Bradley PhD ^{69, 77}, Dr Jennifer Poyner MD ⁷⁹ and Matilde Mori BSc

1217 ¹¹⁰

1218

1219 Project administration:

1220 Owen Jones BSc ¹¹, Victoria Wright BSc ¹⁸, Ellena Brooks MA ²⁰, Carol M Churcher BSc ²⁰,

1221 Mireille Fragakis HND ²⁰, Dr Katerina Galai PhD ^{20, 70}, Dr Andrew Jermy PhD ²⁰, Sarah

1222 Judges BA ²⁰, Georgina M McManus BSc ²⁰, Kim S Smith ²⁰, Dr Elaine Westwick PhD ²⁰, Dr

1223 Stephen W Attwood PhD ²³, Dr Frances Bolt PhD ^{38, 57}, Dr Alisha Davies PhD ⁷⁴, Elen De

1224 Lacy MPH ⁷⁴, Fatima Downing ⁷⁴, Sue Edwards ⁷⁴, Lizzie Meadows MA ⁷⁵, Sarah Jeremiah

1225 MSc ⁹⁷, Dr Nikki Smith PhD ¹⁰⁹ and Luke Foulser ¹¹⁶

1226

1227 Samples and logistics:

1228 Dr Themoula Charalampous PhD ^{12, 46}, Amita Patel BSc ¹², Dr Louise Berry PhD ¹⁵, Dr Tim

1229 Boswell PhD ¹⁵, Dr Vicki M Fleming PhD ¹⁵, Dr Hannah C Howson-Wells PhD ¹⁵, Dr Amelia

1230 Joseph PhD ¹⁵, Manjinder Khakh ¹⁵, Dr Michelle M Lister PhD ¹⁵, Paul W Bird MSc, MRes ¹⁶,

1231 Karlie Fallon ¹⁶, Thomas Helmer ¹⁶, Dr Claire L McMurray PhD ¹⁶, Mina Odedra BSc ¹⁶,

1232 Jessica Shaw BSc ¹⁶, Dr Julian W Tang PhD ¹⁶, Nicholas J Willford MSc ¹⁶, Victoria Blakey

1233 BSc ¹⁷, Dr Veena Raviprakash MD ¹⁷, Nicola Sheriff BSc ¹⁷, Lesley-Anne Williams BSc ¹⁷,

1234 Theresa Feltwell MSc ²⁰, Dr Luke Bedford PhD ²⁶, Dr James S Cargill PhD ²⁷, Warwick

1235 Hughes MSc ²⁷, Dr Jonathan Moore MD ²⁸, Susanne Stonehouse BSc ²⁸, Laura Atkinson

1236 MSc ²⁹, Jack CD Lee MSc ²⁹, Dr Divya Shah PhD ²⁹, Adela Alcolea-Medina Clinical scientist

1237 ^{32, 112}, Natasha Ohemeng-Kumi MSc ^{32, 112}, John Ramble MSc ^{32, 112}, Jasveen Sehmi MSc ³²,

1238 ¹¹², Dr Rebecca Williams BMBS ³³, Wendy Chatterton MSc ³⁴, Monika Pusok MSc ³⁴, William

1239 Everson MSc ³⁷, Anibolina Castigador IBMS HCPC ⁴⁴, Emily Macnaughton FRCPATH ⁴⁴, Dr

1240 Kate El Bouzidi MRCP ⁴⁵, Dr Temi Lampejo FRCPATH ⁴⁵, Dr Malur Sudhanva FRCPATH ⁴⁵,

1241 Cassie Breen BSc ⁴⁷, Dr Graciela Sluga MD, MSc ⁴⁸, Dr Shazaad SY Ahmad MSc ^{49, 70}, Dr

1242 Ryan P George PhD ⁴⁹, Dr Nicholas W Machin MSc ^{49, 70}, Debbie Binns BSc ⁵⁰, Victoria

1243 James BSc ⁵⁰, Dr Rachel Blacow MBCHB ⁵⁵, Dr Lindsay Coupland PhD ⁵⁸, Dr Louise Smith

1244 PhD ⁵⁹, Dr Edward Barton MD ⁶⁰, Debra Padgett BSc ⁶⁰, Garren Scott BSc ⁶⁰, Dr Aidan Cross

1245 MBCHB ⁶¹, Dr Mariyam Mirfenderesky FRCPATH ⁶¹, Jane Greenaway MSc ⁶², Kevin Cole ⁶⁴,
1246 Phillip Clarke ⁶⁷, Nichola Duckworth ⁶⁷, Sarah Walsh ⁶⁷, Kelly Bicknell ⁶⁸, Robert Impey MSc
1247 ⁶⁸, Dr Sarah Wyllie PhD ⁶⁸, Richard Hopes ⁷⁰, Dr Chloe Bishop PhD ⁷², Dr Vicki Chalker PhD
1248 ⁷², Dr Ian Harrison PhD ⁷², Laura Gifford MSc ⁷⁴, Dr Zoltan Molnar PhD ⁷⁷, Dr Cressida
1249 Auckland FRCPATH ⁷⁹, Dr Cariad Evans PhD ^{85, 109}, Dr Kate Johnson PhD ^{85, 109}, Dr David G
1250 Partridge FRCP, FRCPATH ^{85, 109}, Dr Mohammad Raza PhD ^{85, 109}, Paul Baker MD ⁸⁶, Prof
1251 Stephen Bonner PhD ⁸⁶, Sarah Essex ⁸⁶, Leanne J Murray ⁸⁶, Andrew I Lawton MSc ⁸⁷, Dr
1252 Shirelle Burton-Fanning MD ⁸⁹, Dr Brendan Al Payne MD ⁸⁹, Dr Sheila Waugh MD ⁸⁹, Andrea
1253 N Gomes MSc ⁹¹, Maimuna Kimuli MSc ⁹¹, Darren R Murray MSc ⁹¹, Paula Ashfield MSc ⁹²,
1254 Dr Donald Dobie MBCHB ⁹², Dr Fiona Ashford PhD ⁹³, Dr Angus Best PhD ⁹³, Dr Liam
1255 Crawford PhD ⁹³, Dr Nicola Cumley PhD ⁹³, Dr Megan Mayhew PhD ⁹³, Dr Oliver Megram
1256 PhD ⁹³, Dr Jeremy Mirza PhD ⁹³, Dr Emma Moles-Garcia PhD ⁹³, Dr Benita Percival PhD ⁹³,
1257 Megan Driscoll BSc ⁹⁶, Leah Ensell BSc ⁹⁶, Dr Helen L Lowe PhD ⁹⁶, Laurentiu Maftei BSc ⁹⁶,
1258 Matteo Mondani MSc ⁹⁶, Nicola J Chaloner BSc ⁹⁹, Benjamin J Cogger BSc ⁹⁹, Lisa J Easton
1259 MSc ⁹⁹, Hannah Huckson BSc ⁹⁹, Jonathan Lewis MSc, PgD, FIBMS ⁹⁹, Sarah Lowdon BSc
1260 ⁹⁹, Cassandra S Malone MSc ⁹⁹, Florence Munemo BSc ⁹⁹, Manasa Mutingwende MSc ⁹⁹,
1261 Roberto Nicodemi BSc ⁹⁹, Olga Podplomyk FD ⁹⁹, Thomas Somassa BSc ⁹⁹, Dr Andrew
1262 Beggs PhD ¹⁰⁰, Dr Alex Richter PhD ¹⁰⁰, Claire Cormie ¹⁰², Joana Dias MSc ¹⁰², Sally Forrest
1263 BSc ¹⁰², Dr Ellen E Higginson PhD ¹⁰², Mailis Maes MPhil ¹⁰², Jamie Young BSc ¹⁰², Dr Rose
1264 K Davidson PhD ¹⁰³, Kathryn A Jackson MSc ¹⁰⁷, Dr Lance Turtle PhD, MRCP ¹⁰⁷, Dr
1265 Alexander J Keeley MRCP ¹⁰⁹, Prof Jonathan Ball PhD ¹¹³, Timothy Byaruhanga MSc ¹¹³, Dr
1266 Joseph G Chappell PhD ¹¹³, Jayasree Dey MSc ¹¹³, Jack D Hill MSc ¹¹³, Emily J Park MSc
1267 ¹¹³, Arezou Fanaie MSc ¹¹⁴, Rachel A Hilson MSc ¹¹⁴, Geraldine Yaze MSc ¹¹⁴ and Stephanie
1268 Lo ¹¹⁶
1269
1270 Sequencing and analysis:
1271 Safiah Afifi BSc ¹⁰, Robert Beer BSc ¹⁰, Joshua Maksimovic FD ¹⁰, Kathryn McCluggage
1272 Masters ¹⁰, Karla Spellman FD ¹⁰, Catherine Bresner BSc ¹¹, William Fuller BSc ¹¹, Dr Angela
1273 Marchbank BSc ¹¹, Trudy Workman HNC ¹¹, Dr Ekaterina Shelest PhD ^{13, 81}, Dr Johnny
1274 Debebe PhD ¹⁸, Dr Fei Sang PhD ¹⁸, Dr Marina Escalera Zamudio PhD ²³, Dr Sarah Francois
1275 PhD ²³, Bernardo Gutierrez MSc ²³, Dr Tetyana I Vasylyeva DPhil ²³, Dr Flavia Flaviani PhD
1276 ³¹, Dr Manon Ragonnet-Cronin PhD ³⁹, Dr Katherine L Smollett PhD ⁴², Alice Broos BSc ⁵³,
1277 Daniel Mair BSc ⁵³, Jenna Nichols BSc ⁵³, Dr Kyriaki Nomikou PhD ⁵³, Dr Lily Tong PhD ⁵³,
1278 Ioulia Tsatsani MSc ⁵³, Prof Sarah O'Brien PhD ⁵⁴, Prof Steven Rushton PhD ⁵⁴, Dr Roy
1279 Sanderson PhD ⁵⁴, Dr Jon Perkins MBCHB ⁵⁵, Seb Cotton MSc ⁵⁶, Abbie Gallagher BSc ⁵⁶,
1280 Dr Elias Allara MD, PhD ^{70, 102}, Clare Pearson MSc ^{70, 102}, Dr David Bibby PhD ⁷², Dr Gavin
1281 Dabrera PhD ⁷², Dr Nicholas Ellaby PhD ⁷², Dr Eileen Gallagher PhD ⁷², Dr Jonathan Hubb
1282 PhD ⁷², Dr Angie Lackenby PhD ⁷², Dr David Lee PhD ⁷², Nikos Manesis ⁷², Dr Tamyo Mbisa
1283 PhD ⁷², Dr Steven Platt PhD ⁷², Katherine A Twohig ⁷², Dr Mari Morgan PhD ⁷⁴, Alp Aydin
1284 MSci ⁷⁵, David J Baker BEng ⁷⁵, Dr Ebenezer Foster-Nyarko PhD ⁷⁵, Dr Sophie J Prosolek
1285 PhD ⁷⁵, Steven Rudder ⁷⁵, Chris Baxter BSc ⁷⁷, Sílvia F Carvalho MSc ⁷⁷, Dr Deborah Lavin
1286 PhD ⁷⁷, Dr Arun Mariappan PhD ⁷⁷, Dr Clara Radulescu PhD ⁷⁷, Dr Aditi Singh PhD ⁷⁷, Miao
1287 Tang MD ⁷⁷, Helen Morcette BSc ⁷⁹, Nadua Bayzid BSc ⁹⁶, Marius Cotic MSc ⁹⁶, Dr Carlos E
1288 Balcazar PhD ¹⁰⁴, Dr Michael D Gallagher PhD ¹⁰⁴, Dr Daniel Maloney PhD ¹⁰⁴, Thomas D
1289 Stanton BSc ¹⁰⁴, Dr Kathleen A Williamson PhD ¹⁰⁴, Dr Robin Manley PhD ¹⁰⁵, Michelle L
1290 Michelsen BSc ¹⁰⁵, Dr Christine M Sambles PhD ¹⁰⁵, Dr David J Studholme PhD ¹⁰⁵, Joanna
1291 Warwick-Dugdale BSc ¹⁰⁵, Richard Eccles MSc ¹⁰⁷, Matthew Gemmell MSc ¹⁰⁷, Dr Richard
1292 Gregory PhD ¹⁰⁷, Dr Margaret Hughes PhD ¹⁰⁷, Charlotte Nelson MSc ¹⁰⁷, Dr Lucille Rainbow

1293 PhD ¹⁰⁷, Dr Edith E Vamos PhD ¹⁰⁷, Hermione J Webster BSc ¹⁰⁷, Dr Mark Whitehead PhD
1294 ¹⁰⁷, Claudia Wierzbicki BSc ¹⁰⁷, Dr Adrienn Angyal PhD ¹⁰⁹, Dr Luke R Green PhD ¹⁰⁹, Dr Max
1295 Whiteley PhD ¹⁰⁹, Emma Betteridge BSc ¹¹⁶, Dr Iraad F Bronner PhD ¹¹⁶, Ben W Farr BSc ¹¹⁶,
1296 Scott Goodwin MSc ¹¹⁶, Dr Stefanie V Lensing PhD ¹¹⁶, Shane A McCarthy ^{116, 102}, Dr Michael
1297 A Quail PhD ¹¹⁶, Diana Rajan MSc ¹¹⁶, Dr Nicholas M Redshaw PhD ¹¹⁶, Carol Scott ¹¹⁶,
1298 Lesley Shirley MSc ¹¹⁶ and Scott AJ Thurston BSc ¹¹⁶

1299

1300 Software and analysis tools:

1301 Dr Will Rowe PhD⁴³, Amy Gaskin MSc ⁷⁴, Dr Thanh Le-Viet PhD ⁷⁵, James Bonfield BSc ¹¹⁶,
1302 Jennifer Liddle ¹¹⁶ and Andrew Whitwham BSc ¹¹⁶

1303

1304 1 Barking, Havering and Redbridge University Hospitals NHS Trust, 2 Barts Health NHS Trust, 3
1305 Belfast Health & Social Care Trust, 4 Betsi Cadwaladr University Health Board, 5 Big Data Institute,
1306 Nuffield Department of Medicine, University of Oxford, 6 Blackpool Teaching Hospitals NHS
1307 Foundation Trust, 7 Bournemouth University, 8 Cambridge Stem Cell Institute, University of
1308 Cambridge, 9 Cambridge University Hospitals NHS Foundation Trust, 10 Cardiff and Vale University
1309 Health Board, 11 Cardiff University, 12 Centre for Clinical Infection and Diagnostics Research,
1310 Department of Infectious Diseases, Guy's and St Thomas' NHS Foundation Trust, 13 Centre for
1311 Enzyme Innovation, University of Portsmouth, 14 Centre for Genomic Pathogen Surveillance,
1312 University of Oxford, 15 Clinical Microbiology Department, Queens Medical Centre, Nottingham
1313 University Hospitals NHS Trust, 16 Clinical Microbiology, University Hospitals of Leicester NHS Trust,
1314 17 County Durham and Darlington NHS Foundation Trust, 18 Deep Seq, School of Life Sciences,
1315 Queens Medical Centre, University of Nottingham, 19 Department of Infectious Diseases and
1316 Microbiology, Cambridge University Hospitals NHS Foundation Trust, 20 Department of Medicine,
1317 University of Cambridge, 21 Department of Microbiology, Kettering General Hospital, 22 Department
1318 of Microbiology, South West London Pathology, 23 Department of Zoology, University of Oxford, 24
1319 Division of Virology, Department of Pathology, University of Cambridge, 25 East Kent Hospitals
1320 University NHS Foundation Trust, 26 East Suffolk and North Essex NHS Foundation Trust, 27 East
1321 Sussex Healthcare NHS Trust, 28 Gateshead Health NHS Foundation Trust, 29 Great Ormond Street
1322 Hospital for Children NHS Foundation Trust, 30 Great Ormond Street Institute of Child Health (GOS
1323 ICH), University College London (UCL), 31 Guy's and St. Thomas' Biomedical Research Centre, 32
1324 Guy's and St. Thomas' NHS Foundation Trust, 33 Hampshire Hospitals NHS Foundation Trust, 34
1325 Health Services Laboratories, 35 Heartlands Hospital, Birmingham, 36 Hub for Biotechnology in the
1326 Built Environment, Northumbria University, 37 Hull University Teaching Hospitals NHS Trust, 38
1327 Imperial College Healthcare NHS Trust, 39 Imperial College London, 40 Infection Care Group, St
1328 George's University Hospitals NHS Foundation Trust, 41 Institute for Infection and Immunity, St
1329 George's University of London, 42 Institute of Biodiversity, Animal Health & Comparative Medicine, 43
1330 Institute of Microbiology and Infection, University of Birmingham, 44 Isle of Wight NHS Trust, 45
1331 King's College Hospital NHS Foundation Trust, 46 King's College London, 47 Liverpool Clinical
1332 Laboratories, 48 Maidstone and Tunbridge Wells NHS Trust, 49 Manchester University NHS
1333 Foundation Trust, 50 Microbiology Department, Buckinghamshire Healthcare NHS Trust, 51
1334 Microbiology, Royal Oldham Hospital, 52 MRC Biostatistics Unit, University of Cambridge, 53 MRC-
1335 University of Glasgow Centre for Virus Research, 54 Newcastle University, 55 NHS Greater Glasgow
1336 and Clyde, 56 NHS Lothian, 57 NIHR Health Protection Research Unit in HCAI and AMR, Imperial
1337 College London, 58 Norfolk and Norwich University Hospitals NHS Foundation Trust, 59 Norfolk
1338 County Council, 60 North Cumbria Integrated Care NHS Foundation Trust, 61 North Middlesex
1339 University Hospital NHS Trust, 62 North Tees and Hartlepool NHS Foundation Trust, 63 North West
1340 London Pathology, 64 Northumbria Healthcare NHS Foundation Trust, 65 Northumbria University, 66
1341 NU-OMICS, Northumbria University, 67 Path Links, Northern Lincolnshire and Goole NHS Foundation
1342 Trust, 68 Portsmouth Hospitals University NHS Trust, 69 Public Health Agency, Northern Ireland, 70
1343 Public Health England, 71 Public Health England, Cambridge, 72 Public Health England, Colindale,

1344 73 Public Health Scotland, 74 Public Health Wales, 75 Quadram Institute Bioscience, 76 Queen
1345 Elizabeth Hospital, Birmingham, 77 Queen's University Belfast, 78 Royal Brompton and Harefield
1346 Hospitals, 79 Royal Devon and Exeter NHS Foundation Trust, 80 Royal Free London NHS
1347 Foundation Trust, 81 School of Biological Sciences, University of Portsmouth, 82 School of Health
1348 Sciences, University of Southampton, 83 School of Medicine, University of Southampton, 84 School of
1349 Pharmacy & Biomedical Sciences, University of Portsmouth, 85 Sheffield Teaching Hospitals NHS
1350 Foundation Trust, 86 South Tees Hospitals NHS Foundation Trust, 87 Southwest Pathology Services,
1351 88 Swansea University, 89 The Newcastle upon Tyne Hospitals NHS Foundation Trust, 90 The
1352 Queen Elizabeth Hospital King's Lynn NHS Foundation Trust, 91 The Royal Marsden NHS
1353 Foundation Trust, 92 The Royal Wolverhampton NHS Trust, 93 Turnkey Laboratory, University of
1354 Birmingham, 94 University College London Division of Infection and Immunity, 95 University College
1355 London Hospital Advanced Pathogen Diagnostics Unit, 96 University College London Hospitals NHS
1356 Foundation Trust, 97 University Hospital Southampton NHS Foundation Trust, 98 University Hospitals
1357 Dorset NHS Foundation Trust, 99 University Hospitals Sussex NHS Foundation Trust, 100 University
1358 of Birmingham, 101 University of Brighton, 102 University of Cambridge, 103 University of East
1359 Anglia, 104 University of Edinburgh, 105 University of Exeter, 106 University of Kent, 107 University of
1360 Liverpool, 108 University of Oxford, 109 University of Sheffield, 110 University of Southampton, 111
1361 University of St Andrews, 112 Viapath, Guy's and St Thomas' NHS Foundation Trust, and King's
1362 College Hospital NHS Foundation Trust, 113 Virology, School of Life Sciences, Queens Medical
1363 Centre, University of Nottingham, 114 Watford General Hospital, 115 Wellcome Centre for Human
1364 Genetics, Nuffield Department of Medicine, University of Oxford, 116 Wellcome Sanger Institute, 117
1365 West of Scotland Specialist Virology Centre, NHS Greater Glasgow and Clyde, 118 Whittington
1366 Health NHS Trust
1367