

Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/162340>

How to cite:

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

1
2
3
4 **1 SARS-CoV-2 transmissibility compared between**
5
6 **2 variants of concern and vaccination status**
7
8

9 **3 Liang Wang^{1,*}, Xavier Didelot², Yuhai Bi^{1,3,*}, George F Gao^{1,3,*}**

10
11 **4 ¹CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of**
12
13 **5 Microbiology, Center for Influenza Research and Early-warning (CASCIRE), CAS-**
14
15 **6 TWAS Center of Excellence for Emerging Infectious Diseases (CEEID), Chinese**
16
17 **7 Academy of Sciences, Beijing 100101, China**

18
19 **8 ²School of Life Sciences and Department of Statistics, University of Warwick,**
20
21 **9 Coventry CV47AL, United Kingdom**

22
23 **10 ³University of Chinese Academy of Sciences, Beijing 101408, China**

24
25 **11 * Correspondence: wangliang@im.ac.cn (L.W); beeyh@im.ac.cn (Y.B.);**
26
27 **12 gaof@im.ac.cn (G.F.G)**

13 Abstract

14 Since the start of the SARS-CoV-2 pandemic in late 2019, several variants of concern
15 (VOC) have been reported to have increased transmissibility. In addition, despite the
16 progress of vaccination against SARS-CoV-2 worldwide, all vaccines currently in used
17 are known to protect only partially from infection and onward transmission. We
18 combined phylogenetic analysis with Bayesian inference under an epidemiological
19 model to infer the reproduction number (R_t) and also trace person-to-person
20 transmission. We ~~also~~ examined the impact of phylogenetic uncertainty and sampling
21 bias on the estimation. Our result indicated that ~~the~~ lineage B had a significantly higher
22 transmissibility than lineage A, and contributed to the global pandemic to a large extent.
23 In addition, although the transmissibility of VOCs ~~has been increased compared with~~
24 ~~larger than~~ other exponentially growing lineages ~~with exponential growth rate~~, this
25 difference is not very high. The probability of detecting onward transmission from
26 patients infected with SARS-CoV-2 VOCs who had received at least one dose of
27 vaccine was approximate 1.06% (3/284), which was slightly lower but not statistically
28 ~~not~~ significantly different from a probability of 1.21% (10 /828) for unvaccinated
29 individuals. In addition to VOCs, exponentially growing lineages ~~with exponential~~
30 ~~growth rate~~ in each country should also be paid attention account for when tailoring
31 prevention and control strategies. One dose of vaccination could not efficiently prevent
32 the onward transmission of SARS-CoV-2 VOCs. ~~In order to prevent this~~ Consequently,
33 non-pharmaceutical interventions (such as ~~low-cost and efficient strategies, like~~
34 wearing masks and social distancing ~~ete~~) should still be implemented in each country

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

35 during the vaccination period.

36

37 **Keywords**

38 SARS-CoV-2, variants of concern, vaccine, transmissibility, onward transmission

39

For Peer Review

40 Introduction

41 Coronavirus diseases 2019 (COVID-19), the biggest pandemic so far in the 21st century,
42 is caused by a novel type of coronaviruses named SARS-CoV-2 (also known as 2019-
43 nCoV, or HCoV-19)[1]. As of 10th October 2021, there are more than 238 million
44 confirmed cases with more than four million deaths², posing a global threat to public
45 health. During the SARS-CoV-2 pandemic, several types of SARS-CoV-2 variants of
46 concern (VOC) with increased transmissibility emerged, such as B.1.1.7 (WHO label:
47 Alpha), B.1.351 (WHO label: Beta), P.1 (WHO label: Gamma), and B.1.617.2 (WHO
48 label: Delta)[2-5], the global spread of these VOCs has also further thoroughly taxed
49 the medical systems and global economies.

50
51 Although VOCs deserves worldwide attention, those lineages with exponential growth
52 in each country cannot be ignored. Since the advantages of transmissibility for VOCs
53 were mainly concluded by comparing them to all other lineages as a whole[2, 5], it will
54 cause the advantage of transmissibility for some lineages to be overwhelmed. In
55 addition, VOCs have also been reported to be harder to neutralize by convalescent and
56 vaccine sera than others[6-11], indicating they could still infect vaccinated individuals,
57 which therefore could increase the probability of transmission to others. Together with
58 the increased breakthrough infection rates[12], more efforts are needed to identify the
59 transmissibility of lineages with exponential growth other than VOCs in each country
60 and survey the extent of onward transmission caused by vaccinated persons being
61 infected by SARS-CoV-2 VOCs, which is also an indicator for policy makers to tailor

1
2
3
4 62 further prevention and control measures during the vaccination and post-vaccination
5
6 63 process.
7
8

9 64

11 65 **Materials and methods**

14 66 **Data collection and selection**

16
17 67 SARS-COV-2 genomic sequences were download from GISAID several times (data
18
19 68 for estimating lineage A and B was downloaded at 9th April 2020, data for UK was
20
21
22 69 downloaded at 21st December 2020, data for South Africa and Brazil was downloaded
23
24
25 70 at 16th March 2021, data for India was downloaded at 13th May 2021). For estimating
26
27 71 the extent of onward transmission caused by vaccinated persons being infected by
28
29
30 72 SARS-CoV-2 VOCs, genomic sequences and corresponding patients' vaccination
31
32
33 73 status were download from GISAID at 18th June 2021. Totally, we got 408 SARS-CoV-
34
35 74 2 genomic sequences, all of which came from patients who had received at least one
36
37
38 75 dose of vaccine before being infected with SARS-CoV-2 VOCs.
39

40 76

42
43 77 Only viral genomes collected before the implementation of national non-
44
45 78 pharmaceutical interventions would be included in the analysis of R_t estimation for
46
47
48 79 lineage A and B. In addition, countries that include lineage A and B, and the number of
49
50
51 80 completely viral genomes within each lineage ≥ 80 would be included in the subsequent
52
53 81 analysis. Since only the United States and Australia met the above criteria, the
54
55
56 82 estimation of the transmissibility of lineage A and B was only based on the data of these
57
58
59 83 two countries. The cut-off dates for the collection time in the USA and Australia are
60

1
2
3
4 84 20th and 25th January 2020, respectively, as there were no nationwide epidemic
5
6
7 85 prevention measures were implemented before the date. Due to the high volume of
8
9
10 86 genomic data from sub-lineages in the UK, South Africa, Brazil, and India, the amount
11
12 87 of calculation would be too large, especially for reconstruction of dated phylogeny. ~~In~~
13
14 88 ~~this case, we~~ We therefore filtered and ~~also~~ sub-sampled the data for datasets from each
15
16
17 89 sub-lineage. First, the viral genomes of patients who had not had a history of
18
19
20 90 international travel are retained, according to their epidemiological data. Second, the
21
22 91 viral genomes should also meet the criteria as follow: length ≥ 29 KB, and the ratio of
23
24
25 92 N in the genome $\leq 1\%$. Third, based on the collection date, if more than 10 genomes
26
27
28 93 were available in a specific date, we randomly select 10 of them, otherwise all genomes
29
30
31 94 would be included. For identifying onward transmission caused by patients being
32
33 95 infected with VOCs after receiving at least one dose of vaccine, we first filtered the
34
35
36 96 data based on several following criteria. Only complete SARS-CoV-2 genomes from
37
38 97 patients receiving at least one dose of vaccine were retained for further analysis. We
39
40
41 98 then discarded genomic data with no exact collection date (accurate to days). Due to
42
43
44 99 the aim of our study is to identify direct transmission events, we then also collected
45
46
47 100 viral genomic sequences that were highly similar to those SARS-CoV-2 genomes from
48
49
50 101 patients receiving at least one dose of vaccine, as we assumed that SARS-CoV-2
51
52 102 genome sequences from two patients that directly transmitted SARS-CoV-2 to each
53
54
55 103 other were with high sequence similarity. For each SARS-CoV-2 genome from patients
56
57
58 104 receiving at least one dose of vaccine (query), we also used BLAST to find 10 most
59
60 105 similar complete genomes (target) and then retained those with exact collection date

1
2
3
4 106 (accurate to days) which were also from the same country as each query and their
5
6 107 collection times were within 22 days (maximum infectious period)[13] after the
7
8
9 108 collection time of the query. The query and target sequences were then put together and
10
11
12 109 removed redundancy for further analysis. For SARS-CoV-2 Alpha VOC, genomic
13
14 110 sequences were split into different datasets based on the country, and only dataset
15
16
17 111 contained more than 70 SARS-CoV-2 genomes was used for further analysis, as the
18
19 112 computational cost was extremely large if we combined data from all countries. Since
20
21
22 113 there are still several countries with limited genomic sequences, we then merged them
23
24
25 114 into a dataset. Other VOCs were considered as independent dataset and were not further
26
27
28 115 split anymore. Finally, only 284 genomic sequences of SARS-CoV-2 VOCs, all of
29
30
31 116 which came from patients who had received at least one dose of vaccine before
32
33 117 infection, and 828 genomic sequences of SARS-CoV-2 VOCs that close related to the
34
35 118 above sequences but all of which came from patients who did not receive vaccine at all
36
37
38 119 were retained for further analysis. Before further analysis, genomic sequences were
39
40
41 120 aligned using Mafft v7.310[14]. Then, we trimmed the uncertain regions in 3' and 5'
42
43 121 terminals and also masked 30 sites (Supplementary Table 1) that are highly homoplastic
44
45 122 and have no phylogenetic signal as previous noted ([https://virological.org/t/issues-with-](https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473)
46
47
48 123 [sars-cov-2-sequencing-data/473](https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473)).

49
50
51 124

52 53 125 **Reconstruction of dated phylogeny**

54
55
56 126 Since recombination could affect the evolutionary signal, we searched for
57
58 127 recombination events in these SARS-CoV-2 genomes using RDP4[15]. No evidence
59
60

1
2
3
4 128 for recombination has been found in our dataset. We used jModelTest v2.1.6[16] to
5
6
7 129 find the best substitution model for each dataset according to the Bayesian information
8
9
10 130 criterion. The best substitution model for each dataset was listed in Supplementary
11
12 131 Table 2. The list of genomic sequences used in this study were provided in
13
14 132 Supplementary Table 3 &4. The list of genomic sequences used in this study were
15
16
17 133 openly shared via the GISAID initiative[17]. We then used the Bayesian Markov Chain
18
19
20 134 Monte Carlo (MCMC) approach implemented in BEAST v1.10.4[18] to derive a dated
21
22 135 phylogeny for each dataset. At least three replicate runs for each 100 million MCMC
23
24
25 136 steps were performed for each dataset, among which sampled parameters and trees
26
27
28 137 every 10,000 steps. For data from lineage A and B in USA and Australia during the
29
30
31 138 early phase of COVID-19, the estimation of the most appropriate combination of
32
33 139 molecular clock and coalescent models for Bayesian phylogenetic analysis was
34
35
36 140 determined using both path-sampling and stepping-stone models[19]. In order to reduce
37
38
39 141 the amount of calculation, we assumed that data from sub-lineages followed a strict
40
41
42 142 molecular clock and with an exponential population growth tree prior, as genomic
43
44
45 143 sequences used in each dataset were all from the same sub-lineage and they all had an
46
47
48 144 exponential growth. For dataset of identifying onward transmission caused by patients
49
50
51 145 being infected with VOCs after receiving at least one dose of vaccine, as genomic
52
53
54 146 sequences used in each dataset were all from the same lineage, we assumed that they
55
56
57 147 followed a strict molecular clock. The estimation of the most appropriate coalescent
58
59
60 148 models for Bayesian phylogenetic analysis was determined using both path-sampling
149 and stepping-stone models[19]. The model comparison result for datasets from lineage

1
2
3
4 150 A and B in USA and Australia were listed in Supplementary Table 5. Tracer 1.7.1[20]
5
6 151 was then used to check the convergence of MCMC chain (effective sample size >200)
7
8
9 152 and to compute marginal posterior distributions of parameters, after discarding 10% of
10
11
12 153 the MCMC chain as burn-in. We determined whether there was sufficient temporal
13
14
15 154 signal in each dataset, as it was the prerequisite for getting a reliable inference when
16
17 155 performed phylodynamic analysis. Bayesian evaluation of temporal signal (BETS)[21]
18
19
20 156 was used to evaluate the temporal signal in each dataset. BETS relies on the comparison
21
22
23 157 of marginal likelihoods of two models: the heterochronous model (with tip date) and
24
25 158 isochronous (without tip date) model. Analyses were performed with at least three
26
27
28 159 independent replicates of 100 million MCMC steps each, sampling parameters and trees
29
30
31 160 every 10,000 steps with the best substitution model and most appropriate combination
32
33 161 of molecular clock and coalescent models determined above for each dataset. The
34
35 162 marginal likelihoods were estimated by PS. The Bayes factor (BF) was then calculated
36
37
38 163 based on the likelihoods of two models (heterochronous and isochronous). If the log
39
40
41 164 $BF > 5$ (heterochronous model against isochronous model), it indicated there was
42
43 165 sufficient temporal signal in this dataset. The log BF for each dataset was listed in
44
45
46 166 Supplementary Table 6, the result suggested that the temporal signal was sufficiently
47
48 167 strong.

168

169 **Transmission Analysis**

170 As viral genomes were incompletely sampled and the pandemic is currently ongoing,
171 TransPhylo v1.4.4[22] was used to infer the transmission tree using the dated

1
2
3
4 172 phylogeny generated above as input. For B.1.617.2 (Delta) dataset of identifying
5
6 173 onward transmission caused by patients being infected with VOCs after receiving at
7
8
9 174 least one dose of vaccine, we split them into four subtrees (Supplementary Figure 1) to
10
11 175 reduce the amount of computation. The process of split tree into several subtrees did
12
13 176 not affect the result, as direct transmission always occurred in patients within close-
14
15 177 related branches. The generation time (*i.e.* the time gap from infection to onward
16
17 178 transmission, denoted as G) of COVID-19 was previously estimated as 4.8 ± 1.7
18
19 179 days[23], and we used these values to compute the shape and scale parameter of a
20
21 180 gamma distribution of G using the R package *epitrix*[24]. The distribution of sampling
22
23 181 time (*i.e.* the time gap from infection to detection and sampling) was set equal to the
24
25 182 distribution of generation time. For each dataset, we performed the TransPhylo analysis
26
27 183 several replicated runs for each 500,000 iterations simultaneously estimating the
28
29 184 transmission tree, the proportion of sampling, the within-host coalescent time N_{eg} , and
30
31 185 the two parameters of the negative binomial offspring distribution (which represents
32
33 186 the number of secondary cases caused by each infection), and then merge them together.
34
35 187 Therefore, R_t could be inferred as the median of the offspring distribution. All results
36
37 188 were generated after discarding the first part of the MCMC chains as burn-in. The
38
39 189 MCMC mixing and convergence was assessed based on the effective sample size of
40
41 190 each parameter (>200) and by visual examination of the MCMC traces (-Supplementary
42
43 191 Tables 7 & 8). The probabilities of direct transmission from one host to another were
44
45 192 estimated as the proportion of MCMC samples in which this direct transmission event
46
47 193 occurred. The expected numbers of intermediates from one host to another were
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 194 estimated as the average across the MCMC samples of the number of intermediates
5
6
7 195 between the two hosts. The probability of onward transmission for VOCs caused by
8
9 196 unvaccinated persons is calculated by taking the number of direct transmission event
10
11
12 197 caused by unvaccinated persons and dividing by the total number of unvaccinated
13
14 198 persons. The probability of onward transmission for VOCs caused by people receiving
15
16
17 199 at least one dose of vaccine is calculated by taking the number of direct transmission
18
19
20 200 event caused by people receiving at least one dose of vaccine and dividing by the total
21
22 201 number of people receiving at least one dose of vaccine.
23

24
25 202

26 27 203 **Evaluating the robustness of the estimation**

28
29
30 204 Since dated phylogeny was used to estimate the transmissibility for each lineage, we
31
32
33 205 should test whether and how the phylogenetic uncertainty and sampling bias affect the
34
35
36 206 estimation of R_t . We first tested how the phylogenetic uncertainty affect the result,
37
38 207 because only the maximum clade credibility (MCC) tree was used to estimate the
39
40
41 208 transmissibility. We used data from our previous study[25]. Ten dated phylogenetic
42
43
44 209 trees were randomly selected from the MCMC chains. The parameter setting was the
45
46
47 210 same as previous study description. The estimation of R_t from random selected tree
48
49
50 211 from MCMC chain were always lower than for the MCC tree (Supplementary Figure
51
52
53 212 2). As the MCC tree is more accurate than to trees sampled in MCMC chains, this result
54
55
56 213 suggested that the uncertainty of the phylogeny would cause an underestimation of the
57
58 214 R_t . ~~In this case~~ Consequently, the use of the MCC tree for estimation of R_t would reduce
59
60 215 the impact of phylogenetic uncertainty on the results as much as possible. In addition,

1
2
3
4 216 the sampling bias was also a key factor affecting the phylogenetic uncertainty. In order
5
6
7 217 to test if the sampling bias affect the estimation of R_t , we also repeatedly randomly sub-
8
9
10 218 sampled the data five times for each dataset using same criteria (if more than 10
11
12 219 genomes were available in a specific date, we randomly select 10 of them, otherwise
13
14 220 all genomes would be included) and then performed the same analysis.
15
16
17
18

19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

222 **Results**

223 **Lineage B has a higher transmissibility than lineage A**

224 The mean R_t for lineage A from Australia and USA were estimated as 1.75 (95%
225 credible intervals (CI) 1.43-2.11) and 1.74 (95% CI 1.61-1.89), respectively (Figure
226 1A). However, the mean R_t for lineage B from Australia and USA were estimated as
227 2.33 (95% CI 2.05-2.64) and 3.18 (95% CI 2.76-3.63), respectively (Figure 1A). Firstly,
228 the R_t of lineage B is significantly greater than that of lineage A, indicating higher
229 transmissibility of lineage B compared to lineage A. This might be the reason why
230 strains from lineage B rapidly became dominantly all over the world (Figure 1B).
231 Secondly, the R_t of lineage A from the two countries are very close, however, the R_t of
232 lineage B varied greatly between Australia and USA. We then found that the
233 composition of lineage was significantly different between the datasets from these two
234 countries (Figure 1C and D, $p < 0.01$, Fisher's exact test, two-sided). We speculated that
235 different sub-lineages within lineage B might have different transmissibility and then
236 tested the hypothesis by conducting further analysis. Since the data from lineage A was
237 limited, the evaluation of transmissibility for each sub-lineage was mainly focused on

1
2
3
4 238 those from lineage B and other emerging lineages in the same country during the same
5
6 239 periods.

7
8
9 240

10
11 241 **Some dominant lineages in the UK have similar transmissibility to B.1.1.7**

12
13
14 242 The composition of lineages in the UK is shown in Figure 2A. B.1.177 was the
15
16
17 243 dominant strain before 2021. We also found that the number of viral genomes from
18
19
20 244 England far exceeds that from other parts of the UK (Figure 2B). Besides, according to
21
22 245 the accumulation of number of viral genomes from each lineage in England, we could
23
24 246 find that only three lineages (B.1.177, B.1.1.37, B.1.1.7) grew exponentially after
25
26
27 247 October 2020 (Figure 2B). The R_t for B.1.177, B.1.1.37, B.1.1.7 were estimated as 1.08
28
29
30 248 (95% CI 1.072-1.09), 1.068 (95% CI 1.05-1.086), and 1.186 (95% CI 1.158-1.213)
31
32
33 249 (Figure 2C). The B.1.177, B.1.1.37 had similar R_t which were both close to 1. However,
34
35 250 B.1.1.7 had a significantly higher transmissibility than these two lineages. We next
36
37
38 251 tested if the significantly high R_t could be affected by sampling bias. After five
39
40
41 252 independently repeated sampling and subsequent analysis, we found that all these R_t for
42
43 253 B.1.1.7 were close to each other, ranging from 1.178 to 1.194. ~~Besides~~Furthermore, all
44
45 254 the 95% credible intervals from repeated sampling ~~also~~ did not ~~have any~~
46
47
48 255 ~~intersection~~intersect with those from lineage B.1.177 and B.1.1.37. Thus, the sampling
49
50
51 256 bias had limited effect on the estimation of R_t for each lineage. We also found that
52
53
54 257 B.1.177 had a similar transmissibility than B.1.1.37 (Student's t test, two-sided with
55
56 258 Holm–Bonferroni adjusted $p = 0.1$) (Figure 2D).

57
58
59 259
60

260 **Slightly lower transmissibility for B.1.1.54 than B.1.351 in South Africa**

261 The composition of lineages in South Africa is shown in Figure 3A. Lineage B.1.1.54
262 was the dominant strain before October 2020. Since then, the dominant strain in South
263 Africa was switched to lineage B.1.351 gradually. According to the accumulation of
264 number of viral genomes from each lineage in South Africa, we could find that only
265 lineage B.1.1.54 and B.1.351 grew exponentially after July 2020 (Figure 3B). We could
266 find the R_t for B.1.351 and B.1.54 during July 2020 and February 2021 were estimated
267 as 1.05 (95% CI 1.044-1.065) and 1.02 (95% CI 1.011-1.034), respectively (Figure 3C).
268 The difference of transmissibility between B.1.351 and B.1.54 was also significant
269 (Student's t test, two-sided $p < 0.001$) (Figure 3D). ~~In this case~~ Consequently, isolates
270 from B.1.351 had a slightly higher transmissibility than those from B.1.154.

272 **P.2 had a slightly lower transmissibility than P.1 in Brazil**

273 The composition of lineages in Brazil is shown in Figure 4A. Lineage B.1.1.33 and
274 B.1.1.28 were the dominated before January 2021. Since October 2020, two novel
275 lineages (P.1 and P.2) had gradually appeared and had shown exponential growth
276 (Figure 4B). We could find the R_t for P.1 and P.2 during December 2020 to February
277 2021 were estimated as 1.07 (95% credible intervals 1.054-1.084) and 1.06 (95%
278 credible intervals 1.049-1.070) (Figure 4C), respectively. The difference of
279 transmissibility between P.1 and P.2 was also significant (Student's t test, two-sided
280 $p = 0.016$) (Figure 4D). ~~In this case~~ Consequently, isolates from P.1 had a slightly higher
281 transmissibility than those from P.2.

1
2
3
4 2825
6 283 **B.1.617.2 has a higher transmissibility than other dominant lineages in India**

7
8
9 284 The top five dominant lineages and their corresponding proportion in India are shown
10
11 285 in Figure 5A. Since July 2020, several other lineages, like B.1, B.1.36, B.1.36.29,
12
13
14 286 emerged and grew exponentially in India (Figure 5B). ~~In this case~~Consequently, only
15
16
17 287 these five lineages were used to estimate their R_t . The R_t was estimated as 1.013 (95%
18
19 288 CI 1.006-1.021), 1.018 (95% CI 1.009 1.027), 1.019 (95% CI 1.010-1.027), 1.033 (95%
20
21 289 CI 1.026-1.040), 1.123 (95% CI 1.106-1.140) for B.1, B.1.36, B.1.36.29, B.1.617.1,
22
23
24 290 B.1.617.2, respectively (Figure 5C). After 5 independently repeated sampling and
25
26
27 291 followed analysis for each lineage, we found that both B.1.617.1 and B.1.617.2 had
28
29 292 significantly higher transmissibility than B.1, B.1.36, and B.1.36.29 (all Student's t test,
30
31 293 two-sided with Holm–Bonferroni adjusted $p < 0.001$) (Figure 5D). Furthermore,
32
33 294 B.1.617.2 also had a significantly higher transmissibility than B.1.617.1 (Student's t test,
34
35 295 two-sided with Holm–Bonferroni adjusted $p < 0.001$). In addition, the transmissibility of
36
37 296 both B.1.36, and B.1.36.29 is significantly higher than that of B.1 (both Student's t test,
38
39 297 two-sided with Holm–Bonferroni adjusted $p < 0.001$) (Figure 5D). However, similar
40
41 298 transmissibility was found between B.1.36 and B.1.36.29 (Student's t test, two-sided
42
43 299 with Holm–Bonferroni adjusted $p = 0.057$) (Figure 5D).
44
45
46
47
48
49
50

51 300

52
53 301 **Assessment of extent of onward transmission caused by partially vaccinated**
54
55 302 **individuals infected with SARS-CoV-2 VOCs**56
57
58 303 We found a total of 14 direct transmission events. Four of them, concerning three types
59
60

1
2
3
4 304 of VOCs, were transmitted by vaccinated patients among three countries (Table 1). For
5
6 305 convenience, we labelled patients involved in these four direct transmission pairs
7
8 306 identified in this study. V1/V2 and V3/V4 from Belgium and Spain are considered to
9
10 307 be transmitted by each other with a bidirectional probability for direct transmission of
11
12 308 0.99 and 0.85, respectively. However, we could not determine the direction of the
13
14 309 transmission, as the probabilities of direct transmission from both directions were
15
16 310 similar. We also found that these four patients had not been infected by others, as the
17
18 311 bidirectional probability for direct transmission between them to others (except the
19
20 312 patients who are considered to be their corresponding direct transmission pair) are all
21
22 313 extremely low (Supplementary Figure 3). In the dataset of P.1, we also found two
23
24 314 patient pairs with bidirectional probability for direct transmission as 0.76 and 0.65,
25
26 315 respectively. Furthermore, the direction of transmission was more likely from patients
27
28 316 receiving vaccines to those without receiving vaccines, as the probability of direct
29
30 317 transmission from one direction (from patients receiving vaccines to those without
31
32 318 receiving vaccines) were both >0.5 and significantly higher than that from the opposite
33
34 319 direction. Next, we tested if the phylogenetic uncertainty affected the estimation of
35
36 320 direct transmission events. We could find that the posterior probability of the branches
37
38 321 containing V1/V2, V5/N1, and V6/N2 were 1, 0.99 and 0.99, indicating the extremely
39
40 322 low phylogenetic uncertainty on these branches, further suggesting that the direct
41
42 323 transmission events estimated based on these branches are highly reliable. However,
43
44 324 the posterior probability of the branch containing V3 and V4 was only 0.33, suggesting
45
46 325 V3 and V4 did not always clustered together. We could therefore only conclude that
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 326 we found definite evidence for three direct transmission events, being transmitted by
5
6 327 patients receiving at least one dose of vaccines, with high probability. The probability
7
8
9 328 of detecting onward transmission caused by patients being infected by SARS-CoV-2
10
11 329 VOCs after receiving at least one dose of vaccine was estimated to be 1.06% (3/284).
12
13
14 330 We also calculated the probability of onward transmission caused by patients being
15
16
17 331 infected by SARS-CoV-2 VOCs who had not received any vaccine in the same dataset.
18
19 332 Ten direct transmission events were identified in the same datasets (Table 2). After
20
21
22 333 checking the phylogenetic robustness of branch containing these patients, we found that
23
24
25 334 the posterior probability of these branches all >0.9 , indicating high phylogenetic
26
27
28 335 robustness (Supplementary Figure 3). The direct transmission events identified on these
29
30
31 336 branches were therefore robust. The probability of detecting transmission from patients
32
33
34 337 infected by SARS-CoV-2 VOCs who had not received any vaccine was 1.21% (10/828).
35
36
37 338 The probability after vaccination was therefore slightly lower, but not significantly
38
39
40 339 different (Fisher exact test, $p>0.5$). This result suggested the vaccine has no obvious
41
42
43 340 effect on suppressing the continued spread of VOC, and so it needs to be implemented
44
45
46 341 in parallel with existing NPIs.
47
48
49 342

343 **Discussion**

344 Assessing the transmissibility of pathogens is essential to tailor prevention and control
345 strategies. As the COVID-19 pandemic spread, several VOCs have been found. The
346 emergence of these VOCs has caused a significant threat to public health. A previous
347 study had documented that B.1.1.7, B.1.351, P.1, and B.1.617.2 have an increased
348
349
350

1
2
3
4 348 transmissibility of 29% (95%CI: 24-33), 25% (95%CI: 20-30), 38% (95%CI: 29-48),
5
6
7 349 and 97% (95%CI: 76-117) compared to other lineages[5]. However, this conclusion
8
9 350 was based on comparing non-VOC as a whole with VOC. For some dominant lineages,
10
11 351 the number of cases added per day may be much higher than that of other lineages, but
12
13
14 352 due to its large base, the number of cases from these dominant lineages will not increase
15
16
17 353 exponentially. However, if these dominant lineages are grouped together with those
18
19
20 354 lineages in which number of cases have increased exponentially, but the number of
21
22 355 cases is not high, the advantages of transmissibility for those exponentially growing
23
24
25 356 lineages will be overwhelmed. ~~In this case~~Consequently, in order to account for not to
26
27 357 ~~ignore these~~ exponentially growing lineages, it will be very important to list them
28
29
30 358 separately as an assessment of their transmissibility.
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

360 Our results show that lineage B has a significantly higher transmissibility than lineage
361 A (Figure 1A). Together with the fact that lineage B was the dominant types of SARS-
362 CoV-2 all over the world, it seems that the high transmissibility of lineage B contributed
363 to the global pandemic to a large extent. However, we also found that the
364 transmissibility for lineage B from Australia and USA differed significantly.
365 Considering the significantly different composition of sub-lineages among these two
366 countries, we speculated that different sub-lineage within lineage B would have
367 different transmissibility. We estimated the transmissibility of VOCs and the dominant
368 lineages with exponential growth during same period in each country, so that the impact
369 of non-pharmaceutical interventions on the estimation of R_t will be consistent among

1
2
3
4 370 different lineages. Our results also indicated although VOCs had advantage of
5
6 371 transmissibility, there are still some lineages in each country with not much lower
7
8
9 372 transmissibility. These lineages should also need to be taken seriously in the
10
11 373 formulation of prevention and control policies.
12

13
14 374

15
16 375 Although vaccine manufacturers have been continuously producing vaccines, unequal
17
18 376 distribution of vaccines will still cause many people to be unable to get vaccinated in
19
20 377 the short term. In addition, even if there is an adequate supply of vaccines and
21
22 378 vaccination is being gradually progressed, it takes a relatively long period to achieve
23
24 379 complete vaccination in each country. It means that every country will have a certain
25
26 380 period of time, during which many people received only one dose of the vaccine,
27
28 381 leading to insufficient antibodies produced in their bodies. However, it was still
29
30 382 unknown whether and to what extent people receiving at least one dose of vaccines
31
32 383 could also transmit VOCs to others. We ~~found~~estimated the probability of onward
33
34 384 transmission caused by patients being infected by SARS-CoV-2 VOCs after receiving
35
36 385 at least one dose of vaccine ~~would to~~ be 1.231.06%. The similar probability of onward
37
38 386 transmission caused by patients being infected by SARS-CoV-2 VOCs without
39
40 387 receiving any vaccine indicated that only one dose of vaccine could not prevent
41
42 388 individuals from infections of SARS-CoV-2 VOCs. However, the overall extent of
43
44 389 onward transmission caused by patients being infected by SARS-CoV-2 VOCs after
45
46 390 receiving at least one dose of vaccine could be underestimated. First, not all the viral
47
48 391 genomic sequences and clinical information of patients are available. Second, the
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 392 criteria used in this study was very strict to reduce the false positive rate. Previous study
5
6 393 using household contact data demonstrated that vaccination (most of individuals
7
8
9 394 receiving one dose of vaccine) can reduce the probability of onward transmission by 50%
10
11 395 (from 10% to 5%)[26]. However, they did not distinguish between VOCs and non-
12
13
14 396 VOCs. Our results indicated that partially vaccination could not efficiently prevent the
15
16
17 397 onward transmission of SARS-CoV-2 VOCs.
18

19
20 398

21
22 399 Although the extent of onward transmission caused by patients being infected by
23
24 400 SARS-CoV-2 VOCs after receiving at least one dose of vaccine was low, the prevent
25
26
27 401 and control measures should not be loosed intemperately for following reasons. First,
28
29
30 402 the low extent of onward transmission was partially contributed to non-
31
32 403 pharmacological interventions implemented in each country. If the prevent and control
33
34
35 404 measures were abolished, the human contact frequency would be increased and then
36
37
38 405 also increase the probability of SARS-CoV-2 infection and further onward transmission.
39
40 406 Second, breakthrough infections have been identified in several countries[12, 27, 28],
41
42
43 407 indicating the vaccines against SARS-CoV-2 could not be totally neutralized. The
44
45 408 coexistence of SARS-CoV-2 and its antibodies in the human body and the continued
46
47
48 409 spread of the virus among incompletely immunized individuals will make it easier to
49
50
51 410 generate vaccine-escaped variants, which would thoroughly threaten the public health.
52
53 411 Therefore, non-pharmaceutical interventions (such as ~~some low-cost and efficient~~
54
55 412 ~~strategies, like~~ wearing masks and social distancing ~~ete~~) should be implemented in each
56
57
58 413 country before the vaccination is completed.
59
60

414 Key Points

- 415 ● ~~Except~~ In addition to VOCs, lineages with exponential growth ~~rate~~ should also be
416 paid attention in each country.
- 417 ● One dose of vaccination could not efficiently prevent the onward transmission of
418 SARS-CoV-2 VOCs
- 419 ● Non-pharmaceutical interventions (~~such as low-cost and efficient strategies, like~~
420 ~~wearing masks and social distancing etc~~) should continue to still be implemented
421 in each country during the vaccination period.

422 Biographical note

423 Liang Wang is an assistant professor at Institute of Microbiology, Chinese Academy of
424 Sciences

425 Xavier Didelot is a professor at School of Life Sciences and Department of Statistics,
426 University of Warwick

427 Yuhai Bi is a professor at Institute of Microbiology, Chinese Academy of Sciences

428 George F. Gao is a professor at Institute of Microbiology, Chinese Academy of
429 Sciences

430

431 References

- 432 1. Jiang S, Shi Z, Shu Y et al. A distinct name is needed for the new coronavirus, Lancet
433 2020;395:949.
- 434 2. Davies NG, Abbott S, Barnard RC et al. Estimated transmissibility and impact of SARS-
435 CoV-2 lineage B.1.1.7 in England, Science 2021;372:149-+.

- 1
2
3
4 436 3. Volz E, Mishra S, Chand M et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7
5
6 437 in England, *Nature* 2021;593:266-269.
7
8
9 438 4. Faria NR, Mellan TA, Whittaker C et al. Genomics and epidemiology of the P.1 SARS-
10
11 439 CoV-2 lineage in Manaus, Brazil, *Science* 2021;372:815-821.
12
13
14 440 5. Campbell F, Archer B, Laurenson-Schafer H et al. Increased transmissibility and global
15
16 441 spread of SARS-CoV-2 variants of concern as at June 2021, *Euro Surveill* 2021;26.
17
18
19 442 6. Supasa P, Zhou D, Dejnirattisai W et al. Reduced neutralization of SARS-CoV-2 B.1.1.7
20
21 443 variant by convalescent and vaccine sera, *Cell* 2021;184:2201-2211 e2207.
22
23
24 444 7. Abu-Raddad LJ, Chemaitelly H, Butt AA et al. Effectiveness of the BNT162b2 Covid-19
25
26 445 Vaccine against the B.1.1.7 and B.1.351 Variants, *N Engl J Med* 2021;385:187-189.
27
28
29 446 8. Hoffmann M, Arora P, Gross R et al. SARS-CoV-2 variants B.1.351 and P.1 escape from
30
31 447 neutralizing antibodies, *Cell* 2021;184:2384-2393 e2312.
32
33
34 448 9. Zhou D, Dejnirattisai W, Supasa P et al. Evidence of escape of SARS-CoV-2 variant
35
36 449 B.1.351 from natural and vaccine-induced sera, *Cell* 2021;184:2348-2361 e2346.
37
38
39 450 10. Liu C, Ginn HM, Dejnirattisai W et al. Reduced neutralization of SARS-CoV-2 B.1.617 by
40
41 451 vaccine and convalescent serum, *Cell* 2021;184:4220-4236 e4213.
42
43
44 452 11. Wall EC, Wu M, Harvey R et al. Neutralising antibody activity against SARS-CoV-2 VOCs
45
46 453 B.1.617.2 and B.1.351 by BNT162b2 vaccination, *Lancet* 2021;397:2331-2333.
47
48
49 454 12. Kustin T, Harel N, Finkel U et al. Evidence for increased breakthrough rates of SARS-CoV-
50
51 455 2 variants of concern in BNT162b2-mRNA-vaccinated individuals, *Nat Med* 2021;27:1379-1384.
52
53
54 456 13. Li F, Li YY, Liu MJ et al. Household transmission of SARS-CoV-2 and risk factors for
55
56 457 susceptibility and infectivity in Wuhan: a retrospective observational study, *Lancet Infect Dis*
57
58
59
60

- 1
2
3
4 458 2021;21:617-628.
5
6
7 459 14. Katoh K, Misawa K, Kuma K et al. MAFFT: a novel method for rapid multiple sequence
8
9 460 alignment based on fast Fourier transform, *Nucleic Acids Res* 2002;30:3059-3066.
10
11
12 461 15. Martin DP, Murrell B, Golden M et al. RDP4: Detection and analysis of recombination
13
14 462 patterns in virus genomes, *Virus Evol* 2015;1:vev003.
15
16
17 463 16. Darriba D, Taboada GL, Doallo R et al. jModelTest 2: more models, new heuristics and
18
19 464 parallel computing, *Nat Methods* 2012;9:772.
20
21
22 465 17. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative
23
24 466 contribution to global health, *Glob Chall* 2017;1:33-46.
25
26
27 467 18. Suchard MA, Lemey P, Baele G et al. Bayesian phylogenetic and phylodynamic data
28
29 468 integration using BEAST 1.10, *Virus Evol* 2018;4:vey016.
30
31
32 469 19. Baele G, Li WL, Drummond AJ et al. Accurate model selection of relaxed molecular clocks
33
34 470 in bayesian phylogenetics, *Mol Biol Evol* 2013;30:239-243.
35
36
37 471 20. Rambaut A, Drummond AJ, Xie D et al. Posterior Summarization in Bayesian
38
39 472 Phylogenetics Using Tracer 1.7, *Syst Biol* 2018;67:901-904.
40
41
42 473 21. Duchene S, Lemey P, Stadler T et al. Bayesian Evaluation of Temporal Signal in
43
44 474 Measurably Evolving Populations, *Mol Biol Evol* 2020;37:3363-3379.
45
46
47 475 22. Didelot X, Fraser C, Gardy J et al. Genomic Infectious Disease Epidemiology in Partially
48
49 476 Sampled and Ongoing Outbreaks, *Mol Biol Evol* 2017;34:997-1007.
50
51
52 477 23. Challen R, Brooks-Pollock E, Tsaneva-Atanasova K et al. Meta-analysis of the SARS-
53
54 478 CoV-2 serial interval and the impact of parameter uncertainty on the COVID-19 reproduction
55
56 479 number, *medRxiv* 2020:2020.2011.2017.20231548.
57
58
59
60

- 1
2
3
4 480 24. Jombart T, Cori A, Kamvar ZN et al. epitrix: small helpers and tricks for epidemics analysis.
5
6
7 481 <https://CRAN.R-project.org/package=epitrix>.
8
9 482 25. Wang L, Didelot X, Yang J et al. Inference of person-to-person transmission of COVID-19
10
11 483 reveals hidden super-spreading events during the early outbreak phase, Nat Commun
12
13 484 2020;11:5006.
14
15
16
17 485 26. Harris RJ, Hall JA, Zaidi A et al. Effect of Vaccination on Household Transmission of
18
19 486 SARS-CoV-2 in England, N Engl J Med 2021;385:759-760.
20
21
22 487 27. Brinkley-Rubinstein L, Peterson M, Martin R et al. Breakthrough SARS-CoV-2 Infections
23
24 488 in Prison after Vaccination, N Engl J Med 2021;385:1051-1052.
25
26
27 489 28. Hacisuleyman E, Hale C, Saito Y et al. Vaccine Breakthrough Infections with SARS-CoV-
28
29 490 2 Variants, N Engl J Med 2021;384:2212-2218.
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

493 **Figure Legend**

494 **Figure 1.** Difference in transmissibility between lineages A and B.

495 A. The distribution of R_t for each lineage. The black line in each distribution indicated
496 the 95% CI.

497 B. The cumulative number of SARS-CoV-2 genomes for each lineage all over the
498 world.

499 C. The heatmap of number of viral genomes for each sub-lineage in lineage A.

500 D. The heatmap of number of viral genomes for each sub-lineage in lineage B.

501 **Figure 2.** ~~Difference in transmissibility for lineages in the UK. Lineage B of SARS-~~
502 ~~CoV-2 has a higher transmissibility than lineage A.~~

503 A. The pie chart of SARS-CoV-2 lineage composition in the UK. The circle size was
504 proportion to the number of SARS-CoV-2 genomes.

505 B. The cumulative number of SARS-CoV-2 genomes for each lineage in different
506 region in the UK. The dash line indicated the earliest collection date of the data used
507 for estimating the transmissibility for each lineage.

508 C. The distribution of R_t for each lineage. The black line in each distribution indicated
509 the 95% CI.

510 D. The boxplot of repeated estimation of transmissibility by using 5 independent re-
511 sampling data for each lineage. Upper bound, center, and lower bound of box
512 represent the 75th percentile, the 50th percentile (median), and the 25th percentile,
513 respectively.

514 **Figure 3.** Difference in transmissibility for lineages in South Africa.

- 1
2
3
4 515 A. The donut chart of SARS-CoV-2 lineage composition in South Africa.
5
6 516 B. The cumulative number of SARS-CoV-2 genomes for each lineage in South Africa.
7
8
9 517 The dash line indicated the earliest collection date of the data used for estimating
10
11 518 the transmissibility for each lineage.
12
13
14 519 C. The distribution of R_t for each lineage. The black line in each distribution indicated
15
16 520 the 95% CI.
17
18
19 521 D. The boxplot of repeated estimation of transmissibility by using 5 independent re-
20
21 522 sampling data for each lineage. Upper bound, center, and lower bound of box
22
23 523 represent the 75th percentile, the 50th percentile (median), and the 25th percentile,
24
25 524 respectively.
26
27
28
29

30 **Figure 4.** Difference in transmissibility for lineages in Brazil.

- 31
32 526 A. The donut chart of SARS-CoV-2 lineage composition in Brazil.
33
34
35 527 B. The cumulative number of SARS-CoV-2 genomes for each lineage in Brazil. The
36
37 528 dash line indicated the earliest collection date of the data used for estimating the
38
39 529 transmissibility for each lineage.
40
41
42
43 530 C. The distribution of R_t for each lineage. The black line in each distribution indicated
44
45 531 the 95% CI.
46
47
48 532 D. The boxplot of repeated estimation of transmissibility by using 5 independent re-
49
50 533 sampling data for each lineage. Upper bound, center, and lower bound of box
51
52 534 represent the 75th percentile, the 50th percentile (median), and the 25th percentile,
53
54 535 respectively.
55
56
57

58 **Figure 5.** Difference in transmissibility for lineages in India.
59
60

- 1
2
3
4 537 A. The donut chart of SARS-CoV-2 lineage composition in India.
5
6 538 B. The cumulative number of SARS-CoV-2 genomes for each lineage in India. The
7
8
9 539 dash line indicated the earliest collection date of the data used for estimating the
10
11
12 540 transmissibility for each lineage.
13
14 541 C. The distribution of R_t for each lineage. The black line in each distribution indicated
15
16
17 542 the 95% CI.
18
19 543 D. The boxplot of repeated estimation of transmissibility by using 5 independent re-
20
21
22 544 sampling data for each lineage. Upper bound, center, and lower bound of box
23
24
25 545 represent the 75th percentile, the 50th percentile (median), and the 25th percentile,
26
27
28 546 respectively.
29
30 547 E. **Figure 6.** Validation of direct transmission pairs. A. The bidirectional direct
31
32
33 548 transmission probability of patients involved in direct transmission pairs and others
34
35
36 549 (excluding their corresponding direct transmission patient). Upper bound, center,
37
38
39 550 and lower bound of box represent the 75th percentile, the 50th percentile (median),
40
41
42 551 and the 25th percentile, respectively. Whiskers represent $1.5\times$ interquartile range
43
44
45 552 and points are outliers. B. The number of intermediates between patients involved
46
47
48 553 in direct transmission pairs and others (excluding their corresponding direct
49
50
51 554 transmission patient).
52
53
54 555
55
56 556
57
58 557
59
60 558

559

560 **Table 1.** The statistics of direct transmission pairs (transmission from patients receiving
 561 at least one dose of vaccines to others) identified in our study.

562

VOCs	Country	Patient_1 ID	Patient_2 ID	Probability of Patient_1 transmit to Patient_2	Probability of Patient_2 transmit to Patient_1	Bidirectional probability for direct transmission
B.1.1.7	Belgium	V1	V2	0.42551111	0.56824444	0.99375555
B.1.1.7	Spain	V3	V4	0.47526851	0.37742592	0.85269444
P.1	Brazil	V5	N1	0.58384444	0.17171111	0.75555555
P.1	French Guiana	V6	N2	0.64104444	0.01077777	0.65182222

563

564 **Table 2.** The statistics of direct transmission pairs (transmission between patients who
 565 both did not receive vaccine) identified in our study.

VOCs	Country	Patient_1 ID	Patient_2 ID	Probability of Patient_1 transmit to Patient_2	Probability of Patient_2 transmit to Patient_1	Bidirectional probability for direct transmission
------	---------	--------------	--------------	--	--	---

				Patient_2	Patient_1	transmission
B.1.1.7	Belgium	N3	N4	0.587111	0.156022	0.743133
B.1.1.7	Estonia	N5	N6	0.306932	0.34694	0.653872
B.1.1.7	Italy	N7	N8	0.253991	0.28041	0.534402
B.1.1.7	Italy	N9	N10	0.219726	0.28953	0.509256
B.1.1.7	Spain	N11	N12	0.510713	0.423852	0.934565
B.1.1.7	Spain	N13	N14	0.470843	0.392519	0.863361
B.1.1.7	Spain	N15	N16	0.281417	0.254824	0.536241
B.1.1.7	USA	N17	N18	0.475644	0.039222	0.514867
B.1.351	Belgium	N19	N20	0.382056	0.137167	0.519222
P.1	Brazil	N21	N22	0.2872	0.255111	0.542311

566

567

568 **Supplementary Information**

569 **Supplementary Figure 1.** The division of subtrees for Delta dataset.

570 **Supplementary Figure 2.** The 95% CI distribution of R_t using MCC tree and ten
571 randomly selected trees from the MCMC chains.

572 **Supplementary Figure 3.** Overview of the direct transmission events identified in our
573 datasets. The MCC tree is showed for each dataset. Branches with a posterior
574 probability >0.9 are shown by a purple circle. The size of the circle is proportional to
575 the posterior probability. Branches of patients involved in direct transmission identified
576 in this study were marked in red. Patients receiving at least one dose of vaccine were
577 highlighted in green. A. Analysis of dataset of SARS-CoV-2 B.1.1.7 (Alpha) in
578 Belgium; B. Analysis of dataset of SARS-CoV-2 B.1.1.7 (Alpha) in Spain; C. Analysis
579 of dataset of SARS-CoV-2 B.1.1.7 (Alpha) in USA; D. Analysis of dataset of SARS-
580 CoV-2 B.1.1.7 (Alpha) in other countries; E. Analysis of dataset of SARS-CoV-2
581 B.1.351 (Beta); F. Analysis of dataset of SARS-CoV-2 P.1 (Gamma); G. Analysis of
582 dataset of SARS-CoV-2 B.1.617.2 (Delta).

583 **Supplementary Table 1.** List of 30 masked sites in SARS-CoV-2 genome.

584 **Supplementary Table 2.** The best substitution model for dataset from each dataset.

585 **Supplementary Table 3.** The acknowledgement table of viral genomes used for
586 estimating R_t .

587 **Supplementary Table 4.** The acknowledgement table of viral genomes used for
588 evaluating the onward transmission caused by patients being infected with SARS-CoV-
589 2 VOCs after receiving at least one dose of vaccine.

1
2
3
4 590 **Supplementary Table 5.** Log-marginal likelihood estimates from model selection by
5
6 591 using the path-sampling (PS) and stepping-stone (SS) approaches for lineage A and B.
7
8

9 592 **Supplementary Table 6.** Bayesian evaluation for the temporal signal of dataset from
10
11 593 each dataset.
12
13

14 594 **Supplementary Table 7.** The estimation of R_t and corresponding effective size of each
15
16 595 dataset.
17
18

19 596 **Supplementary Table 8.** The parameters of offspring distribution estimated for
20
21 597 different dataset.
22
23
24

25 598

27 599 **Acknowledgements**

30 600 We gratefully acknowledge the authors from the originating laboratories and the
31
32 601 submitting laboratories where genetic sequence data were generated and shared via
33
34 602 GISAID, enabling this research.
35
36

37 603 **Funding**

40 604 This work was supported by the Strategic Priority Research Programme of the Chinese
41
42 605 Academy of Sciences (Grant numbers XDB29010102 and XDA19090118), National
43
44 606 Natural Science Foundation of China (NSFC) (Grant numbers 32041010 and
45
46 607 31900155). Y.B. is supported by the NSFC Outstanding Young Scholars (Grant number
47
48 608 31822055) and Youth Innovation Promotion Association of CAS (Grant number
49
50 609 2017122). X.D. is supported by the NIHR Health Protection Research Unit in Genomics
51
52 610 and Enabling Data.
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

612 **Conflict of Interest**

613 All the authors declared no conflict of interests

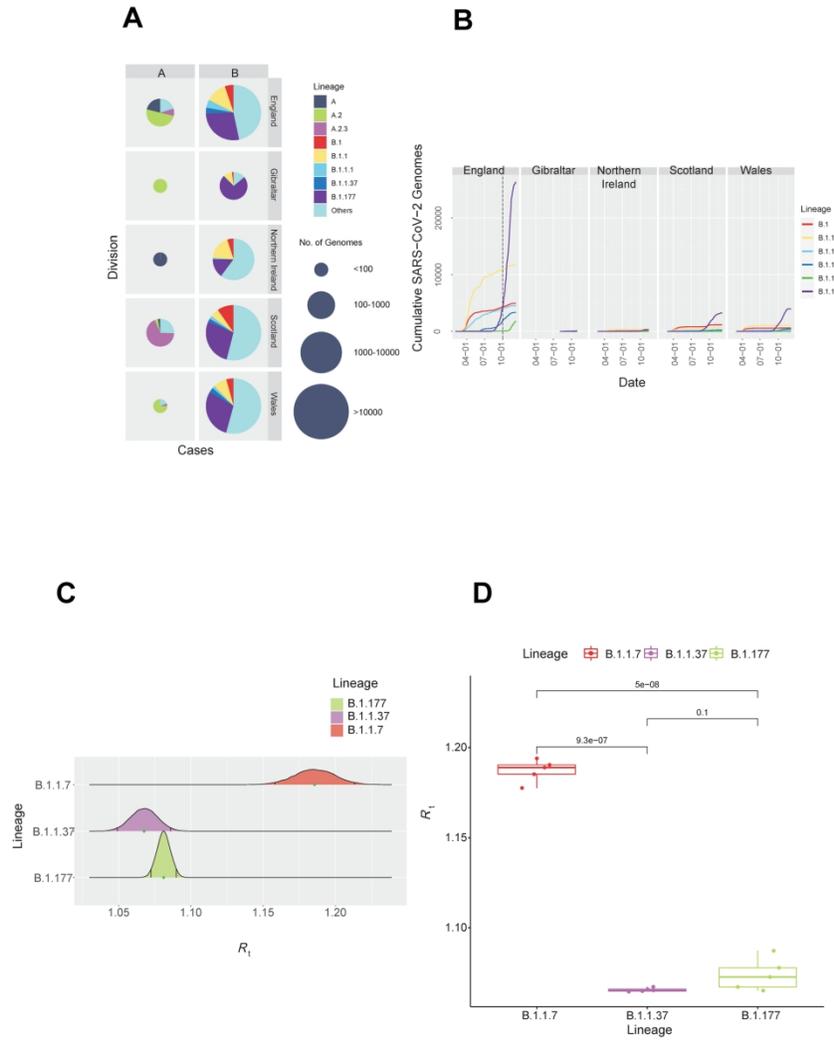
614

615

616

For Peer Review

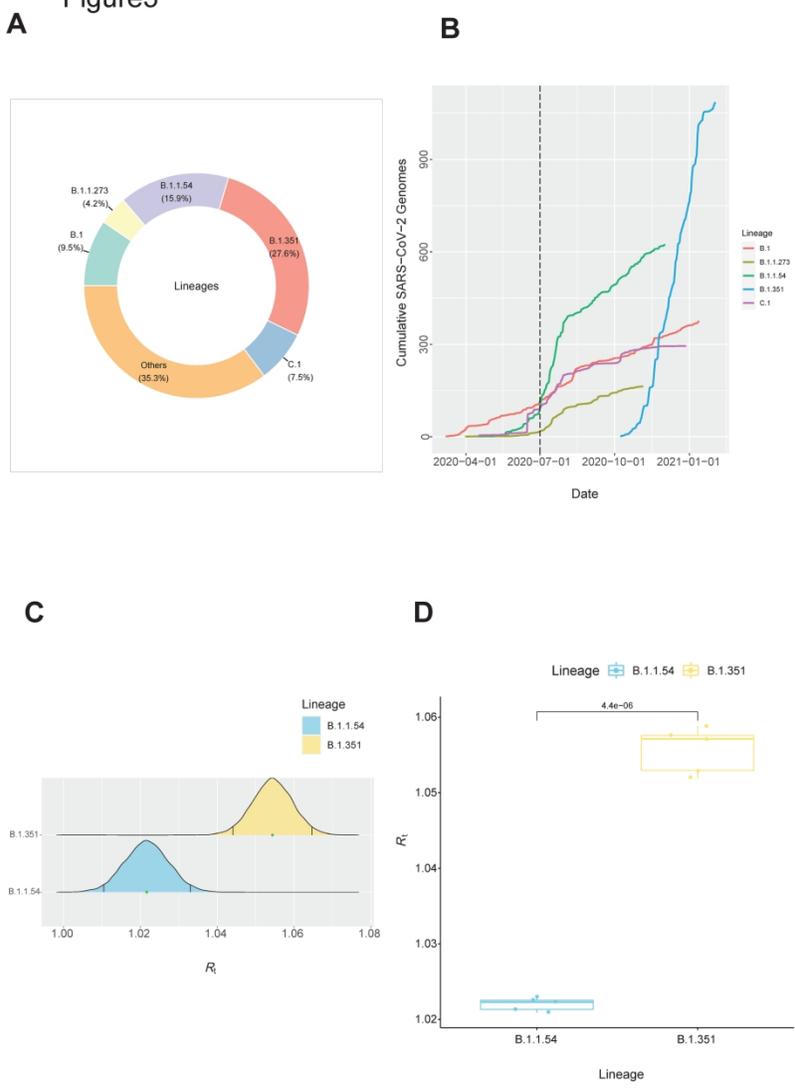
Figure 2



209x297mm (300 x 300 DPI)

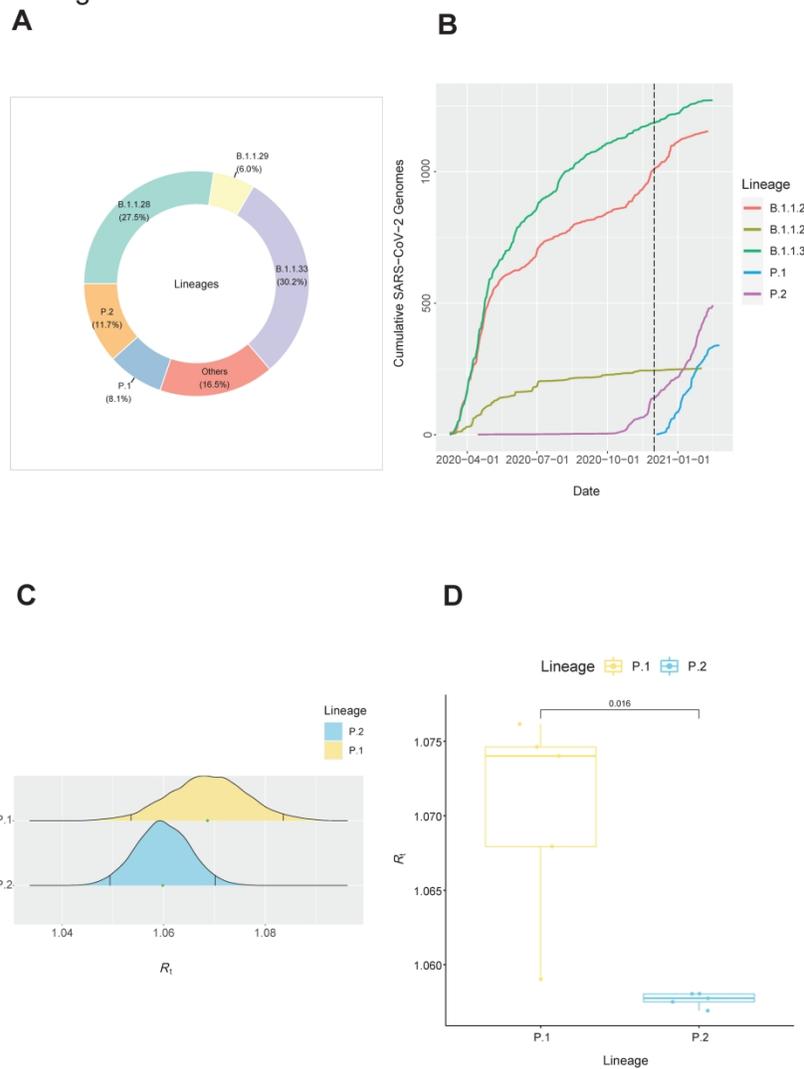
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure3



209x297mm (300 x 300 DPI)

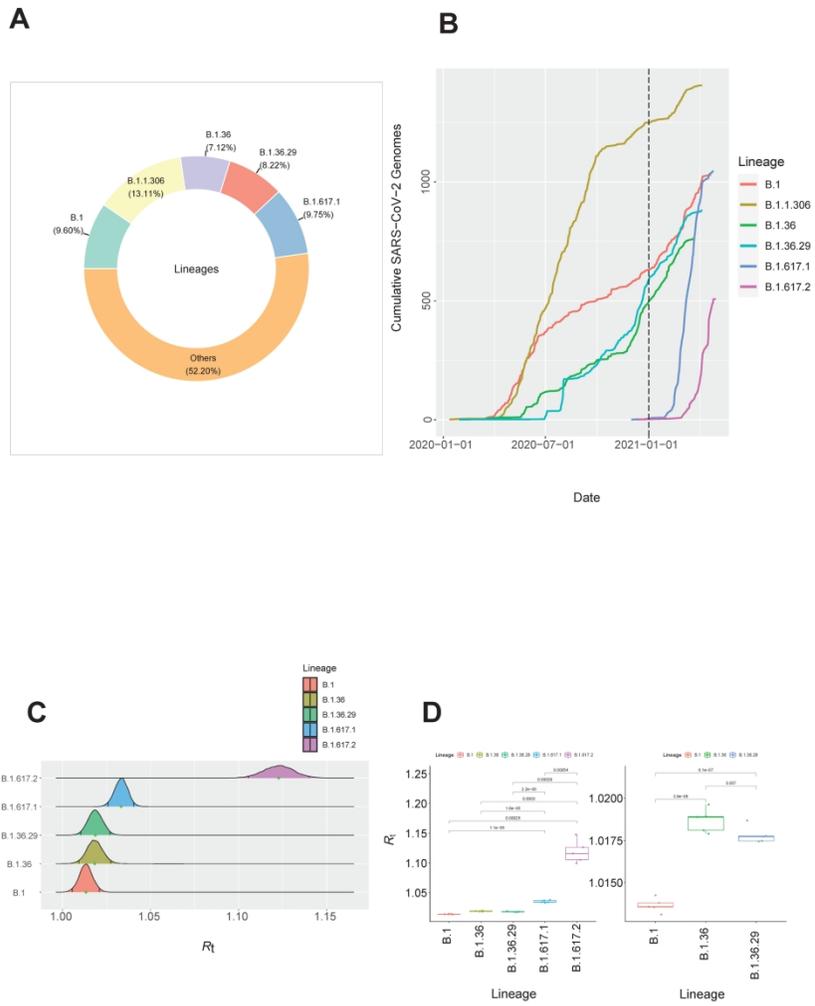
Figure 4



209x297mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

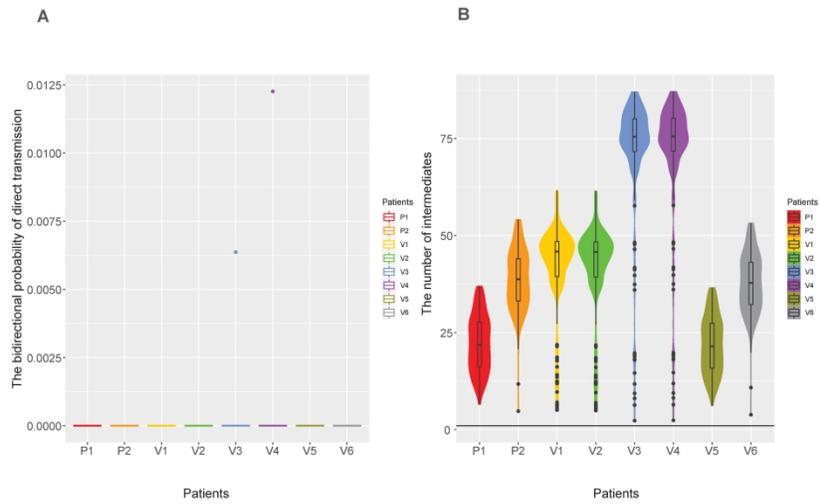
Figure5



209x297mm (300 x 300 DPI)

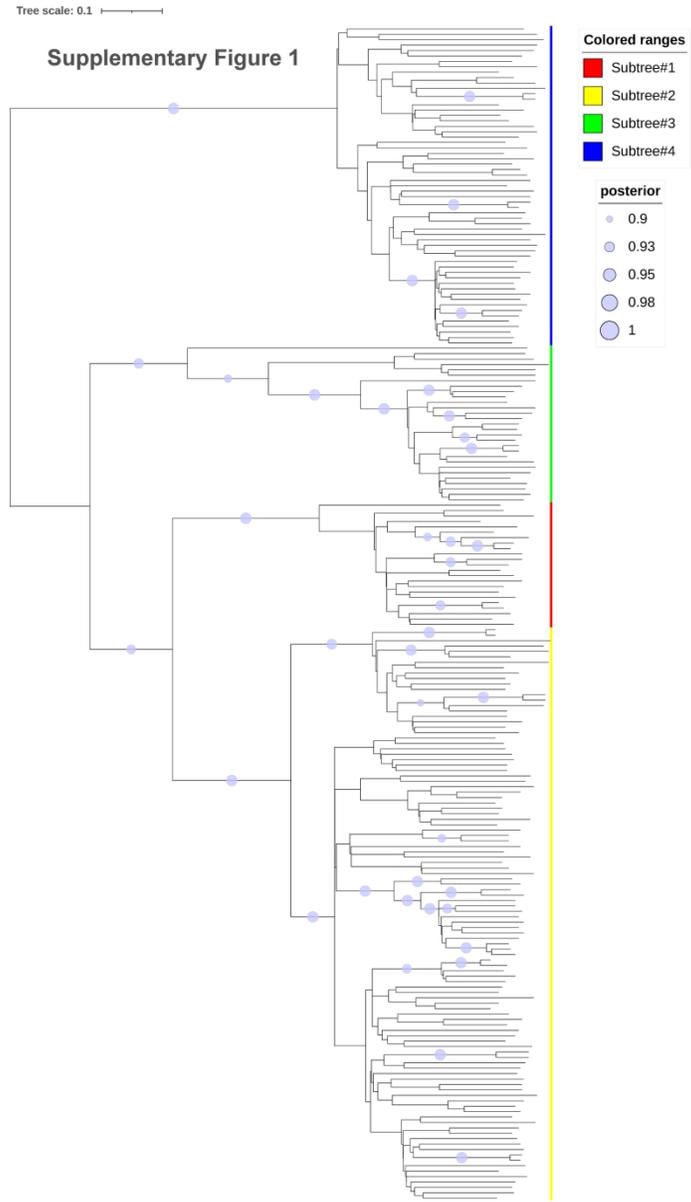
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 6

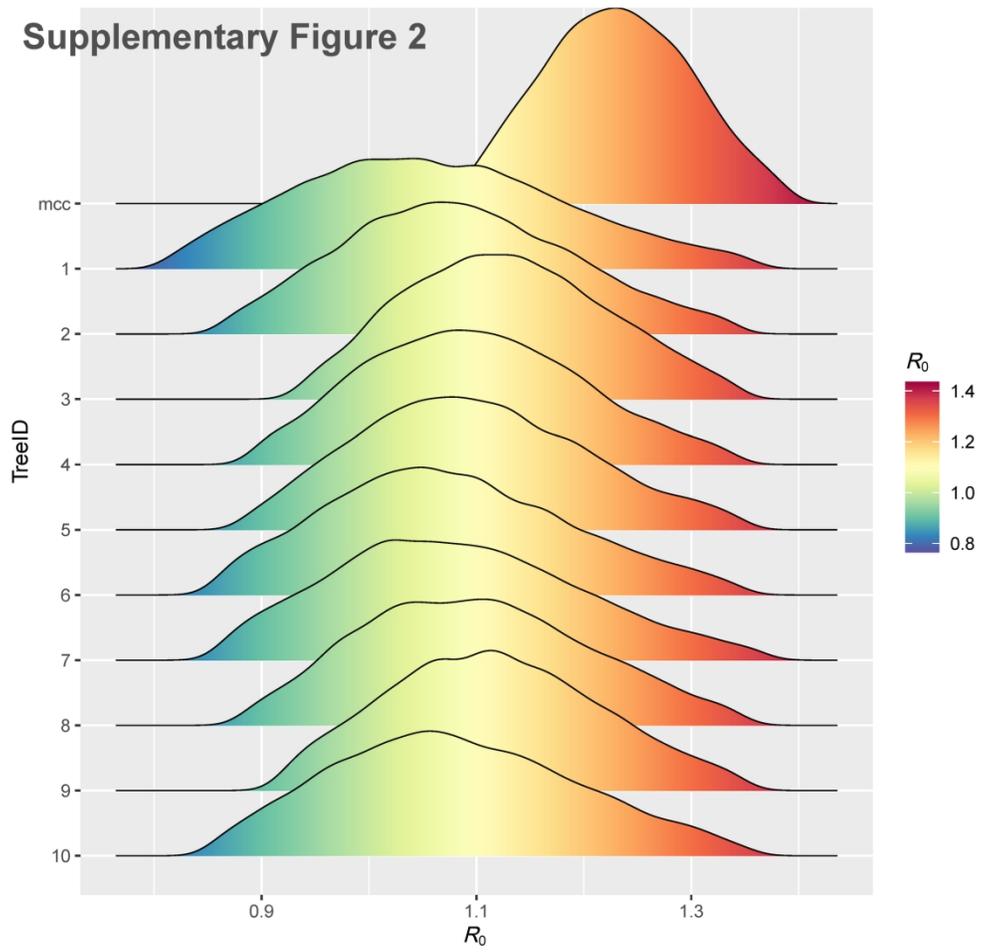


209x297mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

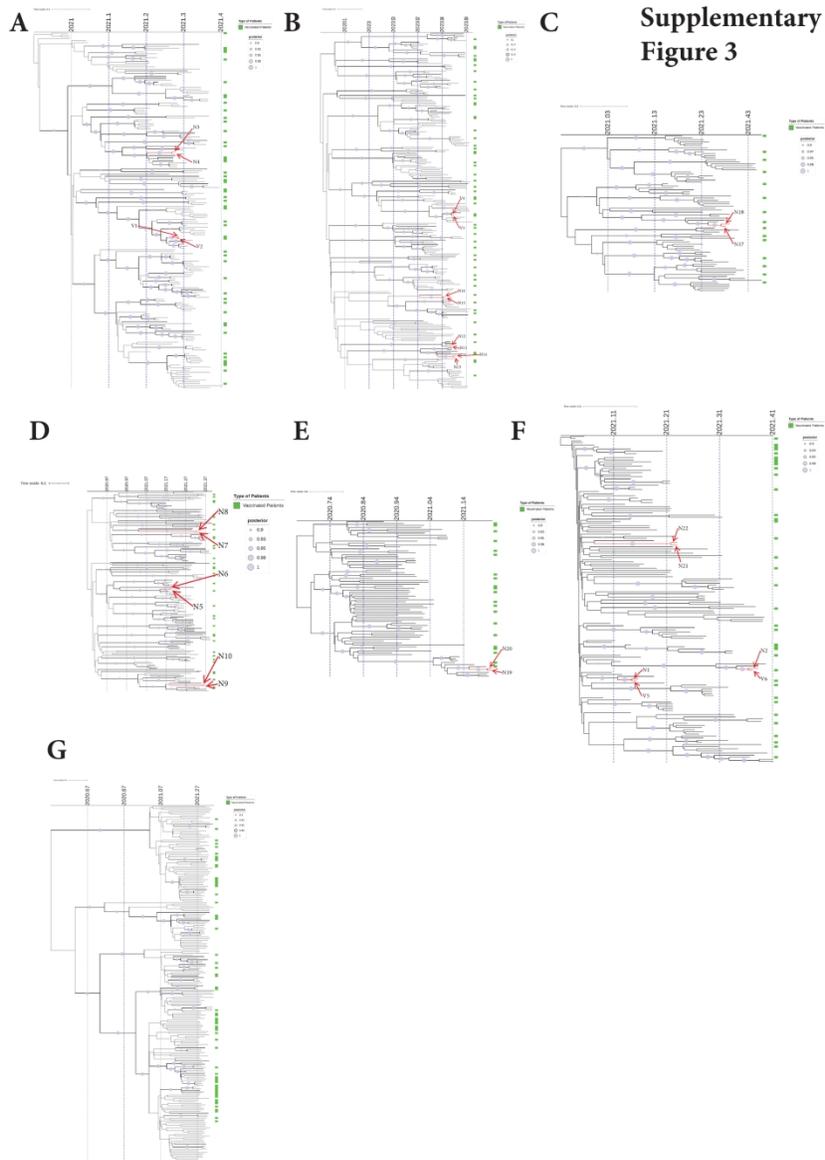


194x340mm (300 x 300 DPI)



177x177mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



209x297mm (300 x 300 DPI)