

1 Predicting the time course of replacements 2 of SARS-CoV-2 variants using relative 3 reproduction numbers

4
5 Chayada Piantham¹ and Kimihito Ito^{2*}

6 ¹*Graduate School of Infectious Diseases, Hokkaido University; chayada@czc.hokudai.ac.jp*

7 ²*International Institute for Zoonosis Control, Hokkaido University; itok@czc.hokudai.ac.jp*

8 Abstract

9 The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has continuously evolved
10 since its introduction to the human population in 2019. Natural selection selects variants with
11 higher effective reproduction numbers, increasing the overall transmissibility of the circulating
12 viruses. In order to establish effective control measures for a new variant, it is crucial to know its
13 transmissibility and replacement time course in early phases of the variant replacement. In this
14 paper, we conduct retrospective prediction tests of the variant replacement from Alpha to Delta
15 in England. Our method firstly estimated the relative reproduction number, the ratio of the
16 reproduction number of a variant to that of another, from partial observations up to a given time
17 point. Secondly, the replacement time course after the time point was predicted based on the
18 estimates of relative reproduction number. Thirdly, the estimated relative reproduction number
19 and the predicted time course were evaluated by being compared to those estimated using the
20 entire observations. We found that it is possible to estimate the relative reproduction number of
21 Delta with respect to Alpha when the frequency of Delta was more than or equal to 0.25. Using
22 these relative reproduction numbers, predictions targeting on 1st June 2021, the date when the
23 frequency of Delta reached 0.90, had maximum absolute prediction errors of 0.015 for
24 frequencies of Delta and 0.067 for the average relative reproduction number of circulating
25 viruses with respect to Alpha. These results suggest that our method allows us to predict the time
26 course of variant replacement in future from partial datasets observed in early phases of variant
27 replacement.

28

29 **Keywords:** SARS-CoV-2, relative reproduction number, Delta, Alpha, variant replacement

30 Introduction

31 Since its first emergence in the human population in 2019, the severe acute respiratory syndrome
32 coronavirus 2 (SARS-CoV-2) has been generating new variants. Natural selection selects new
33 variants that has higher effective reproduction numbers than other circulating variants. As a
34 result, the average transmissibility in the viral population increases over time [1]. As of date,
35 Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), and Omicron (B.1.1.529) have
36 been assigned as variants of concern (VOCs) because of their increased transmissibility
37 compared to previous variants [2].

38

39 It is important to know how transmissible the new variants are compared to previously
40 circulating variants because the average reproduction number of the circulating virus changes
41 when new variants become dominant. Several studies have analyzed the reproduction numbers of
42 new variants that have replaced old ones. Volz et al. estimated the effective reproduction number
43 of Alpha in England to be 1.5–2.0 times higher than that of non-VOCs using a logistic growth
44 model for variant frequencies [3]. Leung et al. estimated the basic reproduction number of Alpha
45 to be 1.75 times higher than that of previously circulating variants in England using a renewal-
46 equation-based model [4]. Ito et al. estimated the effective reproduction number of Delta to be
47 1.35 times higher than that of Alpha from variant frequencies observed in Japan by using an
48 approximated version of the renewal-equation-based model [5]. Using the same method, Ito et al.
49 estimated the effective reproduction number of Omicron to be 3.15 times higher than that of
50 Delta in Denmark [6], and Nishiura et al. estimated the effective reproduction of Omicron to be
51 4.2 times higher than that of Delta in South Africa [7].

52

53 In order to prepare control measures against new variants, it is crucial to predict the time course
54 of the variant replacements in advance. The prediction of variant selection has been widely
55 studied in seasonal influenza viruses [8]. Łuksza and Lässig developed a fitness model using
56 mutations on epitopes and non-epitopes to predict selected variants [9]. Huddleston et al.
57 predicted the future frequency of variants using its current frequency, the antigenic novelty of

58 epitopes, and the mutational load in non-epitopes [10]. Piantham and Ito modeled the fixation
59 probability of variants using variant frequency and statistics on patient ages [11]. In the case of
60 seasonal influenza, the main driving force of natural selection was the population immunity
61 acquired from previous infections. In contrast, most of the human population are considered
62 naïve to SARS-CoV-2 infections, and a new method to predict the time course of variant
63 replacement of SARS-CoV-2 needs to be developed.

64
65 The transmissibility of an infectious agent can be measured by its reproduction number. The
66 effective reproduction number at time t (R_t) is defined as the average number of people
67 someone infected at time t could expect to produce if conditions should remain unchanged
68 [12]. When more than one variant of the infectious agent is circulating, the relative reproduction
69 number can be used to measure the relative transmissibility of a variant compared to a baseline
70 variant [4,13]. However, the method requires the number of new cases in addition to the
71 frequencies of variants, and it is not applicable for predicting future time course of variant
72 replacement. Using approximations, Ito et al. proposed a method to determine the relative
73 reproduction number without knowing the number of new cases [5]. This method allows us to
74 predict the future time course of variant replacements.

75
76 Nucleotide sequences of SARS-CoV-2 variants collected worldwide have been submitted to and
77 accumulated on the GISAID database [14]. It is known that different geographical locations have
78 different distributions of variants [15]. As of 28th February 2022, a total of 8,753,735 sequences
79 have been registered on the database worldwide. Of these, 1,786,080 (20%) were submitted from
80 England, which has their population account for 0.71% of the world population. These numbers
81 indicate that England is one of the locations having highest sequencing capacity. In England, the
82 Alpha–Delta replacement was observed from March 2021 to June 2021. The sequence
83 information during the Alpha–Delta replacement in England is one of the best datasets to
84 evaluate the predictability of variant replacement in SARS-CoV-2.

85
86 In this study, we conduct retrospective prediction tests using the nucleotide sequences collected
87 in England during the Alpha–Delta replacement. For each given time point, we use partial
88 sequence data observed only up to that time point to estimate the relative reproduction number of

89 Delta with respect to (w.r.t.) Alpha. The estimated relative reproduction number is then used to
90 predict the future time course of variant replacement. The estimated relative reproduction
91 numbers and the predicted time courses are evaluated by being compared to those estimated
92 using the entire dataset.

93 Materials and Methods

94 Nucleotide sequences

95 Nucleotide sequences of SARS-CoV-2 viruses collected from England during 1st January 2021 to
96 31st July 2021 were downloaded from the GISAID database on 16th November 2021 [14]. Of
97 these, 411,123 sequences had complete information about date of sample collection in the
98 metadata. The PANGO lineage names [16] of these sequences were collected from metadata and
99 recorded with their collection dates (Supplementary Table S1). Sequences that are labeled as
100 “B.1.1.7” or sublineage names starting with “Q.” were classified as the Alpha variant. Sequences
101 that are labeled as “B.1.617.2” or sublineage names starting with “AY.” were classified as the
102 Delta variant. There were 11,773 sequences of lineages other than Alpha and Delta, and these
103 were ignored in subsequent analyses. A total of 399,350 sequences of Alpha (192,250) and Delta
104 (207,100) were used for counting the daily numbers of sequences belonging to Alpha and Delta
105 (Supplementary Fig. S1).

106 Model of Advantageous Selection

107 We estimated the relative reproduction number of a variant w.r.t. a baseline variant using an
108 approximated version of the renewal-equation-based model [5], which is based on Fraser’s time-
109 since-infection model [12]. Briefly, suppose the viral population consisted of only variant X at
110 time t_0 . Variant Y was introduced into the population at calendar time $t_Y > t_0$ with an initial
111 frequency of $q_Y(t_Y)$. Note that $q_Y(t) = 0$ for $t_0 \leq t < t_Y$. The date of variant introduction, t_Y ,
112 was set to the first date when more than one sequence of that variant was counted. We assume
113 the effective reproduction number of variant Y was k times higher than that of variant X . Let
114 $f(\tau)$ be the probability mass function of generation time τ . We assume $f(\tau)$ follows the gamma
115 distribution with a shape parameter of 3.42 and a scale parameter of 1.36 [17]. We discretize

116 $f(\tau)$ to $g(j) = \int_j^{j+1} f(\tau)d\tau$ for $1 \leq j \leq 20$. We truncate the generation time distributions at $\tau =$
 117 1 and $\tau = 20$ by adding $\int_0^1 f(\tau)d\tau$ to $g(1)$ and $\int_{21}^{\infty} f(\tau)d\tau$ to $g(20)$ so that $\sum_{j=1}^{20} g(j) = 1$. Let
 118 $I(t)$ be the total number of new infections by either X or Y at calendar time t , the effective
 119 reproduction numbers of variant X and Y can be calculated as

$$120 \quad R_X(t) = \frac{q_X(t)I(t)}{\sum_{j=1}^{20} g(j)q_X(t-j)I(t-j)} \quad (1)$$

$$121 \quad R_Y(t) = \frac{q_Y(t)I(t)}{\sum_{j=1}^{20} g(j)q_Y(t-j)I(t-j)}. \quad (2)$$

122 Since the effective reproduction number of variant Y is k times higher than that of variant X , the
 123 effective reproduction number of variant Y at time t is given by

$$124 \quad R_Y(t) = kR_X(t). \quad (3)$$

125 Assuming that the viral population at time t comprises of only variants X and Y , the frequency of
 126 variant Y at calendar time t , $q_Y(t)$, can be calculated as

$$127 \quad q_Y(t) = \frac{q_Y(t)I(t)}{q_X(t)I(t) + q_Y(t)I(t)}. \quad (4)$$

128 We assume that the numbers of new infections do not vary greatly for 20 days, i.e.

$$129 \quad I(t-1) \cong \dots \cong I(t-20) \quad (5)$$

130 for $t > t_0$. Using this approximation with Equations (1), (2), and (3), we can rewrite Equation
 131 (4) using $q_Y(t-j)$ for $1 \leq j \leq 20$ as

$$132 \quad q_Y(t) = \frac{k \sum_{j=1}^{20} g(j)q_Y(t-j)}{\sum_{j=1}^{20} g(j)q_X(t-j) + k \sum_{j=1}^{20} g(j)q_Y(t-j)}. \quad (6)$$

133 The average relative reproduction number of circulating viruses at time t w.r.t. variant X is given
 134 by

$$135 \quad q_X(t) + kq_Y(t). \quad (7)$$

136 Parameter estimation from the number of sequences

137 Let $N_X(t)$ and $N_Y(t)$ be the number of sequences of variant X and Y observed at calendar time t ,
 138 respectively. Suppose that variant Y is sampled and sequenced following a beta-binomial
 139 distribution having distribution parameters of $\alpha = q_Y(t)M$ and $\beta = (1 - q_Y(t))M$. Note that
 140 this beta-binomial distribution has a mean of $(N_X(t) + N_Y(t))q_Y(t)$ and a variance of

141 $\frac{(N_X(t)+N_Y(t)q_X(t)q_Y(t)(N_X(t)+N_Y(t)+M)}{M+1}$. The beta-binomial distribution becomes the binomial
142 distribution when $M = \infty$. The following equation gives the likelihood function of parameters k ,
143 $q_Y(t_Y)$, and M for observing $N_X(t)$ and $N_Y(t)$ sequences of variant X and Y at calendar time t :

$$144 \quad L(k, q_Y(t_Y), M; N_X(t), N_Y(t)) = \binom{N_X(t) + N_Y(t)}{N_Y(t)} \frac{B(N_Y(t) + \alpha, N_X(t) + \beta)}{B(\alpha, \beta)} \quad (8)$$

145 where $\alpha = q_Y(t)M$, $\beta = (1 - q_Y(t))M$, and $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$.

146
147 The likelihood for observing $N_Y(t)$ sequences of variant Y during the period on calendar times
148 t_1, \dots, t_n is given by the product of the above formula for $1 \leq t \leq n$. The estimates of k , $q_Y(t_Y)$,
149 and M were obtained by maximizing the likelihood function using observed numbers of
150 sequences in England from $t = t_Y$ until the latest t in which $q_Y(t) < 1$. The 95% confidence
151 intervals (95% CI) of parameters were determined using the profile likelihood method [18].
152 From the maximum likelihood estimates of k and $q_Y(t)$, the average relative reproduction
153 number of circulating viruses w.r.t. Alpha at time t was estimated from Equation (7).

154 Prediction of variant frequency and average relative reproduction 155 number

156 Frequencies of Delta and average relative reproduction numbers of circulating viruses w.r.t.
157 Alpha in future were predicted using the maximum likelihood estimates of parameters calculated
158 from early observations. First, we estimated k , $q_Y(t_Y)$, and M using all observations of $N_X(t)$
159 and $N_Y(t)$ for $t_Y \leq t \leq t_n$ where t_n is the latest calendar time t such that $q_Y(t) < 1$, by
160 maximizing likelihood of products of Equation (8). Using the maximum likelihood estimates
161 and their 95% CIs, we determined the calendar times t when the estimated frequency $q_Y(t)$
162 exceeded 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75,
163 0.80, 0.85, 0.90, and 0.95. For each date t determined above, we calculated the maximum
164 likelihood estimates of k , $q_Y(t_Y)$, and M using observations no later than t . Frequencies of Delta
165 and average relative reproduction numbers of circulating viruses w.r.t. Alpha in future were
166 predicted by substituting k and $q_Y(t_Y)$ in Equations (6) and (7), respectively.

167 Results

168 Estimation of relative reproduction number from entire observations

169 Parameters of the model were estimated using the entire observations from 18th March to 4th July
170 2021 in England (Table 1). The relative reproduction number (k) of Delta w.r.t. Alpha was
171 estimated to be 1.67 (95% CI: 1.64, 1.67) with a beta-binomial distribution parameter (M) of
172 286.82 (95% CI: 201.83, 403.22).

173
174 Figure 1a shows the observed and estimated frequencies of Delta during the Alpha–Delta
175 replacement in England. The blue curve and black dots around the blue curve represent the
176 maximum likelihood estimates and 95% CI of frequencies of Delta. The gray area represents
177 95% equal-tailed intervals of the beta distribution with the parameters $q_Y(t)$ and M in the 95%
178 CI. Figure 1b shows the maximum likelihood estimations and 95% CI of the average relative
179 reproduction number of circulating viruses w.r.t. Alpha during the same period. Dashed vertical
180 lines in both panels indicate the dates when frequencies of Delta exceeded each 0.05 increment
181 from 0.05 to 0.95 (Table 2). It took 48 days for Delta to reach from frequencies of 0.05 to 0.95.

182 Estimation of relative reproduction number from partial observations

183 Table 3 shows the parameters of our model estimated using the partial data collected no later
184 than each maximum likelihood date in Table 2. The maximum likelihood estimate of k using
185 observations of the entire period in the Alpha–Delta replacement was 1.67 (Table 1). We call this
186 estimate the ‘final estimate’. The final estimate was within 95% CIs of estimations in six (0.15,
187 0.25, 0.30, 0.35, 0.40, and 0.45) out of nine estimations using the partial observations before
188 Delta reached frequencies of 0.50. Three early estimations made at frequencies of 0.05, 0.10, and
189 0.20 failed to cover the final estimate of k in their 95% CIs. The five estimations made at
190 frequencies greater than or equal to 0.25 covered the final estimate of k in their 95% CIs. This
191 implied that it was possible to accurately estimate the relative reproduction number of Delta
192 w.r.t. Alpha when frequencies of Delta became 0.25 or later. It took 34 days for Delta to reach a
193 frequency of 0.95 from when it was 0.25 (Table 2). Therefore, we would be able to estimate the
194 relative reproduction number of Delta w.r.t. Alpha more than one month before its fixation. The

195 estimated values of M also became steady at around 300.00 at frequencies greater than or equal
196 to 0.25 (Table 3).

197 Prediction of variant frequency and average relative reproduction 198 number

199 We conducted retrospective prediction tests on the future frequency of Delta and the average
200 relative reproduction number of circulating viruses w.r.t. Alpha using model parameters in Table
201 3, which were estimated from partial observations. Figure 2 shows predicted trajectories of the
202 Alpha–Delta replacement using partial observations up to different time points in Table 2. The
203 maximum likelihood predictions made at frequencies of 0.05, 0.10, and 0.20 overestimated the
204 future frequencies of Delta (Figures 2a–b, 2d), while predictions made at frequencies greater
205 than or equal to 0.25 fitted well with future observations (Figures 2e–i).

206

207 According to the final estimate using the entire observations, Delta exceeded frequencies of 0.50,
208 0.70, and 0.90 on 13th May, 20th May, and 1st June 2021, respectively (Table 2). We evaluated
209 the accuracy of predictions by analyzing predictions targeted on these dates (Figure 3). As the
210 relative reproduction numbers were overestimated when predictions were made before
211 frequencies of Delta reached 0.25, the frequencies of Delta on the target dates were also
212 overestimated in these predictions (Figure 3a–c). In contrast, predictions made when frequencies
213 of Delta were greater than or equal to 0.25 were close to the final estimate of frequency using
214 entire observations. When frequencies of Delta were below 0.25, the predictions targeted on 13th
215 May, 20th May, and 1st June 2021 had median errors of 0.074, 0.069, and 0.036 with maximum
216 absolute errors of 0.143, 0.121, and 0.055, respectively (Table 4). When frequencies of Delta
217 were greater than or equal to 0.25, the predictions targeted on 13th May, 20th May, and 1st June
218 2021 had median errors of 0.018, 0.011, and 0.000 with maximum absolute errors of 0.023,
219 0.023, and 0.015, respectively (Table 5). Predictions made when frequencies of Delta were
220 greater than or equal to 0.25 have significantly smaller prediction errors than those made when
221 frequencies of Delta were less than 0.25 (*p*-values of 0.016, 0.003, and 0.001 using two-sided
222 Wilcoxon rank sum test for predictions targeted on 13th May, 20th May, and 1st June 2021,
223 respectively).

224
225 From the final estimate using entire observations, the average relative reproduction numbers of
226 circulating viruses w.r.t. Alpha on 13th May, 20th May, and 1st June 2021 were estimated to be
227 1.34, 1.47, and 1.60, respectively (Table 2). In the same way as the frequency of Delta, the
228 average relative reproduction numbers of circulating viruses w.r.t. Alpha on the target dates were
229 overestimated when predictions were made before frequencies of Delta reached 0.25 (Figure 3d–
230 f). When the frequencies of Delta were below 0.25, predictions targeted on 13th May, 20th May,
231 and 1st June 2021 had median prediction errors of 0.088, 0.098, and 0.086 and maximum
232 absolute errors of 0.168, 0.173, and 0.143, respectively (Table 4). When frequencies of Delta
233 were greater than or equal to 0.25, predictions targeted on 13th May, 20th May, and 1st June 2021
234 had median errors of 0.014, 0.010, and –0.019 with maximum absolute errors of 0.023, 0.044,
235 and 0.067, respectively (Table 5). Predictions made when frequencies of Delta were greater than
236 or equal to 0.25 have significantly smaller prediction errors than those made when frequencies of

237 Delta were less than 0.25 (p -values of 0.016, 0.003, and 0.001 using two-sided Wilcoxon rank
238 sum test for predictions targeted on 13th May, 20th May, and 1st June 2021, respectively).
239

240 Discussion

241 We analyzed the replacement from the Alpha variant to the Delta variant in England using
242 nucleotide sequences on the GISAID database collected between 1st January 2021 to 31st July
243 2021. The estimated relative reproduction number, k , of Delta w.r.t. Alpha was 1.67 (95% CI:
244 1.64–1.67) with a beta-binomial distribution parameter (M) of 286.82 (95% CI: 201.83–403.22)
245 (Table 1). The relative reproduction number was accurately estimated from early observations
246 once frequencies of Delta reached 0.25 (Table 3). Using these relative reproduction numbers,
247 predictions targeting on 1st June 2021, the date when the frequency of Delta reached 0.90, had
248 maximum absolute prediction errors of 0.015 for frequencies of Delta and 0.067 for the average
249 relative reproduction number of circulating viruses with respect to Alpha (Table 5).

250
251 Several studies have estimated the relative reproduction of Delta w.r.t. Alpha. Ito et al. estimated
252 the relative reproduction number of Delta w.r.t. Alpha in Japan to be 1.35 [5]. Hansen estimated
253 the relative reproduction number of Delta w.r.t. Alpha in Denmark to be 2.17 [19]. In this study,
254 the relative reproduction number of Delta w.r.t. Alpha was estimated to be 1.67 (95% CI: 1.64–
255 1.67) (Table 1). Figgins and Bedford found that the relative reproduction number of Delta and
256 Alpha w.r.t. non-VOC variants in the United States were different depending on the states [20].
257 The differences in relative reproduction numbers of Delta w.r.t. Alpha among countries or states
258 may be attributed to the differences in the vaccine usage or the ethnicity of the target
259 populations.

260
261 Our model assumes that the sequences on GISAID database was sampled following a beta-
262 binomial distribution. We could use the binomial distribution in the model instead of the beta-
263 binomial distribution. The model using beta-binomial distribution resulted in lower Akaike
264 information criteria (AIC) compared to the model using only binomial distribution
265 (Supplementary Table S2). This means that the observed variance was larger than the variance of
266 the binomial distribution. The additional variance to the binomial distribution might be attributed

267 to the difference between variant frequencies among subpopulations. For example, different
268 regions may show different progresses in the variant replacement. The same may be true for
269 different age groups. The lower AIC in the beta-binomial distribution model may mean that the
270 target population was not well-mixed.

271
272 When a new variant having higher effective reproduction number than the circulating variants
273 emerges, the new variant will be selected by the natural selection. As a result, the average
274 relative reproduction number of the circulating viruses w.r.t. old variants increases. Our
275 retrospective prediction tests show that the average relative reproduction number of circulating
276 viruses w.r.t. Alpha were accurately predicted when observed frequencies of Delta exceeded
277 0.25. However, some predictions made at higher frequencies (0.55, 0.70, 0.75, and 0.85) were
278 slightly lower than the final estimation using entire observations (Figure 3e–f). One possible
279 reason for these is the relaxation of lockdown restrictions from 17th May 2021. Assuming that
280 variant replacement occurs earlier and sampling is more dense in the city area than countryside,
281 the relaxation of lockdown restrictions mixed viral population in both areas, making the
282 observations of Alpha in city area more than expected.

283
284 Prediction tests conducted in this study did not consider delay in sequence submissions. To
285 simulate real-time predictions with submission delays, we conducted the same analysis using
286 observations that were submitted no later than the day of prediction. The results were not as good
287 as the analysis which assumed that there was no submission delay (Supplementary Fig. S2).
288 Even if predictions were made when frequencies of Delta were greater than or equal to 0.25, the
289 frequencies of Delta were overestimated with maximum absolute errors of 0.084, 0.080, and
290 0.041 for the predictions targeted on 13th May, 20th May, and 1st June 2021, respectively
291 (Supplementary Table S3). The maximum absolute errors of average relative reproduction
292 number of circulating viruses w.r.t. Alpha targeting the same dates were 0.101, 0.113, and 0.099,
293 respectively (Supplementary Table S3). The possible reason for these overestimations is the
294 difference in submission delay between Delta and Alpha sequences. The submission delay of
295 nucleotide sequences belonging to Alpha had a mean of 11.05 days and a median of 10 days
296 during the period from the introduction of Delta (18th March 2021) to the date when Delta
297 reached a frequency of 0.50 (13th May 2021). The submission delay of Delta during the same

298 period had a mean of 10.44 days and a median of 9 days. This means that sequences of Delta
299 were submitted to the GISAID database more quickly than those of Alpha during the early stages
300 of the Alpha–Delta replacement. Sequencing prioritizing a new variant may have led to the
301 overestimations, and is the limitation of our real-time prediction using sequence databases. An
302 explicit inclusion of submission delays for each variant in the model may solve this problem for
303 real-time predictions. Use of data sources without variant prioritization, such as results of PCR
304 tests, can also solve this problem.

305
306 Our model assumes that there was no difference between the generation times of both variants
307 with a mean value of 4.64 days [17]. However, Hart et al. estimated the generation time of Delta
308 (4.7 days) to be shorter than that of Alpha (5.5 days) [21]. To allow differences between
309 generation times of variants, it is necessary to extend the model to also estimate the relative
310 generation times of the variant w.r.t. that of the baseline variant [22].

311 References

- 312
- 313 1 Tao, K. *et al.* The biological and clinical significance of emerging SARS-CoV-2 variants. *Nat*
314 *Rev Genet* **22**, 757-773, doi:10.1038/s41576-021-00408-x (2021).
 - 315 2 World Health Organization. *Tracking SARS-CoV-2 variants*,
316 <<https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>> (2022).
 - 317 3 Volz, E. *et al.* Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*
318 **593**, 266-269, doi:10.1038/s41586-021-03470-x (2021).
 - 319 4 Leung, K., Shum, M. H., Leung, G. M., Lam, T. T. & Wu, J. T. Early transmissibility assessment
320 of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November
321 2020. *Euro Surveill* **26**, doi:10.2807/1560-7917.ES.2020.26.1.2002106 (2021).
 - 322 5 Ito, K., Piantham, C. & Nishiura, H. Predicted dominance of variant Delta of SARS-CoV-2
323 before Tokyo Olympic Games, Japan, July 2021. *Euro Surveill* **26**, doi:10.2807/1560-
324 7917.ES.2021.26.27.2100570 (2021).
 - 325 6 Ito, K., Piantham, C. & Nishiura, H. Relative instantaneous reproduction number of Omicron
326 SARS-CoV-2 variant with respect to the Delta variant in Denmark. *J Med Virol*,
327 doi:10.1002/jmv.27560 (2021).
 - 328 7 Nishiura, H. *et al.* Relative reproduction number of SARS-CoV-2 Omicron (B.1.1.529) compared
329 with Delta variant in South Africa. *J Clin Med* **11**, doi:10.3390/jcm11010030 (2021).
 - 330 8 Morris, D. H. *et al.* Predictive modeling of influenza shows the promise of applied evolutionary
331 biology. *Trends Microbiol* **26**, 102-118, doi:10.1016/j.tim.2017.09.004 (2018).
 - 332 9 Łuksza, M. & Lässig, M. A predictive fitness model for influenza. *Nature* **507**, 57-61,
333 doi:10.1038/nature13087 (2014).
 - 334 10 Huddleston, J. *et al.* Integrating genotypes and phenotypes improves long-term forecasts of
335 seasonal influenza A/H3N2 evolution. *Elife* **9**, doi:10.7554/eLife.60067 (2020).

- 336 11 Piantham, C. & Ito, K. Modeling the selective advantage of new amino acids on the
337 hemagglutinin of H1N1 influenza viruses using their patient age distributions. *Virus Evolution* **7**,
338 doi:10.1093/ve/veab049 (2021).
- 339 12 Fraser, C. Estimating individual and household reproduction numbers in an emerging epidemic.
340 *PLoS One* **2**, e758, doi:10.1371/journal.pone.0000758 (2007).
- 341 13 Leung, K., Lipsitch, M., Yuen, K. Y. & Wu, J. T. Monitoring the fitness of antiviral-resistant
342 influenza strains during an epidemic: a mathematical modelling study. *Lancet Infect Dis* **17**, 339-
343 347, doi:10.1016/S1473-3099(16)30465-0 (2017).
- 344 14 Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data - from vision to
345 reality. *Euro Surveill* **22**, doi:10.2807/1560-7917.ES.2017.22.13.30494 (2017).
- 346 15 Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121-
347 4123, doi:10.1093/bioinformatics/bty407 (2018).
- 348 16 Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic
349 epidemiology. *Nat Microbiol* **5**, 1403-1407, doi:10.1038/s41564-020-0770-5 (2020).
- 350 17 Nishiura, H., Linton, N. M. & Akhmetzhanov, A. R. Serial interval of novel coronavirus
351 (COVID-19) infections. *Int J Infect Dis* **93**, 284-286, doi:10.1016/j.ijid.2020.02.060 (2020).
- 352 18 Pawitan, Y. *In all likelihood: Statistical modelling and inference using likelihood.* (Oxford
353 University Press, 2001).
- 354 19 Hansen, P. R. Relative contagiousness of emerging virus variants: An analysis of the Alpha,
355 Delta, and Omicron SARS-CoV-2 variants. *The Econometrics Journal*, doi:10.1093/ectj/utac011
356 (2022).
- 357 20 Figgins, M. D. & Bedford, T. SARS-CoV-2 variant dynamics across US states show consistent
358 differences in effective reproduction numbers. *medRxiv*, 2021.2012.2009.21267544,
359 doi:10.1101/2021.12.09.21267544 (2021).
- 360 21 Hart, W. S. *et al.* Generation time of the alpha and delta SARS-CoV-2 variants: an
361 epidemiological analysis. *Lancet Infect Dis*, doi:10.1016/S1473-3099(22)00001-9 (2022).
- 362 22 Ito, K., Piantham, C. & Nishiura, H. Estimating relative generation times and relative
363 reproduction numbers of Omicron BA.1 and BA.2 with respect to Delta in Denmark. *medRxiv*,
364 2022.2003.2002.22271767, doi:10.1101/2022.03.02.22271767 (2022).
- 365

366 Acknowledgement

367 We gratefully acknowledge the laboratories responsible for obtaining the specimens and the
368 laboratories where genetic sequence data were generated and shared via the GISAID Initiative,
369 on which this research is based. The information on originating laboratories, submitting
370 laboratories, and authors of SARS-CoV-2 sequence data can be found in Supplementary Table
371 S1. We thank Brad Suchoski, Heidi Gurung, and Prasith Baccam from IEM, Inc. in the United
372 States of America for their help converting our R program into the Julia language, which
373 produced an approximate 50x speedup of our computations. This work was supported by the
374 Japan Agency for Medical Research and Development (grant numbers JP20fk0108535,
375 JP21wm0125008). K.I. received funding JSPS KAKENHI (21H03490). C.P. was supported by
376 the World-leading Innovative and Smart Education Program (1801) from the Ministry of

377 Education, Culture, Sports, Science, and Technology, Japan. The funders had no role in the study
378 design, data collection and analysis, decision to publish, or preparation of the manuscript.

379 Author contributions

380 K.I. designed the study. C.P. collected data and conducted the analysis. C.P. and K.I. wrote the
381 paper.

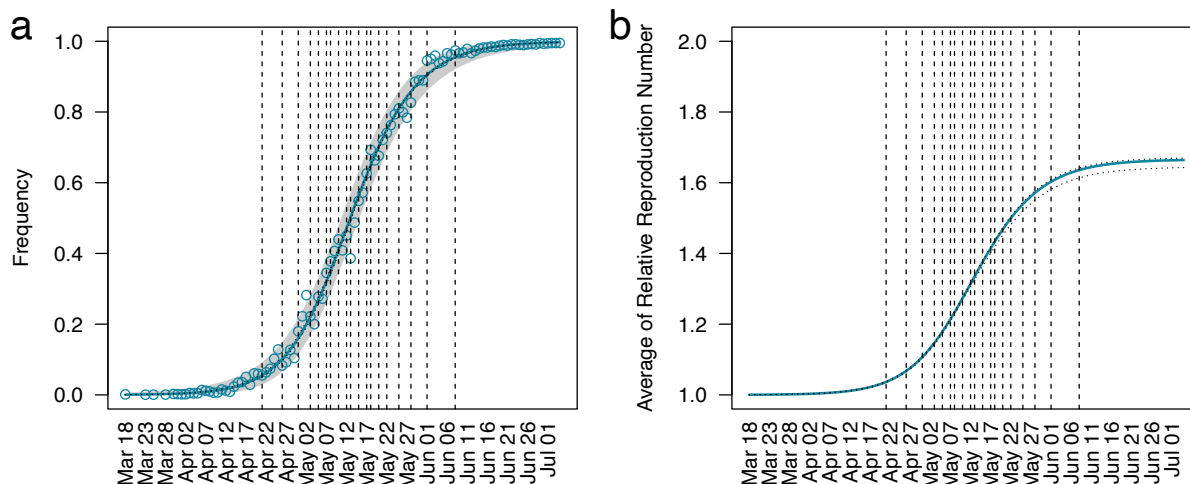
382 Data availability

383 Supplementary Table S1 contains metadata of all nucleotide sequences of SARS-CoV-2 viruses
384 collected from England during 1st January 2021 to 31st July 2021 used for the analysis.

385 Conflict of interest

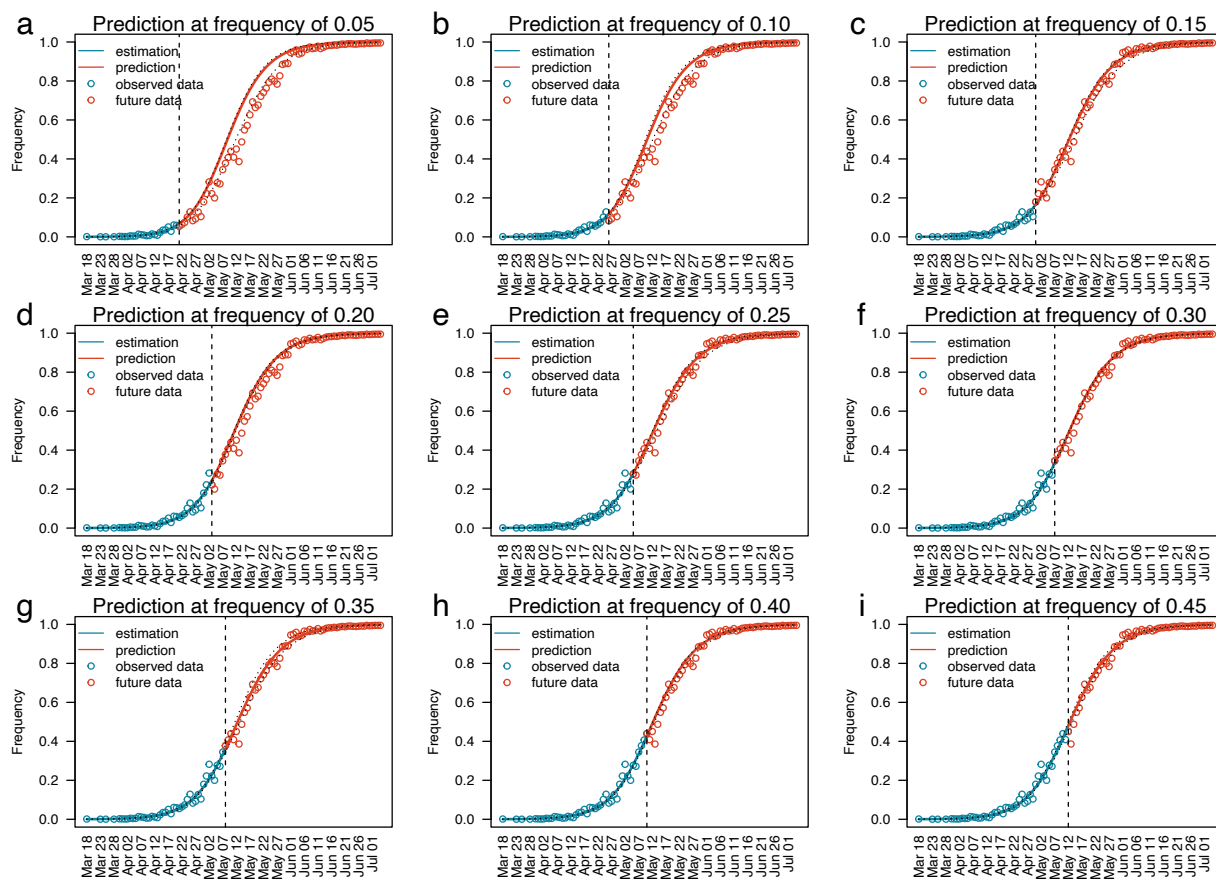
386 We declare that there is no conflict of interest.

387 Figures



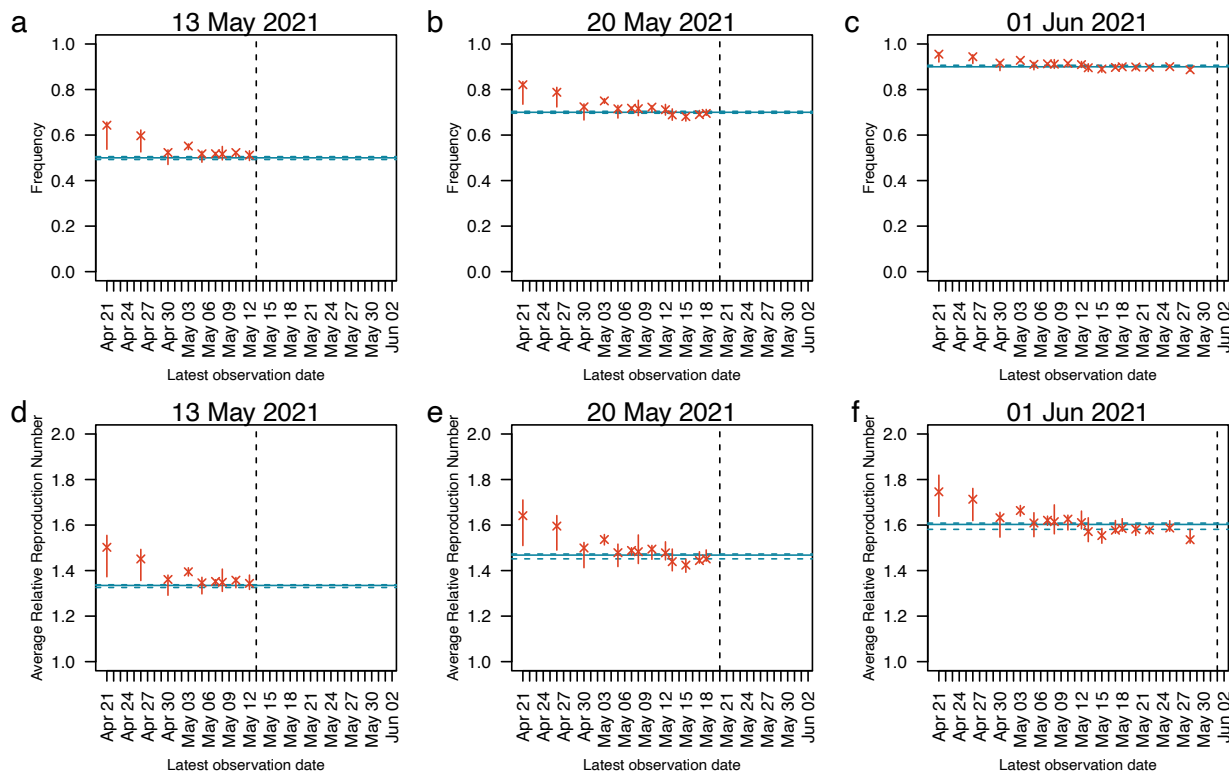
388
389 **Figure 1. Estimated frequencies of the Delta variant and average relative reproduction**
390 **number of circulating viruses w.r.t. Alpha during 18th March to 4th July 2021 using entire**
391 **observations.** In panel a, circles represent variant frequencies observed in the GISAID database.
392 The blue curve represents the maximum likelihood estimates of frequencies of Delta. Black
393 dotted curves around the blue curve represent 95% confidence intervals of the estimated
394 frequencies of Delta. Gray area represents the 95% equal-tailed interval of beta distribution for

395 the 95% confidence intervals of parameters of the estimated beta-binomial distribution. In panel
396 **b**, the blue curve and black dotted curves represents the maximum likelihood estimates and 95%
397 confidence intervals of the average relative reproduction number of circulating viruses with
398 respect to Alpha. Vertical dashed lines in both panels indicate the dates when the estimated
399 frequency reached 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70,
400 0.75, 0.80, 0.85, 0.90, and 0.95.
401



402
403 **Figure 2. Prediction of future frequencies of the Delta variant using partial observations.**
404 Panels **a, b, c, d, e, f, g, h,** and **i** represent predictions estimated using observations until 21st
405 April, 26th April, 30th April, 3rd May, 5th May, 7th May, 8th May, 10th May, and 12th May, 2021,
406 respectively. Blue circles represent observed frequencies used for predictions. Red circles
407 represent future observations that were not used for predictions. The blue curve in each panel
408 represents the maximum likelihood estimates of frequencies, and the red curve represents the
409 frequencies predicted using the estimated parameter. Black dotted curves represent 95%

410 confidence intervals of the estimated frequencies of Delta. The vertical dashed line in each panel
411 represents the date of the last observations used for prediction.
412



413
414 **Figure 3. Predictions of population average of relative reproduction number with respect**
415 **to Alpha.** In each panel, x-axis represents dates until which observations were used in the
416 prediction. Y-axis represents the predicted population average of reproduction number for the
417 date marked by vertical dashed lines (13th May in **a**, 20th May in **b**, and 1st June in **c**). Cross
418 marks represent predicted population average of relative reproduction number with respect to
419 Alpha and vertical bars represent their 95% confidence intervals. The blue horizontal solid lines
420 represent the maximum likelihood estimates using the entire observations, we call the final
421 estimates. The horizontal dashed blue lines represent 95% confidence intervals of the final
422 estimates.

423

424 **Tables**

425 **Table 1. Parameters estimated using the entire observations during the Alpha–Delta**
 426 **replacement**

k (95% CI)	$q_Y(t_Y)$ (95% CI)	M (95% CI)	Log Likelihood
1.67 (1.64, 1.67)	0.0005 (0.0004, 0.0006)	286.82 (201.83, 403.22)	−430.97

427

428 **Table 2. Maximum likelihood estimations for dates when Delta exceeded certain**
 429 **frequencies and the average relative reproduction numbers w.r.t. Alpha on those dates.**

Frequency	Date when exceeded the frequency (95% CI)	Average relative reproduction number w.r.t. Alpha (95% CI)
0.05	2021-04-21 (2021-04-20, 2021-04-21)	1.037 (1.035, 1.039)
0.10	2021-04-26 (2021-04-26, 2021-04-27)	1.068 (1.065, 1.070)
0.15	2021-04-30 (2021-04-29, 2021-04-30)	1.107 (1.103, 1.109)
0.20	2021-05-03 (2021-05-02, 2021-05-03)	1.147 (1.142, 1.147)
0.25	2021-05-05 (2021-05-04, 2021-05-05)	1.179 (1.174, 1.180)
0.30	2021-05-07 (2021-05-06, 2021-05-07)	1.214 (1.209, 1.216)
0.35	2021-05-08 (2021-05-08, 2021-05-09)	1.233 (1.228, 1.235)
0.40	2021-05-10 (2021-05-10, 2021-05-11)	1.273 (1.267, 1.275)
0.45	2021-05-12 (2021-05-12, 2021-05-12)	1.314 (1.306, 1.317)
0.50	2021-05-13 (2021-05-13, 2021-05-14)	1.335 (1.326, 1.338)
0.55	2021-05-15 (2021-05-15, 2021-05-15)	1.376 (1.364, 1.379)
0.60	2021-05-17 (2021-05-17, 2021-05-17)	1.415 (1.401, 1.419)
0.65	2021-05-18 (2021-05-18, 2021-05-19)	1.433 (1.419, 1.438)
0.70	2021-05-20 (2021-05-20, 2021-05-21)	1.468 (1.452, 1.473)
0.75	2021-05-22 (2021-05-22, 2021-05-23)	1.499 (1.481, 1.505)
0.80	2021-05-25 (2021-05-25, 2021-05-25)	1.539 (1.520, 1.545)
0.85	2021-05-28 (2021-05-28, 2021-05-28)	1.571 (1.550, 1.577)
0.90	2021-06-01 (2021-06-01, 2021-06-01)	1.603 (1.581, 1.609)
0.95	2021-06-08 (2021-06-07, 2021-06-08)	1.636 (1.614, 1.642)

430

431 **Table 3. Parameters estimated using partial observations**

Observed frequency	k	q_0	M	Log Likelihood
0.05	1.78 (1.69, 1.85)	0.0003 (0.0002, 0.0004)	611.41 (268.84, 2000.00 [†])	-82.09
0.10	1.76 (1.68, 1.80)	0.0003 (0.0002, 0.0004)	559.33 (264.55, 1476.99)	-102.85
0.15	1.69 (1.62, 1.71)	0.0004 (0.0003, 0.0006)	395.63 (202.88, 842.16)	-123.12
0.20	1.72 (1.69, 1.73)	0.0003 (0.0003, 0.0004)	359.75 (189.17, 725.97)	-137.61
0.25	1.67 (1.62, 1.71)	0.0005 (0.0003, 0.0007)	283.92 (160.53, 603.44)	-148.98
0.30	1.68 (1.66, 1.70)	0.0004 (0.0003, 0.0005)	313.07 (171.38, 579.86)	-157.87
0.35	1.67 (1.63, 1.74)	0.0004 (0.0003, 0.0006)	311.65 (175.27, 602.30)	-161.95
0.40	1.68 (1.64, 1.70)	0.0004 (0.0004, 0.0006)	335.32 (189.19, 637.92)	-170.37
0.45	1.67 (1.65, 1.72)	0.0004 (0.0003, 0.0005)	319.57 (181.41, 575.63)	-180.60
0.50	1.64 (1.60, 1.69)	0.0006 (0.0004, 0.0008)	230.02 (137.25, 411.11)	-190.09
0.55	1.62 (1.59, 1.65)	0.0007 (0.0005, 0.0008)	218.52 (133.36, 439.22)	-199.09
0.60	1.64 (1.63, 1.68)	0.0005 (0.0004, 0.0006)	254.02 (150.72, 429.33)	-207.23
0.65	1.65 (1.64, 1.69)	0.0005 (0.0004, 0.0006)	245.72 (148.15, 413.85)	-213.00
0.70	1.65 (1.63, 1.67)	0.0005 (0.0004, 0.0006)	256.17 (154.91, 426.78)	-222.19
0.75	1.64 (1.63, 1.66)	0.0005 (0.0005, 0.0007)	264.58 (162.71, 451.18)	-230.71
0.80	1.65 (1.64, 1.68)	0.0005 (0.0004, 0.0006)	291.69 (177.99, 484.83)	-244.13
0.85	1.60 (1.59, 1.65)	0.0008 (0.0005, 0.0009)	202.38 (119.43, 612.16)	-264.29
0.90	1.65 (1.63, 1.67)	0.0005 (0.0004, 0.0006)	236.54 (155.20, 372.12)	-284.01
0.95	1.66 (1.65, 1.68)	0.0005 (0.0004, 0.0006)	201.41 (138.15, 306.96)	-328.33

432 †The upper bound of M in the maximum likelihood estimation was set to 2,000.

433

434 **Table 4. Errors of predictions made at frequencies less than 0.25**

Target date	Number of predictions	Errors in predicted frequency		Errors in predicted average relative reproduction number	
		Median	Maximum absolute error	Median	Maximum absolute error
13 th May 2021	4	0.074	0.143	0.088	0.168
20 th May 2021	4	0.069	0.121	0.098	0.173
1 st June 2021	4	0.036	0.055	0.086	0.143

435

436 **Table 5. Errors of predictions made at frequencies greater than or equal to 0.25**

Target date	Number of predictions	Errors in predicted frequency		Errors in predicted average relative reproduction number	
		Median	Maximum absolute error	Median	Maximum absolute error
13 th May 2021	5	0.018	0.023	0.014	0.023
20 th May 2021	9	0.011	0.023	0.010	0.044
1 st June 2021	13	0.000	0.015	-0.019	0.067

437

438