Estimating the increased transmissibility of the B.1.1.7 strain over previously circulating strains in England using frequencies of GISAID sequences and the distribution of serial intervals

Chayada Piantham[1], Kimihito Ito[2*],

[1]*Graduate School of Infectious Diseases, Hokkaido University; chayada@czc.hokudai.ac.jp*
[2]*Research Center for Zoonosis Control, Hokkaido University; itok@czc.hokudai.ac.jp*

Abstract

The B.1.1.7 strain, a variant strain of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is thought to have higher transmissibility than previously circulating strains in England. This paper proposes a method to estimate the selective advantage of a mutant strain over previously circulating strains using the time course of B.1.1.7 strain frequencies and the distribution of serial intervals. The method allows instantaneous reproduction numbers of infections to change during the target period of analysis. The proposed method, instead, assumes that the selective advantage of a mutant strain over previously circulating strains is constant over time. Applying this method to the sequence data of SARS-CoV-2 in England, the instantaneous reproduction number of the B.1.1.7 strain is estimated to be 33.7% (with a 95% confidence interval (CI) from 33.6% to 33.9%) higher than previously circulating strains in England, assuming that serial intervals of COVID-19 follow a lognormal distribution with a mean of 4.7 days and a standard deviation of 2.9 days. The result indicated that the control measure against B.1.1.7 strain needs to be strengthened by 33.7% from that against previously circulating strains. To get the same control effect as before, contact rates between individuals needed to be restricted to 0.748 of the contact rates that had been achieved by the control measure taken for previously circulating strains.

Keywords: COVID-19, B.1.1.7, selective advantage, adaptive evolution, serial interval, GISAID, England, SARS-CoV-2, instantaneous reproduction number

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of COVID-19, has been rapidly evolving since its emergence in 2019. In December 2020, the Public Health England detected a new cluster of SARS-CoV-2 viruses phylogenetically distinct from the other strains circulating in the United Kingdom (Chand et al. 2020). These viruses belong to the lineage B.1.1.7 according to the PANGO nomenclature (Rambaut et al. 2020b). The World Health Organization (WHO) designated them as Variant of Concern, year 2020, month 12, variant 01 (VOC 202012/01) (World Health Organization 2020).

The B.1.1.7 strain was first detected in England in September 2020, and the numbers of infections with this strain increased in October and November in 2020 (Chand et al. 2020). By February 2021, B.1.1.7 strain occupied 95% of the SARS-CoV-2 circulating in England (Davies et al., 2021).

Several studies have compared the transmissibility of the B.1.1.7 strain to that of previously circulating strains. Davies et al. estimated that the reproduction number of the B.1.1.7 strain is 43–90% (with a 95% credible interval of 38–130%) higher than preexisting strains using data from England (Davies et al., 2021). However, different models resulted in different ranges of estimates in their multiplicative increase in reproduction number (R). Grabowski et al. estimated a 83–118% increase with a confidence interval of 71–140% compared to previously circulating strains in England (Grabowski et al. 2021). Volz et al. estimated 40–75% increase in R using data from England (Volz et al. 2021). Their methods use the linear regression of log odds ratio between B.1.1.7 strains and previously circulating strains and estimate increase in transmissibility under the assumption that R is constant over time during the target period of analysis. Washington et al. estimated a 35–45% increase using data from the United States of America using Volz's method (Washington et al., 2021). Chen et al. estimated 49–65% increase in reproduction number using data from Switzerland also under the constant R assumption (Chen et al., 2021). Due to the high transmissibility the B.1.1.7 strain, strong control measures such as lockdown were taken when the strain was introduced (Davies et al., 2021). Thus, the constant R assumption is questionable when analyzing the increase in the reproduction number of B.1.1.7 compared to that of previously circulating strains.

In this paper, we propose a method to estimate the selective advantage of a mutant strain over previously circulating strains. Based on Fraser's method to estimate instantaneous reproduction numbers (Fraser 2007), our method allows reproduction numbers to change during the target period of analysis. Our approach is also based on the Maynard Smith's model of allele frequencies in adaptive evolution, which assumes that the selective advantage of a mutant strain over previously circulating strains is constant over time (Maynard Smith and Haigh 1974). Applying the developed method to the sequence data in England using the serial interval distribution of COVID-19 estimated by Nishiura et al. (Nishiura et al., 2020), we estimate the increase in the instantaneous reproduction number of B.1.1.7 strains compare to that of previously circulating strains in England. Based on the estimate, we discuss its implication to control measures for COVID-19.

Materials and Methods

Sequence data

Nucleotide sequences of SARS-CoV-2 viruses were downloaded from the GISAID EpiCoV database (Shu & McCauley, 2017) on March 1, 2021. Nucleotide sequences of viruses detected in England were selected and aligned to the reference amino acid sequence of the spike protein of SARS-CoV-2 virus (YP_009724390) using DIAMOND (Buchfink et al. 2015). The aligned nucleotide sequences were translated into amino acid sequences, then were aligned with the reference amino acid sequence using MAFFT (Katoh et al. 2002). Amino acid sequences having either an ambiguous amino acid or more than ten gaps were excluded from the rest of analyses. Table 1 shows amino acids on the spike protein used to characterize B.1.1.7 strain, retrieved from the PANGO database (Rambaut et al. 2020b).

Table 1. Amino acids on the spike protein which are used to define B.1.1.7 strains

Position on S protein	Amino acid
69	Deletion
70	Deletion
144	Deletion
501	Tyrosine (Y)
570	Aspartic acid (D)
681	Histidine (H)
716	Isoleucine (I)
982	Alanine (A)
1118	Histidine (H)

We divided amino acid sequences into three groups based on amino acids shown in Table 1. The first group consists of sequences having all of B.1.1.7-defining amino acids in Table 1. We call a virus in this group a “B.1.1.7 strain”. The second group contains sequences that have none of the B.1.1.7-defining substitutions. We call a virus in this group a “non-B.1.1.7 strain”. The third group is a set of sequences that have at least one but incomplete set of the B.1.1.7-defining amino acids. We call a strain in the third group a “B.1.1.7-like strain”. Table 2 shows the number of sequences categorized into each group. Figure 1 shows the daily numbers of GISAID sequences of B.1.1.7 strains, non-B.1.1.7 strains and B.1.1.7-like strains detected in England from September 1, 2020 to February 19, 2021. We used the number of B.1.1.7 strains and non-B.1.1.7 strains for the rest of the analyses. B.1.1.7-like strains were excluded from the analyses, since we do not know whether they had the same transmissibility as B.1.1.7 strains or not. These numbers are provided in Supplementary Table 1.

Table 2. Number of GISAID sequences in England from September 1, 2020 to February 19, 2021

B.1.1.7 strains	Non-B.1.1.7	B.1.1.7-like	Total
71692	65840	2227	139759

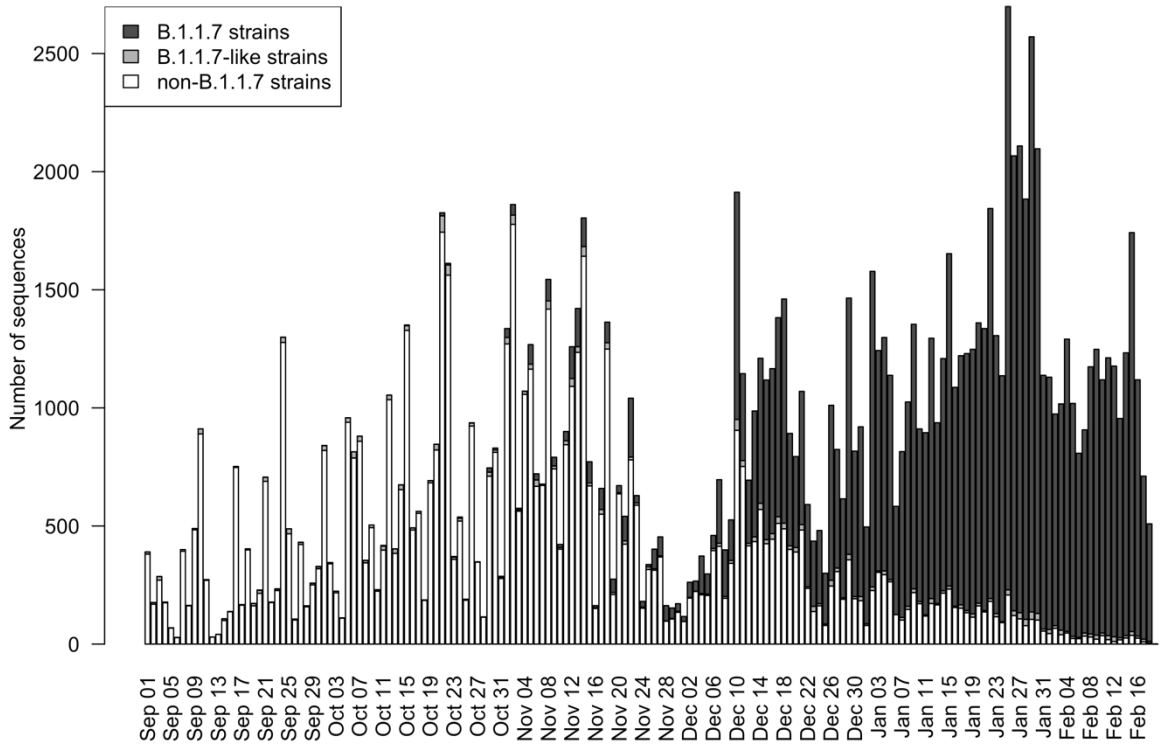

Figure 1. Numbers of nucleotide sequences of B.1.1.7 strains, B.1.1.7-like strains, and non-B.1.1.7 strains in England from September 1, 2020 to February 19, 2021, based on sequences retrieved from the GISAID database on March 1, 2020.

Serial interval distribution

The serial interval is the time from illness onset in a primary case to illness onset in a secondary case (Nishiura et al., 2020). The method we propose in this paper uses discrete distributions of serial intervals. Let $g(i)$ denote the probability that the serial interval of a transmission is i days. We obtained the values of $g(i)$ by discretizing the lognormal distribution of serial intervals of COVID-19 estimated by Nishiura et al. (Nishiura et al., 2020). Thus, the probability mass function of serial intervals for $i \geq 0$ is given by

$$g(i) = \int_{i}^{i+1} f(t)dt\,,$$

where $f(t)$ is the probability density function of a lognormal distribution with a mean of 4.7 days and a standard deviation of 2.9 days.

Model of Advantageous Selection

Let us suppose that we have a large population of viruses consisting of strains of two genotypes A and a, of which frequency in the viral population at a calendar date t are $q_A(t)$ and $q_a(t)$, respectively. Suppose also that genotype A is a mutant of a that emerged at time t_0.

We assume that a virus of genotype A generates $1+s$ times as many secondary transmissions as those of genotype a. Then, s can be considered as the coefficient of selective advantage in adaptive evolution. As described in Maynard Smith and Haigh (1974), the frequency of viruses of allele A after n transmissions, q_n, satisfies the following equation:

$$q_{n+1} = \frac{(1+s)q_n}{(1+s)q_n + (1-q_n)} = \frac{1+s}{1+sq_n} q_n \,. \tag{1}$$

Maynard Smith's formulation of allele frequency can be extended using the concept of instantaneous reproduction numbers of infectious diseases. The instantaneous reproduction number is defined as the average number of people someone infected at time t could expect to infect given that conditions remain unchanged (Fraser 2007). Let $I(t)$ be the number of infections by viruses of either genotype A or a at calendar time t and $g(i)$ be the probability mass function of serial intervals, defined in the previous subsection. Suppose that instantaneous reproduction numbers of genotypes A and a at calendar time t are $R_A(t)$ and $R_a(t)$, respectively. Assuming that the distribution of generation time of a disease can be approximated by the serial interval distribution, the following equations give the discrete version of Fraser's instantaneous reproduction numbers of infections by genotype A and a at time t.

$$R_A(t) = \frac{q_A(t)I(t)}{\sum_{i=0}^{t} g(i)q_A(t-i)I(t-i)} \tag{2}$$

$$R_a(t) = \frac{q_a(t)I(t)}{\sum_{i=0}^{t} g(i)q_a(t-i)I(t-i)} \tag{3}$$

Suppose that $g(i)$ is small enough to be neglected for $i < 1$ and $i > l$, then $g(i)$ can be truncated and the above formula can be treated as follows.

$$R_A(t) = \frac{q_A(t)I(t)}{\sum_{i=1}^{l} g(i)q_A(t-i)I(t-i)} \tag{2}$$

$$R_a(t) = \frac{q_a(t)I(t)}{\sum_{i=1}^{l} g(i)q_a(t-i)I(t-i)} \tag{3}$$

Since a virus of genotype A generates $1+s$ times as many secondary transmissions as those of genotype a, the following equation holds

$$R_A(t) = (1+s)R_a(t) \tag{4}$$

for each calendar time $t \geq t_0$. Next we assume that for all infections at calendar time t, the difference in the number of infections at the time when previous generations infected can be regarded as considerably small, i.e.

$$I(t-1) \approx I(t-2) \approx \cdots \approx I(t-l)\,. \tag{5}$$

The frequency of genotype A in the viral population at calendar time t, $q_A(t)$, can be modeled as follows:

$$\begin{aligned} q_A(t) &= \frac{q_A(t)I(t)}{q_a(t)I(t) + q_A(t)I(t)} \\ &= \frac{\sum_{i=1}^{l} g(i)\, R_A(t) q_A(t-i) I(t-i)}{\sum_{i=1}^{l} g(i) R_a(t) q_a(t-i) I(t-i) + \sum_{i=1}^{l} g(i) R_A(t) q_A(t-i)\, I(t-i)} \\ &= \frac{\sum_{i=1}^{l} g(i)\,(1+s) R_a(t) q_A(t-i)}{\sum_{i=1}^{l} g(i) R_a(t)(1-q_A(t-i)) + \sum_{i=1}^{l} g(i)(1+s) R_a(t) q_A(t-i)} \\ &= \frac{\sum_{i=1}^{l} g(i)\,(1+s) q_A(t-i)}{\sum_{i=1}^{l} g(i)(1-q_A(t-i)) + \sum_{i=1}^{l} g(i)(1+s) q_A(t-i)} \\ &= \frac{\sum_{i=1}^{l} g(i)\,(1+s) q_A(t-i)}{1 + s\sum_{i=1}^{l} g(i) q_A(t-i)}. \end{aligned} \tag{6}$$

Likelihood Function

Let $n(t)$ be the number of sequences of either genotype A or a observed at calendar date t. Let $d_1, \dots, d_k$ be calendar dates such that $n(d_i) > 0$ for $1 \le j \le k$. Suppose that we have $n_A(d_j)$ sequences of genotype A at calendar date d_j. Since genotype A emerged at time t_0, $q_A(d_j) = 0$ *and* $q_a(d_j) = 1$ for $d_j < t_0$. If the frequency of genotype A is q_0 at t_0, then the following equation gives the likelihood function of s, t_0, and q_0 for observing $n_A(d_j)$ sequences of genotype A at calendar date d_j:

$$L\left(s, t_0, q_0\,;\, n_A(d_j)\right) = \binom{n(d_j)}{n_A(d_j)} q_A(d_j)^{n_A(d_j)} \left(1 - q_A(d_j)\right)^{n(d_j) - n_A(d_j)}, \tag{7}$$

for $1 \le j \le k$. The likelihood function of s, t_0, and q_0 for observing $n_A(d_1), \dots, n_A(d_k)$ sequences of genotype A at calendar dates $d_1, \dots, d_k$ is given by the following formula.

$$L\left(s, t_0, q_A(t_0)\,; n_A(d_1), \dots, n_A(d_k)\right) = \prod_{j=1}^{k} L\left(s, t_0, q_A(t_0)\,; n_A(d_j)\right) \tag{8}$$

Parameter estimation from sequence data

The B.1.1.7 strain was first detected in England on September 20, 2020. We assume that t_0 is this day or someday before this day. Parameters s, t_0, and q_0 were estimated by maximizing likelihood of observations on September 1, 2020 and later on. B.1.1.7 strains, viruses having complete subset of B.1.1.7-difining substitutions on its spike protein were considered as genotype A. The non-B.1.1.7 strains, viruses having none of B.1.1.7-defining substitutions were considered to be genotype a. The B.1.1.7-like strains, viruses having an incomplete set of B.1.1.7 substitutions on the spike protein, were excluded from

the analysis. We truncated the distribution of serial intervals so that $g(i) = 0$ if $i < 1$ or $i > 14$ and normalized $g(i)$ to ensure that $\sum_{i=1}^{14} g(i) = 1$. Parameters of s, t_0, and q_0 were estimated by maximizing the likelihood defined in Equation (8). The 95% confidence intervals of parameters were estimated by profile likelihood (Pawitan 2013). The optimization of likelihood function was done by the nloptr package in R (Rowan 1990; Johnson 2020).

Results

The selective advantage of B.1.1.7 strains over non-B.1.1.7 strains, s, was estimated to be 0.337 with its 95% confidence intervals from 0.336 to 0.339 (Table 2). The date when a B.1.1.7 virus have emerged in England, t_0, was estimated to be September 20, 2020 with its 95% confidence interval from September 17, 2020 to September 20, 2020. The initial frequency of B.1.1.7 among non-B.1.1.7 and B.1.1.7 strains at the emergence in England, q_0, was estimated to be 0.00574 with its 95% confidence intervals from 0.00443 to 0.00618.

Table 2. Maximum likelihood estimations of parameters

Parameter	Estimate	95% CI
s	0.337	[0.336, 0.339]
t_0	September 20, 2020	[September 17, 2020, September 20, 2020]
q_0	0.00574	[0.00443, 0.00618]

Figure 2 shows the time course of the frequency of B.1.1.7 strains among all strains except B.1.1.7-like strains detected in England from September 1, 2020 to February 19, 2021. White circles indicate daily frequencies of B.1.1.7 strains among all strains except B.1.1.7-like strains. Solid line indicates the time course of frequency of B.1.1.7 strains calculated using parameters estimated from the data. Dashed lines indicate its lower and upper bounds of its 95% CI.

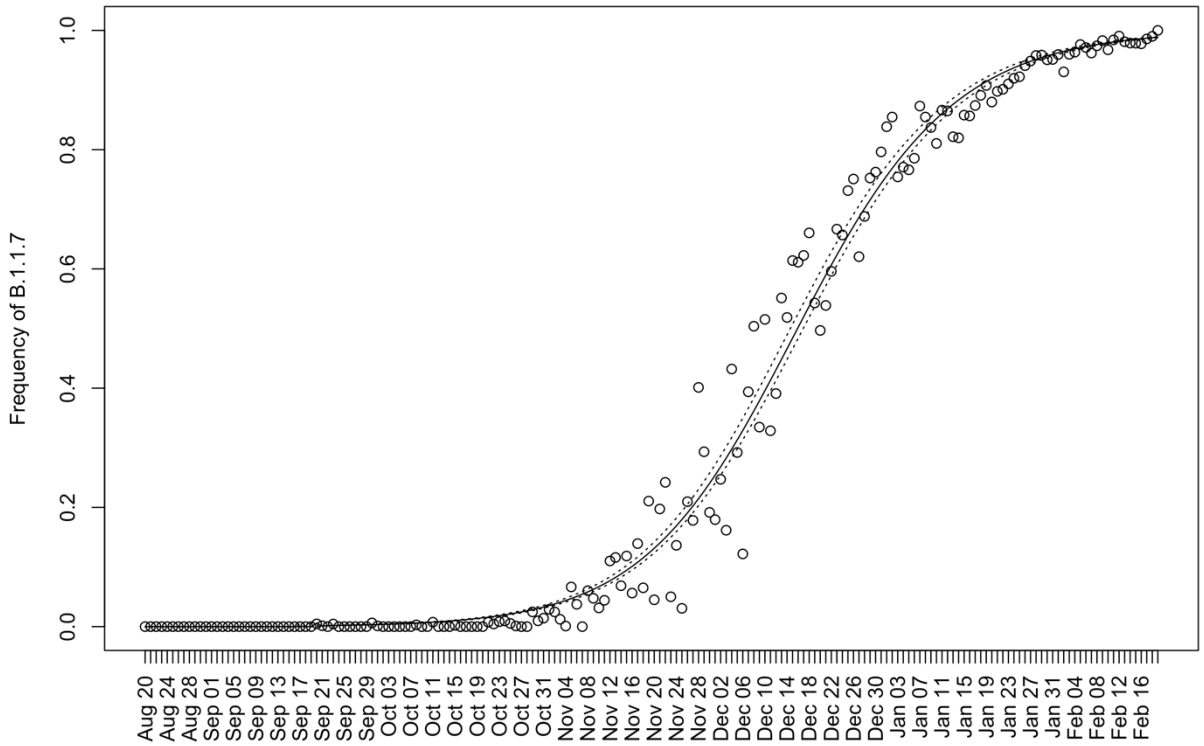

Figure 2. Time course of the frequency of B.1.1.7 strains among all strains except B.1.1.7-like strains detected in England from September 1, 2020 to February 19, 2021. White circles indicate the frequency of B.1.1.7 strains among B.1.1.7 and non-B.1.1.7 strains. The nucleotide sequences were retrieved from GISAID on March 1, 2021. Solid line indicates the time course of frequency of B.1.1.7 strains calculated using parameters estimated from the data. Dashed lines indicate its lower and upper bounds of its 95% CI.

Discussion

In this paper, the selective advantage of the B.1.1.7 strain over non-B.1.1.7 strains in England was estimated to be 0.337 with a 95% CI from 0.336 to 0.339. The date of emergence of B.1.1.7 strains in England was estimated to be September 20, 2020 with its 95% confidence interval from September 17, 2020 to September 20, 2020. The initial frequency of B.1.1.7 among all sequences except B.1.1.7-like strains at the time of emergence in England was estimated to be 0.00574 with its 95% confidence intervals from 0.00443 to 0.00618.

The estimated selective advantage of 0.337 indicates that the B.1.1.7 strain has 33.7% higher instantaneous reproduction numbers than those of previously circulating strains in England. This means that control measures for the B.1.1.7 strain needed to be strengthened by 33.7% compared to that for the previously circulating strains. To get the same control effect as for the previously circulating strains, the

control measure for the B.1.1.7 needed to restrict contact rates between individuals to $1/1.337 = 0.748$ of those that had been achieved by the control measure taken for the previously circulating strains.

Our estimation method is based on the principle that the expected frequency of a mutant strain among all strains can be determined from those in the previous generation using the serial interval distribution of infections. The method assumes that the selective advantage of a mutant strain over previously circulating strains is constant over time, which is based on Maynard Smith's model of allele frequencies in adaptive evolution. Based on Fraser's method for estimating the instantaneous reproduction number, our method allows reproduction numbers of strains to change during the target period of analysis. Thus, the proposed method removes the assumption that the reproduction number is constant over time, which is assumed in previous studies. Our method can estimate selective advantage of viruses in a genotype over the other genotype without estimating the reproduction numbers of viruses of each genotype. Thus the method can be applicable for the analysis on the selection of new variants even when strong control measures such as lockdown were introduced during the target period of analysis. We think this is the largest contribution of this paper to the field of molecular evolution, population genetics, and infectious disease epidemiology.

The selective advantage estimated in this study relies on the serial interval distribution, and thus the results may change depending on the serial interval distribution used in the analysis. In this paper, we used the serial interval distribution estimated by Nishiura et al. (Nishiura et al., 2020). The distribution is a lognormal distribution with a mean serial interval of 4.7 days with a standard deviation of 2.9 days. Several groups have estimated the serial intervals of SARS-CoV-2 using different datasets. Some variations are observed among these estimated values (Rai et al. 2021). Volz et al. (Volz et al. 2021) assumes a fixed serial interval of 6.5 days based on the results by Bi et al. (Bi et al., 2020). However, Ali et al. have reported that the serial interval estimated using data from China before January 22, 2020 was longer than estimates after January 22, 2020 (Ali et al., 2020). The serial interval estimated by Bi et al. contains data before January 22, 2020 and there might be some possibility that the estimated serial interval does not reflect the current situation. This important limitation will be removed when the serial interval distributions is analyzed using a large dataset of COVID-19 transmissions in the world.

As of March 17, 2021, the B.1.1.7 strain has been observed in 93 countries (Rambaut, Holmes, et al., 2020). The selective advantage of the B.1.1.7 strains over previously circulating strains in other countries remains to be our future work. Variant strains originated in Brazil and South Africa also show higher transmissibility than previously circulating strains (World Health Organization 2021). There is an urgent need to estimate the selective advantage of these strains over previously circulating strains. We hope that the methodology developed in this paper provides important information for countries in the world to establish control measures against highly transmissible variants strains such as the B.1.1.7 strain.

Acknowledgement

We gratefully acknowledge the laboratories responsible for obtaining the specimens and the laboratories where genetic sequence data were generated and shared via the GISAID Initiative, on which this research is based. This work was supported by CREST (grant number JPMJCR1413) from Japan Science and Technology Agency (http://www.jst.go.jp/), and the World-leading Innovative and Smart Education (WISE) Program (1801) from the Ministry of Education, Culture, Sports, Science, and Technology, Japan

(http://www.mext.go.jp/). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest

We declare that there is no conflict of interest.

References

1. Ali, S. T., Wang, L., Lau, E. H. Y., Xu, X. K., Du, Z., Wu, Y., Leung, G. M., & Cowling, B. J. (2020). Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. *Science, 369*(6507), 1106-1109. doi:10.1126/science.abc9004
2. Bi, Q., Wu, Y., Mei, S., Ye, C., Zou, X., Zhang, Z., Liu, X., Wei, L., Truelove, S. A., Zhang, T., Gao, W., Cheng, C., Tang, X., Wu, X., Wu, Y., Sun, B., Huang, S., Sun, Y., Zhang, J., Ma, T., Lessler, J., & Feng, T. (2020). Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect Dis, 20*(8), 911-919. doi:10.1016/S1473-3099(20)30287-5
3. Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat Methods, 12*(1), 59-60. doi:10.1038/nmeth.3176
4. Chand, M., Hopkins, S., Dabrera, G., Achison, C., Barclay, W., Ferguson, N., Volz, E., Loman, N., Rambaut, A., & Barrett, J. (2020). *Investigation of novel SARS-COV-2 variant. Variant of Concern 202012/01* (PHE gateway number: GW-1824). Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959438/Technical_Briefing_VOC_SH_NJL2_SH2.pdf
5. Chen, C., Nadeau, S., Topolsky, I., Manceau, M., Huisman, J. S., Jablonski, K. P., Fuhrmann, L., Dreifuss, D., Jahn, K., Beckmann, C., Redondo, M., Kobel, O., Noppen, C., Risch, L., Risch, M., Wohlwend, N., Kas, S., Bodmer, T., Roloff, T., Stange, M., Egli, A., Eckerle, I., Denes, R., Feldkamp, M., Nissen, I., Santacroce, N., Burcklen, E., Aquino, C., de Gouvea, A. C., Moccia, M. D., Grüter, S., Sykes, T., Opitz, L., White, G., Neff, L., Popovic, D., Patrignani, A., Tracy, J., Schlapbach, R., Dermitzakis, E. T., Harshman, K., Xenarios, I., Pegeot, H., Cerutti, L., Penet, D., Blin, A., Elies, M., Althaus, C., Beisel, C., Beerenwinkel, N., Ackermann, M., & Stadler, T. (2021). Quantification of the spread of SARS-CoV-2 variant B.1.1.7 in Switzerland. *medRxiv*, 2021.2003.2005.21252520. doi:10.1101/2021.03.05.21252520
6. Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., Pearson, C. A. B., Russell, T. W., Tully, D. C., Washburne, A. D., Wenseleers, T., Gimma, A., Waites, W., Wong, K. L. M., van Zandvoort, K., Silverman, J. D., Group, C. C.-W., Consortium, C.-G. U., Diaz-Ordaz, K., Keogh, R., Eggo, R. M., Funk, S., Jit, M., Atkins, K. E., & Edmunds, W. J. (2021). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*. doi:10.1126/science.abg3055

7. Fraser, C. (2007). Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS One, 2*(8), e758. doi:10.1371/journal.pone.0000758
8. Grabowski, F., Preibisch, G., Giziński, S., Kochańczyk, M., & Lipniacki, T. (2021). SARS-CoV-2 Variant of Concern 202012/01 Has about Twofold Replicative Advantage and Acquires Concerning Mutations. *Viruses, 13*(3), 392. Retrieved from https://www.mdpi.com/1999-4915/13/3/392
9. Johnson, S. G. The NLopt nonlinear-optimization package. Retrieved from http://ab-initio.mit.edu/nlopt
10. Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res, 30*(14), 3059-3066. doi:10.1093/nar/gkf436
11. Maynard Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genet Res, 23*(1), 23-35. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/4407212
12. Nishiura, H., Linton, N. M., & Akhmetzhanov, A. R. (2020). Serial interval of novel coronavirus (COVID-19) infections. *Int J Infect Dis, 93*, 284-286. doi:10.1016/j.ijid.2020.02.060
13. Pawitan, Y. (2013). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Croydon: Oxford University Press.
14. Rai, B., Shukla, A., & Dwivedi, L. K. (2021). Estimates of serial interval for COVID-19: A systematic review and meta-analysis. *Clin Epidemiol Glob Health, 9*, 157-161. doi:10.1016/j.cegh.2020.08.007
15. Rambaut, A., Holmes, E. C., O'Toole, A., Hill, V., McCrone, J. T., Ruis, C., du Plessis, L., & Pybus, O. G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol, 5*(11), 1403-1407. doi:10.1038/s41564-020-0770-5
16. Rambaut, A., Loman, N., Pybus, O., Barclay, W., Barrett, J., Carabelli, A., Connor, T., Peacock, T., Robertson, D. L., & Volz, E. (2020, 5 February 2021). Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. Retrieved from https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563
17. Rowan, T. (1990). *Functional Stability Analysis of Numerical Algorithms.* (Ph.D. Thesis). University of Texas, Austin.
18. Shu, Y., & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill, 22*(13). doi:10.2807/1560-7917.ES.2017.22.13.30494
19. Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg, L., Hinsley, W. R., Laydon, D. J., Dabrera, G., O'Toole, Á., Amato, R., Ragonnet-Cronin, M., Harrison, I., Jackson, B., Ariani, C. V., Boyd, O., Loman, N. J., McCrone, J. T., Gonçalves, S., Jorgensen, D., Myers, R., Hill, V., Jackson, D. K., Gaythorpe, K., Groves, N., Sillitoe, J., Kwiatkowski, D. P., Flaxman, S., Ratmann, O., Bhatt, S., Hopkins, S., Gandy, A., Rambaut, A., & Ferguson, N. M. (2021). Transmission of SARS-CoV-2 Lineage B.1.1.7

in England: Insights from linking epidemiological and genetic data. *medRxiv*, 2020.2012.2030.20249034. doi:10.1101/2020.12.30.20249034

20. Washington, N. L., Gangavarapu, K., Zeller, M., Bolze, A., Cirulli, E. T., Schiabor Barrett, K. M., Larsen, B. B., Anderson, C., White, S., Cassens, T., Jacobs, S., Levan, G., Nguyen, J., Ramirez, J. M., Rivera-Garcia, C., Sandoval, E., Wang, X., Wong, D., Spencer, E., Robles-Sikisaka, R., Kurzban, E., Hughes, L. D., Deng, X., Wang, C., Servellita, V., Valentine, H., De Hoff, P., Seaver, P., Sathe, S., Gietzen, K., Sickler, B., Antico, J., Hoon, K., Liu, J., Harding, A., Bakhtar, O., Basler, T., Austin, B., Isaksson, M., Febbo, P., Becker, D., Laurent, M., McDonald, E., Yeo, G. W., Knight, R., Laurent, L. C., de Feo, E., Worobey, M., Chiu, C., Suchard, M. A., Lu, J. T., Lee, W., & Andersen, K. G. (2021). Genomic epidemiology identifies emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. *medRxiv*. doi:10.1101/2021.02.06.21251159
21. World Health Organization. (2020). SARS-CoV-2 Variants. *Disease Outbreak News.* Retrieved from https://www.who.int/csr/don/31-december-2020-sars-cov2-variants/en/
22. World Health Organization. (2021). *COVID-19 Weekly Epidemiological Update*. Retrieved from https://www.who.int/publications/m/item/weekly-epidemiological-update---9-february-2021