Estimating the increased transmissibility of

- ² the B.1.1.7 strain over previously circulating
- ³ strains in England using fractions of
- GISAID sequences and the distribution of
 serial intervals
- 6

7 Chayada Piantham¹, Kimihito Ito^{2*},

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

9 ²Research Center for Zoonosis Control, Hokkaido University; itok@czc.hokudai.ac.jp

10 Abstract

11 The B.1.1.7 strain, a variant strain of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-12 2), is thought to have higher transmissibility than previously circulating strains in England. The fraction 13 of the B.1.1.7 strain among SARS-CoV-2 viruses in England have grown rapidly. In this paper, we 14 propose a method to estimate the selective advantage of a mutant strain over previously circulating strains using the time course of the fraction of B.1.1.7 strains. Our approach is based on the Maynard Smith's 15 model of allele frequencies in adaptive evolution, which assumes that the selective advantage of a mutant 16 17 strain over previously circulating strains is constant over time. Applying this method to the sequence data 18 in England using serial intervals of COVID-19, we found that the transmissibility of the B.1.1.7 strain is 19 40% (with a 95% confidence interval (CI) from 40% to 41%) higher than previously circulating strains in 20 England. The date of the emergence of B.1.1.7 strains in England was estimated to be September 20, 2020 21 with its 95% CI from September 11 to September 20, 2020. The result indicated that the control measure 22 against the B.1.1.7 strain needs to be strengthened by 40% from that against previously circulating strains. 23 To get the same control effect, contact rates between individuals need to be restricted to 0.71 of the 24 contact rates that have been achieved form the control measure taken for previously circulating strains. 25 26

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

Introduction 29

30 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of COVID-19, has

31 been rapidly evolving since its emergence in 2019. In December 2020, the Public Health England

32 detected a new cluster of SARS-CoV-2 viruses phylogenetically distinct from the other strains circulating

33 in the United Kingdom (Chand et al., 2020). These viruses belong to the lineage B.1.1.7 according to the

34 PANGO nomenclature (Rambaut, Loman, et al., 2020) and the World Health Organization (WHO)

35 designated them as Variant of Concern, year 2020, month 12, variant 01 (VOC 202012/01) (World Health

- 36 Organization, 2020).
- 37

38 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

39

- 40 strain occupied 95% of the SARS-CoV-2 circulating in England (Davies et al., 2021).
- 41

42 Several studies have compared the transmissibility of the B.1.1.7 strain to that of previously circulating

43 strains. Davies et al. estimated that the reproduction number of the B.1.1.7 strain is 43–90% (with a 95%

44 credible interval of 38–130%) higher than preexisting strains using data from England (Davies et al.,

- 45 2021). However, different models resulted in different ranges of estimates in their multiplicative increase
- 46 in reproduction number (R). Grabowski et al. estimated a 83–118% increase with a confidence interval of
- 47 71–140% compared to previously circulating strains in England (Grabowski et al., 2021). Volz et al.
- 48 estimated 40–75% increase in R using data from England (Volz et al., 2021). Their methods use linear
- regression of log odds ratio between B.1.1.7 strain and previously circulating strains and estimate increase 49
- 50 of transmissibility under 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 51

52 Volz's method (Washington et al., 2021). Chen et al. estimated 49–65% increase in reproduction number

53 using data from Switzerland also under the constant R assumption (Chen et al., 2021). Due to the high

54 transmissibility the B.1.1.7 strain, strong control measure such as lockdown was taken when the strain

55 was introduced (Davies et al., 2021). Thus, the constant R assumption is questionable when analyzing the

56 increase in the reproduction number of B.1.1.7 compared to that of previously circulating strains.

57

58 In this paper, we propose a method to estimate the selective advantage of a mutant strain over previously

59 circulating strains. Our approach is 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 60

strains is constant over time (Maynard Smith & Haigh, 1974). Appling the developed method to the 61

62 sequence data in England using the serial interval distribution of COVID-19 estimated by Nishiura et al.

63 (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. Based on the estimate, we discuss its implication 64

65 to control measures for COVID-19.

66 Materials and Methods

67 Sequence data

- 68 Nucleotide sequences of SARS-CoV-2 viruses was downloaded from GISAID EpiCoV database (Shu &
- 69 McCauley, 2017) on March 1, 2021. Nucleotide sequences that determined from viruses detected in
- 70 England were selected and aligned to the reference amino acid sequence of S 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 of S
- 73 protein 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 S protein of
- characterizing B.1.1.7 strain, retrieved from the PANGO database (Rambaut, Loman, et al., 2020).
- 76

77 Table 1. Amino acids on S 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)

78

79

- to February 19, 2021. These numbers are provided in Supplementary Table 1.
- 91

72 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

⁸⁰ We divided amino acid sequences into three groups based on amino acids shown in Table 1. The first 81 group is sequences having all of B.1.1.7-defining amino acid substitutions in Table 1. We call a virus in 82 this group a "B.1.1.7 strain". The second group are sequences which have none of the B.1.1.7-defining substitutions. We call a virus in the this group a "non-B.1.1.7 strain". The third group are sequences 83 84 which have at least one but incomplete set of the B.1.1.7-defining amino acid substitutions. We called a 85 strain in the third group a "B.1.1.7-like strain". Table 2 shows the number of sequences categorized into 86 each group. We used the number of B.1.1.7 strains and non-B.1.1.7 strains for the rest of the analyses. 87 B.1.1.7-like strains were excluded from the analyses, since we do not know whether they had the same 88 transmissibility as B.1.1.7 strains or not. Figure 1 shows the daily numbers of GISAID sequences of 89 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



Figure 1. Number of sequences in England from September 1, 2020 to February 19, 2021. The
 nucleotide sequences were retrieved from GISAID on March 1, 2020.

97 Serial interval distribution

98 The method we propose in this paper uses discrete distributions of serial intervals. Function g(i) gives 99 the probability that the onset day of a secondary infection is at *i* days after the onset day of its primary 100 case. We obtained the values of g(i) by discretizing the lognormal distribution of serial intervals of 101 COVID-19 estimated by Nishiura et al. (Nishiura et al., 2020). Thus, the probability mass function of

102 serial intervals is given by

103

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

104 where f(t) is the probability density function of a lognormal distribution with a mean of 4.7 days and a 105 standard deviation of 2.9 days.

106 Model of Advantageous Selection

107 Consider we have a large population of viruses consisting of strains of two genotypes A and a, of which

- fraction in the viral population at a calendar date t are $q_A(t)$ and $q_a(t)$, respectively. Suppose also that
- 109 genotype A is mutant of a that emerges at time t_0 .

- 111 We assume that a virus of genotype A generates 1 + s times as many secondary transmissions as those of 112 genotype a. Then, s can be considered as the coefficient of selective advantage in adaptive evolution. As
- 113 described by Maynard Smith and Haigh (1974), the fraction of viruses of allele A after the n
- 114 transmissions, q_n , satisfies

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

115

116 Let g(i) be a discrete distribution of serial intervals defined in the previous subsection. Since the

117 expected fraction of allele A in the population at calendar time t can be represented as s and $q_A(t-i)$ for $0 \le i \le t$ with a probability of g(i), the value of $q_A(t)$ can be represented as follows. 118

119

$$q_A(t) = \sum_{i=0}^t g(i) \frac{1+s}{1+sq_A(t-i)} q_A(t-i).$$
(2)

120

Assuming q(0) = 0, we can approximate the formula. 121

$$q_A(t) \approx \sum_{i=1}^{t} g(i) \frac{1+s}{1+sq_A(t-i)} q_A(t-i).$$
 (3)

122

Likelihood Function 123

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 124 be calendar dates such that $n(d_i) > 0$ for $1 \le j \le k$. Suppose that we have $n_A(d_i)$ samples of genotype A 125 at calendar date d_i . Since genotype A emerged at time t_0 , $q_A(d_i) = 0$ and $q_a(d_i) = 1$ for $d_i < t_0$. Let 126 q_0 be initial frequency of genotype A, i.e., $q_0 = q_A(t_0)$. Then the following equation gives the likelihood 127 128 function of s, t_0 , and q_0 for observing $n_A(d_i)$ samples of viruses of genotype A at calendar date d_i . 129

$$L(s,t_0,q_0;n_A(d_j)) = {\binom{n(d_j)}{n_A(d_j)}} q_A(d_j)^{n_A(d_j)} (1-q_A(d_j))^{n(d_j)-n_A(d_j)},$$
(4)

130

131 for $1 \le j \le k$. The likelihood function of s, t_0 , and q_0 for observing $n_A(d_1), \ldots, n_A(d_k)$ sequences of genotype A at calendar dates d_1, \ldots, d_k is given by the following formula. 132

$$\int_{a}^{b} \int_{a}^{b} \int_{a$$

$$L(s, t_0, q_A(t_0); n_A(d_1), \dots, n_A(d_k)) = \prod_{j=1}^k L(s, t_0, q_A(t_0); n_A(d_j))$$
(5)

133

Parameter estimation from sequence data 134

The B.1.1.7 strain was first detected in England on September 20, 2020. We assume that t_0 is this day or 135

136 someday before this day. Parameters s, t_0 , and q_0 were estimated by maximizing likelihood of

- 137 observations on September 1, 2020 and later on. B.1.1.7 strains, viruses having complete subset of
- 138 B.1.1.7-difining substitutions on its S protein were considered as genotype A. The non-B.1.1.7 strains,
- 139 viruses having none of B.1.1.7-defining substitutions were considered to be genotype *a*. The B.1.1.7-liike
- 140 strains, viruses having an incomplete set of B.1.1.7 substitutions on the S protein, were excluded from the
- 141 analysis. Parameters of s, t_0 , and q_0 were estimated by maximizing log likelihood defined in Equation (5).
- 142 The 95% confidence intervals of parameters were estimated by profile likelihood (Pawitan, 2013). The
- 143 optimization of likelihood function was done by the nloptr package in R (Johnson; Rowan, 1990).

144 Results

- 145 The selective advantage of B.1.1.7 strains over non-B.1.1.7 strains, *s*, was estimated to be 0.40 with its
- 146 95% confidence intervals from 0.40 to 0.41 (Table 2). The date of emergence of B.1.1.7 strains in
- 147 England, t_0 , was estimated to be September 20, 2020 with its 95% confidence interval from September
- 148 11, 2020 to September 20, 2020. The initial fraction of B.1.1.7 among non-B.1.1.7 and B.1.1.7 strains at
- 149 the emergence in England, q_0 , was estimated to be 0.0030 with its 95% confidence intervals from 0.0014
- 150 to 0.0031.
- 151

152 **Table 2. Maximum likelihood estimations of parameters**

Parameter	Estimate	95% CI
S	0.40	[0.40, 0.41]
t_0	September 20, 2020	[September 11, 2020, September 20, 2020]
q_0	0.0030	[0.0014, 0.0031]

153

154 Figure 2 shows the time course of the fraction of B.1.1.7 strains among all strains except B.1.1.7-like

strains detected in the England from September 1, 2020 to February 19, 2021. White circles indicate daily

156 fractions of B.1.1.7 strains among all strains except B.1.1.7-like strains. Solid line indicates the time

157 course of fraction of B.1.1.7 strains calculated using parameters estimated from the data. Dashed lines

158 indicate its lower and upper bound of its 95% CI.





161 detected in England from September 1, 2020 to February 19, 2021. White circles indicate the

162 fraction 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 fraction of B.1.1.7

164 strains calculated using parameters estimated from the data. Dashed lines indicate its lower and

165 **upper bound of its 95% CI.**

166 Discussion

167 In this paper, the selective advantage of the B.1.1.7 strain in England over non-B.1.1.7 strains was

estimated to be 0.40 with a 95% CI from 0.40 to 0.41. The date of emergence of B.1.1.7 strains in

169 England was estimated to be September 20, 2020 with its 95% confidence interval from September 11,

170 2020 to September 20, 2020. The initial fraction of B.1.1.7 among all sequences except B.1.1.7-like

171 strains at the time of emergence in England was estimated to be 0.0030 with its 95% confidence intervals

173

174 The estimated selective advantage of 0.40 indicates that the B.1.1.7 strain is 40% more transmissible than

that of previously circulating strains in England. This means that the control measures for the B.1.1.7

176 strain needs to be strengthened by 40% compared to that for the previously circulating strains. To get the

177 same control effect, contact rates between individuals needs to be restricted below 1/1.40 = 0.71 of the

178 contact rates to achieve the same control measure for the previously circulating strains.

¹⁷² from 0.0014 to 0.0031.

179 Our estimation method is based on the principle that the expected fraction of a mutant strain among all

- 180 strains can be determined from those in the previous generation using the serial interval distribution of
- 181 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. The method is related to Fraser's method for estimating the instantaneous reproduction number (Fraser, 2007). However, it is our future work to mathematically clarify the relationship between the
- 185 selective advantage estimated using our method and the instantaneous reproduction numbers of two
- 186 strains.
- 187

188 The selective advantage estimated in this study relies on the serial interval distribution, and thus the

- results may change depending on 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
- 191 lognormal distribution with a mean serial interval of 4.7 days with a standard deviation of 2.9 days.
- 192 Several groups have estimated the serial intervals of SARS-CoV-2 using different datasets. Some
- 193 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 results by Bi et al. (Bi et al., 2020). However, Ali et
- al. have reported that the serial interval estimated using data from China during before January 22, 2020
- 196 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
- 198 interval does not reflect the current situation. This is the reason why we did not use serial interval
- 199 estimated by Bi et al.
- 200

As of March 17, 2021, the B.1.1.7 strain has now been observed in 93 countries (Rambaut, Holmes, et al., 202). The selective advantage of the B.1.1.7 strains over previously circulating strains in other countries

- 202 2020). The selective advantage of the B.1.1.7 strains over previously circulating strains in other countries
 203 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 developed methodology provides important information for countries in the world to establish control
- 207 measures against variants strains such as the B.1.1.7 strain.

208 Acknowledgement

- 209 We gratefully acknowledge the laboratories responsible for obtaining the specimens and the laboratories
- 210 where genetic sequence data were generated and shared via the GISAID Initiative, on which this research
- 211 is based. This work was supported by CREST (grant number JPMJCR1413) from Japan Science and
- 212 Technology Agency (http://www.jst.go.jp/), and the World-leading Innovative and Smart Education
- 213 (WISE) Program (1801) from the Ministry of Education, Culture, Sports, Science, and Technology, Japan
- 214 (http://www.mext.go.jp/). The funders had no role in the study design, data collection and analysis,
- 215 decision to publish, or preparation of the manuscript.

216 Conflict of Interest

217 We declare that there is no conflict of interest.

218 References

- 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
- 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
- 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
- 235https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_236data/file/959438/Technical_Briefing_VOC_SH_NJL2_SH2.pdf
- 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.,
- 240 Egli, A., Eckerle, I., Denes, R., Feldkamp, M., Nissen, I., Santacroce, N., Burcklen, E.,
- 241 Aquino, C., de Gouvea, A. C., Moccia, M. D., Grüter, S., Sykes, T., Opitz, L., White, G.,
- 242 Neff, L., Popovic, D., Patrignani, A., Tracy, J., Schlapbach, R., Dermitzakis, E. T.,
- 243 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*,
- 246 2021.2003.2005.21252520. doi:10.1101/2021.03.05.21252520
- 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.,
- 249 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-
- 252 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.wadai.com/1000_4015/12/2/202
- 258 <u>https://www.mdpi.com/1999-4915/13/3/392</u>

259	9.	Johnson, S. G. The NLopt nonlinear-optimization package. Retrieved from http://ab-
260		initio.mit.edu/nlopt
261	10.	Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for
262		rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res,
263		30(14), 3059-3066. doi:10.1093/nar/gkf436
264	11.	Maynard Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene.
265		Genet Res, 23(1), 23-35. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/4407212
266	12.	Nishiura, H., Linton, N. M., & Akhmetzhanov, A. R. (2020). Serial interval of novel
267		coronavirus (COVID-19) infections. Int J Infect Dis, 93, 284-286.
268		doi:10.1016/j.ijid.2020.02.060
269	13.	Pawitan, Y. (2013). In All Likelihood: Statistical Modelling and Inference Using
270		Likelihood. Croydon: Oxford University Press.
271	14.	Rai, B., Shukla, A., & Dwivedi, L. K. (2021). Estimates of serial interval for COVID-19:
272		A systematic review and meta-analysis. Clin Epidemiol Glob Health, 9, 157-161.
273		doi:10.1016/j.cegh.2020.08.007
274	15.	Rambaut, A., Holmes, E. C., O'Toole, A., Hill, V., McCrone, J. T., Ruis, C., du Plessis,
275		L., & Pybus, O. G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages
276		to assist genomic epidemiology. Nat Microbiol, 5(11), 1403-1407. doi:10.1038/s41564-
277		020-0770-5
278	16.	Rambaut, A., Loman, N., Pybus, O., Barclay, W., Barrett, J., Carabelli, A., Connor, T.,
279		Peacock, T., Robertson, D. L., & Volz, E. (2020, 5 February 2021). Preliminary genomic
280		characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of
281		spike mutations. Retrieved from https://virological.org/t/preliminary-genomic-
282		characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-
283		spike-mutations/563
284	17.	Rowan, T. (1990). Functional Stability Analysis of Numerical Algorithms. (Ph.D. Thesis).
285		University of Texas, Austin.
286	18.	Shu, Y., & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data -
287		from vision to reality. Euro Surveill, 22(13). doi:10.2807/1560-
288		7917.ES.2017.22.13.30494
289	19.	Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg, L., Hinsley, W.
290		R., Laydon, D. J., Dabrera, G., O'Toole, Á., Amato, R., Ragonnet-Cronin, M., Harrison,
291		I., Jackson, B., Ariani, C. V., Boyd, O., Loman, N. J., McCrone, J. T., Gonçalves, S.,
292		Jorgensen, D., Myers, R., Hill, V., Jackson, D. K., Gaythorpe, K., Groves, N., Sillitoe, J.,
293		Kwiatkowski, D. P., Flaxman, S., Ratmann, O., Bhatt, S., Hopkins, S., Gandy, A.,
294		Rambaut, A., & Ferguson, N. M. (2021). Transmission of SARS-CoV-2 Lineage B.1.1.7
295		in England: Insights from linking epidemiological and genetic data. medRxiv,
296		2020.2012.2030.20249034. doi:10.1101/2020.12.30.20249034
297	20.	Washington, N. L., Gangavarapu, K., Zeller, M., Bolze, A., Cirulli, E. T., Schiabor
298		Barrett, K. M., Larsen, B. B., Anderson, C., White, S., Cassens, T., Jacobs, S., Levan, G.,
299		Nguyen, J., Ramirez, J. M., Rivera-Garcia, C., Sandoval, E., Wang, X., Wong, D.,
300		Spencer, E., Robles-Sikisaka, R., Kurzban, E., Hughes, L. D., Deng, X., Wang, C.,
301		Servellita, V., Valentine, H., De Hoff, P., Seaver, P., Sathe, S., Gietzen, K., Sickler, B.,

302	Antico, J., Hoon, K., Liu, J., Harding, A., Bakhtar, O., Basler, T., Austin, B., Isaksson,
303	M., Febbo, P., Becker, D., Laurent, M., McDonald, E., Yeo, G. W., Knight, R., Laurent,
304	L. C., de Feo, E., Worobey, M., Chiu, C., Suchard, M. A., Lu, J. T., Lee, W., &
305	Andersen, K. G. (2021). Genomic epidemiology identifies emergence and rapid
306	transmission of SARS-CoV-2 B.1.1.7 in the United States. medRxiv.
307	doi:10.1101/2021.02.06.21251159
308	21. World Health Organization. (2020). SARS-CoV-2 Variants. Disease Outbreak News.
309	Retrieved from https://www.who.int/csr/don/31-december-2020-sars-cov2-variants/en/
310	22. World Health Organization. (2021). COVID-19 Weekly Epidemiological Update.
311	Retrieved from https://www.who.int/publications/m/item/weekly-epidemiological-
312	update9-february-2021