Predicted domination of variant Delta of SARS-CoV-2 before Tokyo Olympic games, Japan

Kimihito Ito[1*], Chayada Piantham[2], and Hiroshi Nishiura[3]

[1]*International Institute for Zoonosis Control, Hokkaido University; itok@czc.hokudai.ac.jp*
[2]*Graduate School of Infectious Diseases, Hokkaido University; chayada@czc.hokudai.ac.jp*
[3]*Graduate School of Medicine, Kyoto University; nishiurah@gmail.com*

* *Corresponding author*

Abstract

Using nucleotide sequences of SARS-CoV-2 strains collected in Japan, the relative instantaneous reproduction numbers of the R.1, Alpha, and Delta variants with respect to other strains circulation in Japan were estimated at 1.256, 1.449, and 1.776, respectively. The numbers can range 1.198–1.335 for R.1, 1.342–1.596 for Alpha, and 1.557–2.00 for Delta depending on the assumed serial interval distributions. The frequency of the Delta is expected to take over the Alpha in Japan around July 12, 2021.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of COVID-19, has undergone adaptive evolution since its emergence in the human population in 2019. On May 31, 2021, the World Health Organization (WHO) has designated four variants SARS-CoV-2 as Variants of Concerns (VOCs)—Alpha, a variant lineage first reported in the United Kingdom; Beta, those in South Africa; Gamma, those in Brazil; Delta, those in India (World Health Organization, 2021). Volz et al. estimated 50–100% increase in R of the Alpha variant also using data from England (Volz et al., 2021). In our previous paper, we estimated the selective advantage of the Alpha variant in the England to be 26–45% higher than pre-existing strains using GISAID sequence data collected in England (Piantham, Linton, Nishiura, & Ito, 2021). These analyses use models that assume two circulating variants. As of June 9, 2021, the state of emergency has been declared in ten prefectures in Japan—Tokyo, Kyoto, Osaka, Hyogo, Aichi, Fukuoka, Hokkaido, Okayama, Hiroshima, and Okinawa. Due to the high transmissibility of VOCs, the replacement from local strains to Alpha and Delta variants is progressing, posing a serious public health threat in Japan. In such a situation where multiple variants are co-circulating, a model describing the adaptive evolution among multiple variants is required. Using nucleotide sequences of SARS-CoV-2 in the GISAID database (Shu & McCauley, 2017) submitted from Japan and the serial intervals of COVID-19 estimated by Nishiura et al. (Nishiura, Linton, & Akhmetzhanov, 2020), we estimate the relative instantaneous reproduction numbers of the R.1, Alpha, and Delta variants with

respect to other strains circulating in Japan. We also show the expected temporal changes in variant frequencies of SARS-CoV-2 in Japan until early August 2021.

Model of Advantageous Selection Among Multiple Variants

Suppose that we have a large population of viruses consisting of strains of genotypes a, A_1, …, A_n, of which frequencies in the viral population at a calendar time t are $q_a(t)$, $q_{A_1}(t)$,…, $q_{A_n}(t)$, respectively. Suppose also that genotype a is a wildtype virus that was circulating at the beginning of the target period of analysis and that genotype A_1,…, A_n are mutants that introduced to the population at time t_1,…, t_n, respectively.

We assume that viruses of genotype A_1,…, A_n generate $1 + s_1$,…, $1 + s_n$ times as many secondary transmissions as those of genotype a, respectively. 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 total number of infections by viruses of any genotype of a or $A_1, \dots, A_n$ at calendar time t and $g(j)$ be the probability mass function of serial intervals. Suppose that $g(j)$ is small enough to be neglected for $j < 1$ and $j > l$. The instantaneous reproduction numbers of genotypes a and $A_1, \dots, A_n$ at calendar time t are represented as follows:

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

$$R_{A_i}(t) = \frac{q_{A_i}(t)I(t)}{\sum_{j=1}^{l} g(j)q_{A_i}(t-j)I(t-j)} \tag{2}$$

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

$$R_{A_i}(t) = (1 + s_i)R_a(t) \tag{3}$$

for each calendar time $t \geq t_i$. We call the value of $1 + s_i$ the relative instantaneous reproduction number of A_i with respect to a.

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.

$$I(t-1) \approx I(t-2) \approx \cdots \approx I(t-l)\,. \tag{4}$$

The frequency of genotype A_i in the viral population at calendar time t, $q_{A_i}(t)$, can be modeled as follows:

$$q_{A_i}(t) = \frac{q_{A_i}(t)I(t)}{q_a(t)I(t) + \sum_{i=1}^{n} q_{A_i}(t)I(t)}$$

$$= \frac{\sum_{j=1}^{l} g(j)\, R_{A_i}(t) q_{A_i}(t-j) I(t-j)}{\sum_{j=1}^{l} g(j) R_a(t) q_a(t-j) I(t-j) + \sum_i^n \sum_{j=1}^{l} g(j) R_{A_i}(t) q_{A_i}(t-j)\, I(t-j)}$$

$$= \frac{\sum_{j=1}^{l} g(j)\,(1+s_i) q_{A_i}(t-j)}{1+\sum_{i=1}^{n}\sum_{j=1}^{l} s_i g(j) q_{A_i}(t-j)}. \quad (5)$$

Nucleotide sequences of SARS-CoV-2 collected in Japan

Nucleotide sequences of SARS-CoV-2 were downloaded from the GISAID EpiCoV database (Shu & McCauley, 2017) on June 5, 2021. The spike protein's amino acid sequences of viruses detected in Japan were obtained by applying the method described in our previous paper (Piantham et al., 2021) to Japanese dataset. Each amino acid sequence was assigned to one of four labels indicating variants circulating in Japan, i.e., Alpha, Delta, R.1, or other according to the key amino acids on the spike protein shown in Table 1 (Rambaut et al., 2020). Amino acid sequences before December 1, 2020 were excluded from analysis because none of R.1, Alpha, or Beta linkages was observed in Japan. Amino acid sequences after May 8, 2021 were also excluded from the analysis because they have only a few sequences per day. Finally, daily variant frequencies of a total of 30,538 sequences—consisting of 6,044 R.1, 11,310 Alpha, 168 Delta, and 13,016 other viruses—were obtained and used in the rest of analyses (Supplementary Table 1).

Table 1. Amino acids at key positions in the spike protein to assign a variant label to viruses

Amino acid at key positions on spike protein			Variant label
484	501	452	
K	N	L	R.1
E	Y	L	Alpha
E	N	R	Delta
Q	N	R	Delta
None of above combinations			other

Relative instantaneous reproduction numbers

Let $N(t)$ be the number of sequences of either genotype $A_1, \dots, A_n$, or a observed at calendar time t, and let $d_1, \dots, d_k$ be calendar times such that $N(d_j) > 0$ for $1 \le j \le k$. Suppose that we have $N_{A_i}(d_j)$ sequences of genotype A_i at calendar time d_j for $1 \le i \le n$ and $1 \le j \le k$. Since genotype A_i emerged at time t_i, $q_{A_i}(d_j) = 0$ for $d_j < t_i$. If the genotype A_i emerges with an initial frequency of $q_i(t_i)$ at calendar time t_i, then the likelihood function of parameters $s_1, \dots, s_n$ and $q_1(t_1), \dots, q_n(t_n)$ for observing

$N_{A_1}(d_1), \dots, N_{A_n}(d_1), \dots, N_{A_1}(d_k), \dots, N_{A_n}(d_k)$ sequences of genotype $A_1, \dots, A_n$, at calendar times $d_1, \dots, d_k$ is given by the following formula:

$$L\left(s_1, \dots, s_n, q_1(t_1), \dots, q_n(t_n)\ ;\ N_{A_1}(d_1), \dots, N_{A_n}(d_1), \dots, N_{A_1}(d_k), \dots, N_{A_n}(d_k)\right)$$
$$= \prod_{j=1}^{k} \frac{N(d_j)!}{N_a(d_j)! \prod_i^n N_{A_i}(d_j)!} q_a(d_j)^{N_a(d_j)} \prod_i^n q_{A_i}(d_j)^{N_{Ai}(d_j)} \tag{6}$$

where $N_a(d_j) = N(d_j) - \sum_{i=1}^{n} N_{A_i}(d_j)$ for $1 \le j \le k$.

The R.1, Alpha, and Delta variants were first detected in the GISAID sequence from Japan on December 1, 2020, December 15, 2020, and January 10, 2021, respectively. We assume that $t_{R.1}$, t_{Alpha}, and t_{Delta} are these dates, respectively. The serial intervals were assumed to follow a lognormal distribution with $\mu = 1.38$ and $\sigma = 0.563$ discretized and truncated so that $g(0) = 0$ and $g(j) = 0$ for $j > 0$. Parameters $s_{R.1}$, s_{Alpha}, s_{Delta}, $q_{R.1}$, q_{Alpha}, and q_{Delta}were estimated by maximizing the likelihood defined in Equation (6). The 95% confidence intervals (CIs) of parameters $s_{R.1}$, s_{Alpha}, s_{Delta}, $q_{R.1}(t_{R.1})$, $q_{Alpha}(t_{Alpha})$, and $q_{Delta}(t_{Delta})$ were estimated by profile likelihood (Pawitan, 2013). The optimization of likelihood function and calculation of 95% confidence intervals were performed using the nloptr package in R (Johnson, 2020; Rowan, 1990).

The selective advantages $s_{R.1}$, s_{Alpha}, and s_{Delta} with respect to other strains circulating in Japan were estimated to be 0.256 (with a 95% CI from 0.253 to 0.257), 0.449 (with a 95% CI from 0.446 to 0.450), and 0.776 (with a 95% CI from 0.768 to 0.782), assuming that the serial interval distribution is a lognormal distribution with a log mean of 1.38 and log standard deviation of 0.563 (Table 2). The initial frequencies, $q_{R.1}(t_{R.1})$, $q_{Alpha}(t_{Alpha})$, and $q_{Delta}(t_{Delta})$, were estimated to be 0.00385 (95% CI from 0.00374 to 0.00408), 0.00108 (95% CI from 0.00106 to 0.00112), and 0.000000700 (95% CI from 0.000000627, 0.000000776), respectively (Table 2).

Table 2. Estimated values of parameters and their 95% CIs

Parameter	Estimated values	95% CI
$s_{R.1}$	0.256	[0.253, 0.257]
s_{Alpha}	0.449	[0.446, 0.450]
s_{Delta}	0.776	[0.768, 0.782]
$q_{R.1}$	0.00385	[0.00374, 0.00408]
q_{Alpha}	0.00108	[0.00106, 0.00112]
q_{Delta}	0.000000700	[0.000000627, 0.000000776]

Figure 1 shows temporal changes in frequencies of R.1, Alpha, Delta variants, and other strains circulating in Japan estimated by our model. As can be observed from the trajectories of predicted frequencies for the Alpha and Delta, the Alpha seems to be predominant in Japan at the beginning of June, 2021, but the Delta is predicted to take over the Alpha on July 12, 2021 with a 95% CI from July 8 to July 16, 2021.

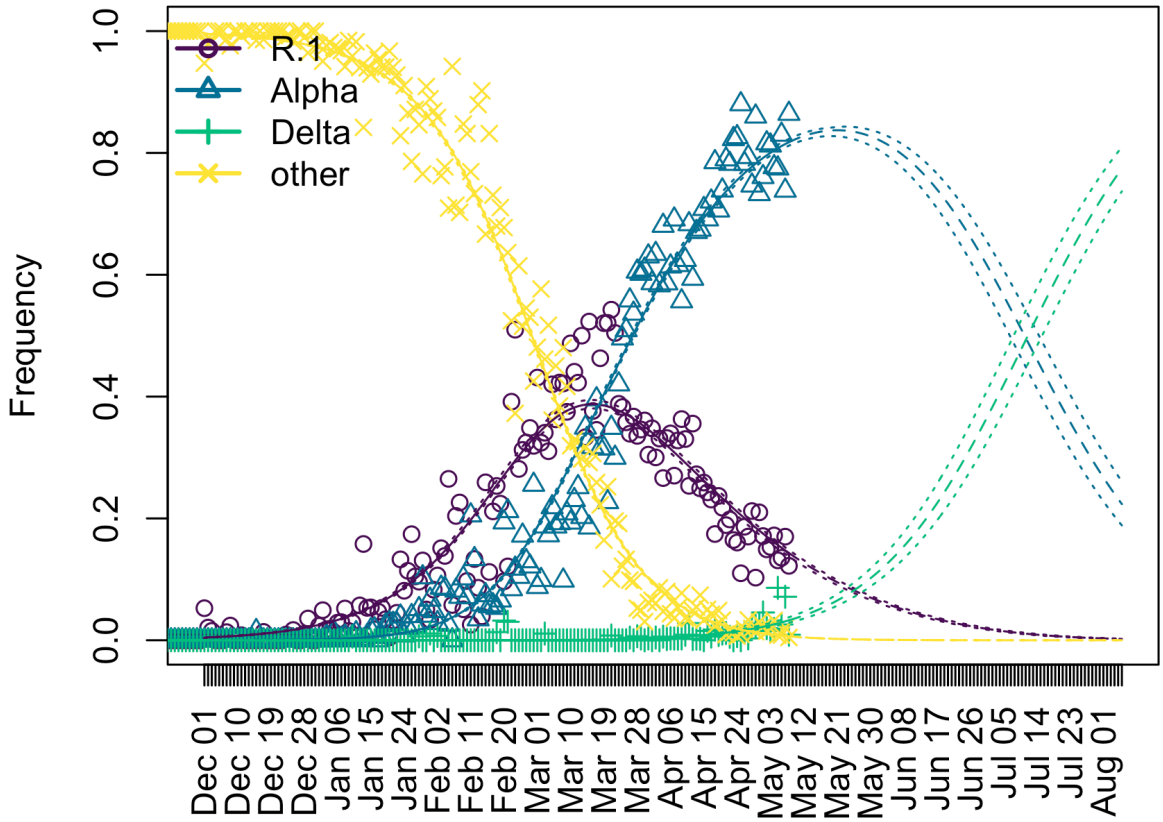

Figure 1. Temporal changes in variant frequencies of R.1 (purple), Alpha (blue), Delta (green), and others (yellow) circulating in Japan from December 1, 2020 to August 6, 2021. The circles, triangles, pluses, and crosses respectively indicate daily frequencies of SARS-CoV-2 sequences of R.1, Alpha, Delta, and others collected in Japan from December 1 2020 to May 8, 2021. The solid lines indicate the maximum likelihood estimates of frequencies of the variants using sequence data collected in Japan from December 1, 2020 to May 8. The dashed lines represent the predicted variant frequencies after May 8, 2021. Dotted lines indicate lower and upper bounds of 95% CI of estimated variant frequencies.

Population average of relative instantaneous reproduction numbers

The population average of relative instantaneous reproduction numbers, $1 + s$, with respect to strains labeled as others at calendar time t was estimated by $1 + s_{R.1}q_{R.1}(t) + s_{Alpha}q_{Alpha}(t) + s_{Delta}q_{Delta}(t)$. Since the maximum likelihood estimates of $s_{R.1}$, $s_{R.1}$, and s_{Delta} can be affected by the log mean and log standard deviation of the lognormal serial interval distribution, sensitivity analyses of parameters $s_{R.1}$, s_{Alpha}, s_{Delta} were performed by using the combinations of μ and σ sampled along the boundary of the 95% confidence area of likelihood surface of the serial interval distribution (Nishiura et al., 2020 ; Piantham et al., 2021). As a result, we estimated the ranges of the maximum likelihood estimates of $s_{R.1}$, $s_{R.1}$, and s_{Delta} to be 0.198–0.335, 0.342–0.596, and 0.557–1.00, respectively.

Figure 2 shows temporal changes in the population average of relative instantaneous reproduction numbers with respect to strains circulating before December 1 in Japan. The population average of relative instantaneous reproduction numbers stayed around one until the end of January 2021, since the strains other than the R.1, Alpha, or Delta variants were predominant around that time. From the beginning of February 2021, the population average of relative instantaneous reproduction has been elevated due to the increase of frequencies of the Alpha variant (Figure 1). The population average of relative instantaneous reproduction numbers reached at 1.20 on March 5, 2021. It should be noted that around this day the decrease in the number of COVID-19 cases in Tokyo leveled off although Tokyo was under the declaration of emergency status. From around the end of May 2021, the population average of relative instantaneous reproduction numbers was elevated because the frequencies of the Alpha started to decrease due to the replacement of the Alpha by the Delta (Figure 1). The elevation of population average of relative instantaneous reproduction numbers will continue until the Alpha is completely replaced by the Delta (Figure 2).

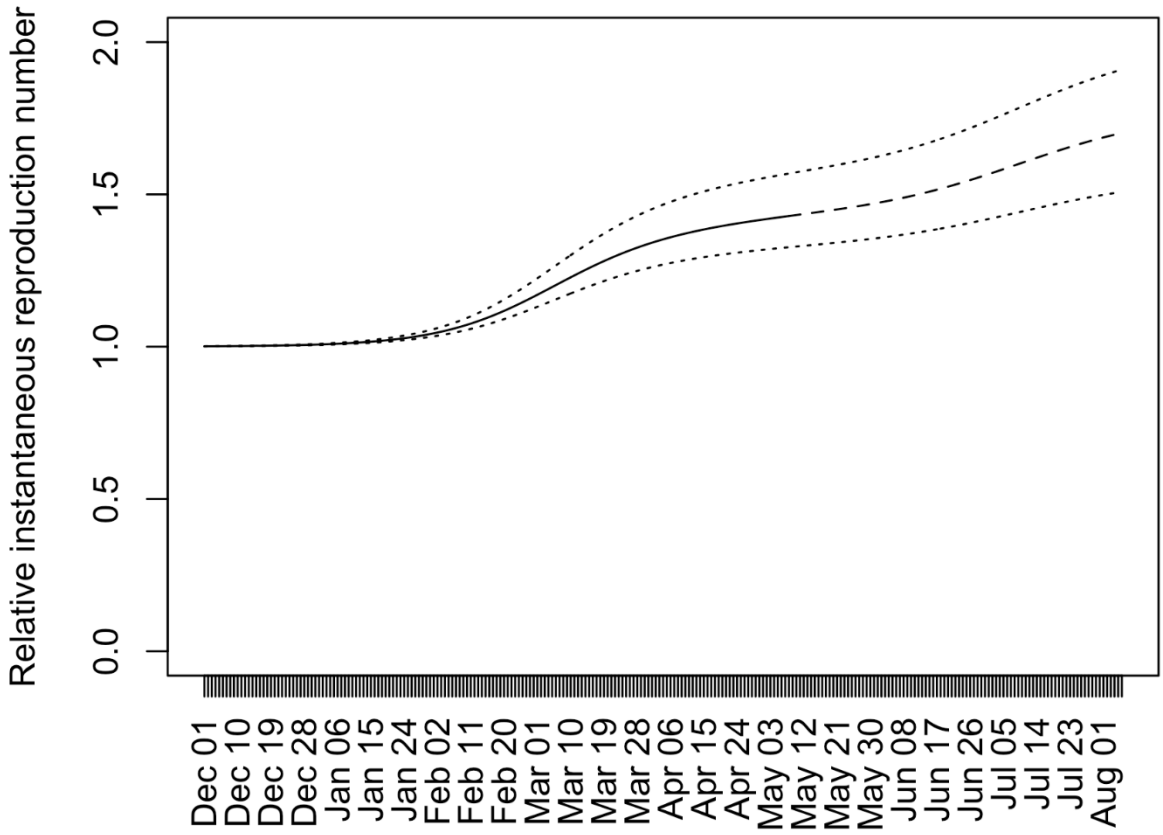

Figure 2. Temporal changes in the population average of the relative instantaneous reproduction numbers with respect to strains circulating before December 1 in Japan. The solid line (until May 8, 2021) and dashed line (after May 8, 2021) indicate the population average of the relative instantaneous reproduction numbers when $s_{R.1} = 0.256$, $s_{Alpha} = 0.449$, and $s_{Delta} = 0.776$, which are calculated by assuming the lognormal serial interval distribution with $\mu = 1.38$ and $\sigma = 0.563$. The lower dotted line indicates the population average of the relative instantaneous reproduction numbers when $s_{R.1} = 0.198$, $s_{Alpha} = 0.342$ and $s_{Delta} = 0.557$, which are the lower bounds of $s_{R.1}$, s_{Alpha}, and s_{Delta} calculated based of the 95% confidence area of the lognormal serial interval distribution. The upper dotted line indicates those when $s_{R.1} = 0.335$, $s_{Alpha} = 0.596$ and $s_{Delta} = 1.00$, which are upper bounds calculated in the same way (Figure 2).

Tokyo Olympic games

We have shown that the Delta variant possesses evidently greater transmissibility than the Alpha variant. The relative instantaneous reproduction number of the Alpha with respect to the strains circulating before December 1 in Japan was estimated at 1.449 and those of the Delta was estimated at 1.776. This means that the Delta possesses 1.226 times higher transmissibility than the Alpha. While the Alpha variant has

just replaced other variants in Japan over the last 5 months, it is very likely that it is just a matter of time for the Delta variant to replace others including the Alpha. An important learning point of this rapid communications is that, unfortunately, the replacement is likely to happen mostly before the Tokyo Olympic games from July 23, 2021. The risk assessment must account for the fact that substantial number of international visitors during the Games will be exposed to the Delta variant, and an increased mobility could help further spread COVID-19 caused by this variant with an elevated transmissibility around the world.

Already we have seen that the interventions have had to be strengthened with the Alpha variant. We have just shown that the focused interventions on drinking and eating services that have been highly effective over the last 1 year was not effective to substantially reduce the reproduction number below the value of 1. However, the Delta variant may require even more.

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. A major part of the sequence data we used in this study were from National Institute of Infectious Diseases, Japan, and prefectural institutes for public health in Japan. This work was supported by the Japan Agency for Medical Research and Development (grant numbers JP20fk0108535). K.I. received funding JSPS KAKENHI (21H03490). P.C. was supported by the World-leading Innovative and Smart Education Program (1801) from the Ministry of Education, Culture, Sports, Science, and Technology, Japan. H.N. received funding from Health and Labor Sciences Research Grants (20CA2024 and 20HA2007); the Japan Agency for Medical Research and Development (JP20fk0108140); JSPS KAKENHI (21H03198) and the Japan Science and Technology Agency (JST) SICORP (e-ASIA) program (JPMJSC20U3). 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

Fraser, C. (2007). Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS One, 2*(8), e758. doi:10.1371/journal.pone.0000758

Johnson, S. G. (2020). The NLopt nonlinear-optimization package. Retrieved from http://github.com/stevengj/nlopt

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

Pawitan, Y. (2013). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Croydon: Oxford University Press.

Piantham, C., Linton, N. M., Nishiura, H., & Ito, K. (2021). Estimating the elevated transmissibility of the B.1.1.7 strain over previously circulating strains in England using GISAID sequence frequencies. *medRxiv*. doi:10.1101/2021.03.17.21253775

Rambaut, A., Holmes, E. C., O'Toole, A., Hill, V., McCrone, J. T., Ruis, C., . . . 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

Rowan, T. (1990). *Functional Stability Analysis of Numerical Algorithms.* (Ph.D. Thesis). University of Texas, Austin.

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

Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg, L., . . . Ferguson, N. M. (2021). Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature, 593*(7858), 266-269. doi:10.1038/s41586-021-03470-x

World Health Organization. (2021). SARS-CoV-2 Variants of Concern and Variants of Interest, updated 31 May 2021. Retrieved from https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/