

# 1 Lineage frequency time series reveal elevated levels of genetic drift 2 in SARS-CoV-2 transmission in England

3 QinQin Yu<sup>1,9,\*</sup>, Joao Ascensao<sup>2,†</sup>, Takashi Okada<sup>1,3,4,5,‡</sup>, The COVID-19 Genomics UK  
4 (COG-UK) consortium<sup>6\*</sup>, Olivia Boyd<sup>7</sup>, Erik Volz<sup>6,7</sup>, and Oskar Hallatschek<sup>1,3,8,†</sup>

5 <sup>1</sup>Department of Physics, University of California, Berkeley, United States

6 <sup>2</sup>Department of Bioengineering, University of California, Berkeley, United States

7 <sup>3</sup>Department of Integrative Biology, University of California, Berkeley, United States

8 <sup>4</sup>Institute for Life and Medical Sciences, Kyoto University, Kyoto, Japan

9 <sup>5</sup>RIKEN iTHEMS, Wako, Saitama, Japan

10 <sup>6</sup><https://www.cogconsortium.uk>

11 <sup>7</sup>MRC Centre for Global Infectious Disease Analysis, Department of Infectious Disease  
12 Epidemiology, Imperial College London, London, United Kingdom

13 <sup>8</sup>Peter Debye Institute for Soft Matter Physics, Leipzig University, Leipzig, Germany

14 <sup>9</sup>Present affiliation: Department of Immunology and Infectious Diseases, Harvard T.H.  
15 Chan School of Public Health, Boston, Massachusetts, United States

16 <sup>‡</sup>These authors contributed equally.

17 <sup>\*</sup>Full list of consortium names and affiliations are in the appendix

18 November 21, 2022

## 19 Abstract

20 Random genetic drift in the population-level dynamics of an infectious disease outbreak results from the  
21 randomness of inter-host transmission and the randomness of host recovery or death. The strength of genetic  
22 drift has been found to be high for SARS-CoV-2 due to superspreading, and this is expected to substantially  
23 impact the disease epidemiology and evolution. Noise that results from the measurement process, such as  
24 biases in data collection across time, geographical areas, etc., can potentially confound estimates of genetic  
25 drift as both processes contribute “noise” to the data. To address this challenge, we develop and validate  
26 a method to jointly infer genetic drift and measurement noise from time-series lineage frequency data. We  
27 apply this method to over 490,000 SARS-CoV-2 genomic sequences from England collected between March  
28 2020 and December 2021 by the COVID-19 Genomics UK (COG-UK) consortium. We find that even after  
29 correcting for measurement noise, the strength of genetic drift is consistently, throughout time, higher than  
30 that expected from the observed number of COVID-19 positive individuals in England by 1 to 3 orders  
31 of magnitude. Corrections taking into account epidemiological dynamics (susceptible-infected-recovered  
32 or susceptible-exposed-infected-recovered models) do not explain the discrepancy. Moreover, the levels of  
33 genetic drift that we observe are higher than the estimated levels of superspreading found by modeling  
34 studies that incorporate data on actual contact statistics in England. We discuss how even in the absence  
35 of superspreading, high levels of genetic drift can be generated via community structure in the host contact  
36 network. Our results suggest that further investigations of heterogeneous host contact structure may be  
37 important for understanding the high levels of genetic drift observed for SARS-CoV-2 in England.

\*[qinqinyu@berkeley.edu](mailto:qinqinyu@berkeley.edu)

†[ohallats@berkeley.edu](mailto:ohallats@berkeley.edu)

## 38 Introduction

39 Random genetic drift is the change in the composition of a population over time due to the randomness  
40 of birth and death processes. In pathogen transmission, births occur as a result of transmission of the  
41 pathogen between hosts and deaths occur as a result of infected host recovery or death. The strength of  
42 genetic drift in pathogen transmission is determined by the disease prevalence, the disease epidemiology  
43 parameters [1], the variance in offspring number (the number of secondary infections that result from an  
44 infected individual) [2], as well as host contact patterns [3]. Many diseases have been found to exhibit high  
45 levels of genetic drift, such as SARS, MERS, tuberculosis, and measles [2, 4, 5]. The strength of genetic  
46 drift affects how the disease spreads through the population [2, 3, 6] how new variants emerge [7, 8, 9, 10,  
47 11], and the effectiveness of interventions [12], making it an important quantity to accurately estimate for  
48 understanding disease epidemiology, evolution, and control.

49 The effective population size is often used to quantify the strength of genetic drift; it is the population size  
50 in an idealized Wright-Fisher model (with discrete non-overlapping generations, a constant population size,  
51 and offspring determined by sampling with replacement from the previous generation) that would reproduce  
52 the observed dynamics [13]. If the effective population size is lower than the true population size, it is an  
53 indication that there are additional sources of stochasticity beyond random sampling with replacement; thus,  
54 a lower effective population size indicates a higher level of genetic drift.

55 Transmission of SARS-CoV-2 has been shown to exhibit high levels of superspreading (high variance in  
56 offspring number) [14, 15, 16] and high levels of genetic drift (low effective population sizes) [17, 18, 19] (see  
57 also Supplementary table S1). However, studies have focused on particular times and locations, and we lack  
58 systematic studies over time and space (see Ref. [20] for a recent first study that uses contact tracing data  
59 to infer changes in SARS-CoV-2 superspreading over time in Hong Kong). Performing a systematic study  
60 may be most feasible with a large-scale surveillance dataset, such as that from the COVID-19 Genomics UK  
61 (COG-UK) consortium, which has sequenced almost 3 million cases of SARS-CoV-2 in both surveillance and  
62 non-surveillance capacities as of October 5, 2022. We focus specifically on this dataset, and specifically on  
63 England, due to its consistently large number of sequenced SARS-CoV-2 cases since early in the pandemic.

64 A challenge to performing a systematic study of the strength of genetic drift for SARS-CoV-2 and other  
65 pathogens is how to handle measurement noise, or noise from the data collection process [21]. Measurement  
66 noise can arise from a variety of factors, including variability in the testing rate across time, geographic  
67 locations, demographic groups, and symptom status, and biases in contact tracing. Methods exist to infer  
68 measurement noise from time-series lineage or allele frequencies [22, 23, 24] (see the Supplementary infor-  
69 mation for a summary of other methods used for inferring genetic drift, and additional references). Note  
70 that here we use the term “lineage” to refer to a group of sequences that are genetically similar to one  
71 another, which are not necessarily the same as the lineages defined by the Pango nomenclature. Intuitively,  
72 in time-series frequency data, genetic drift leads to frequency fluctuations whose magnitudes scale with time,  
73 whereas measurement noise leads to frequency fluctuations whose magnitudes do not scale with time (Figure  
74 1a). Thus, this system has been mapped onto a Hidden Markov Model (HMM) with continuous hidden  
75 and observed states (similar to a Kalman filter), where the hidden states are the true frequencies and the  
76 observed states are the observed frequencies (Figure 1b), and the processes of genetic drift and measurement  
77 noise determine the transition and emission probabilities, respectively [25, 26]. Methods often assume uni-  
78 form sampling of infected individuals from the population [27, 22, 23], but this assumption does not usually  
79 hold outside of surveillance studies. A recent study accounted for overdispersed sampling of sequences in the  
80 inference of fitness coefficients of SARS-CoV-2 variants, but assumes constant overdispersion over time [28];  
81 in reality, the observation process may change over time due to changes in testing intensity between locations  
82 and subpopulations.

83 In this study, we develop a method to jointly infer genetic drift and measurement noise that allows  
84 measurement noise to be overdispersed (rather than uniform) and for the strength of overdispersion to vary  
85 over time (rather than stay constant). By fitting this model to observed lineage frequency trajectories  
86 from simulations, we show that the effective population size and the strength of measurement noise can  
87 be accurately determined in most situations, even when both quantities are varying over time. We then  
88 apply our validated method to estimate the strengths of genetic drift and measurement noise for SARS-  
89 CoV-2 in England across time (from March 2020 until December 2021) and space using over 490,000 SARS-  
90 CoV-2 genomic sequences from COG-UK. We find high levels of genetic drift for SARS-CoV-2 consistently

91 throughout time that cannot be explained by literature values of superspreading. We discuss how community  
 92 structure in the host contact network may partially explain these results. Additionally, we observe that  
 93 sampling of infected individuals from the population is mostly uniform for this dataset, and we also find  
 94 evidence of spatial structure in the transmission dynamics of B.1.177, Alpha, and Delta.

## 95 Results

### 96 Method for jointly inferring genetic drift and measurement noise from time-series 97 lineage frequency data

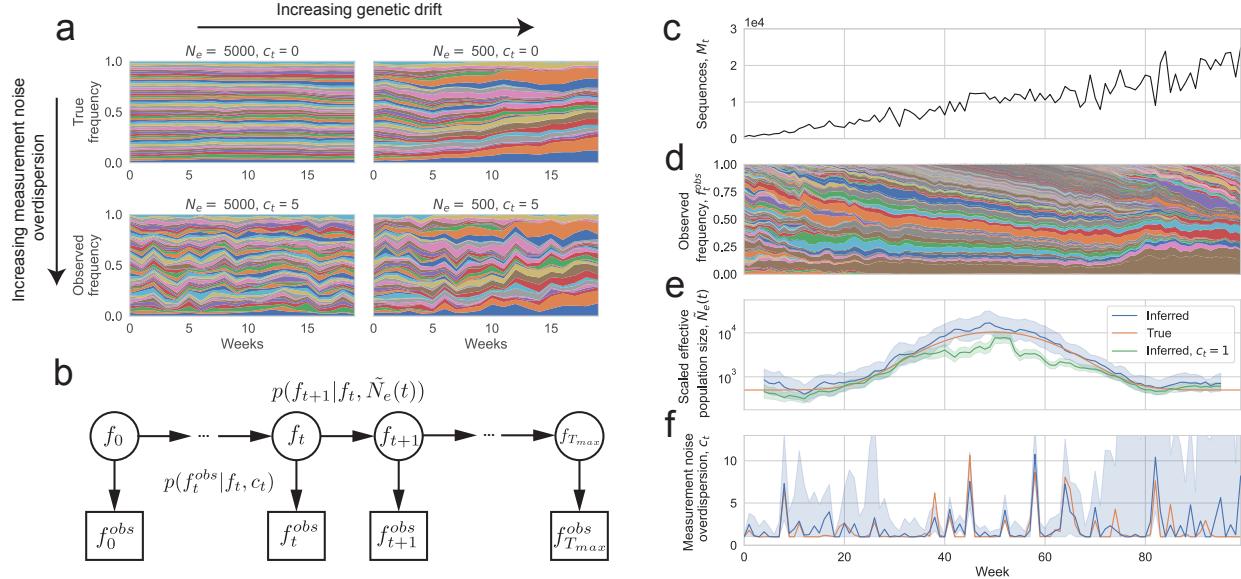


Figure 1: A Hidden Markov Model with continuous hidden and observed states (similar to a Kalman filter) for inferring genetic drift and measurement noise from lineage frequency time series. (a) Illustration of how genetic drift and measurement noise affect the observed frequency time series. Muller plot of lineage frequencies from Wright-Fisher simulations with effective population size 500 and 5000, with and without measurement noise. In simulations with measurement noise, 100 sequences were sampled per week with the measurement noise overdispersion parameter  $c_t = 5$  (parameter defined in text). All simulations were initialized with 50 lineages at equal frequency. A lower effective population size leads to larger frequency fluctuations whose variances add over time, whereas measurement noise leads to increased frequency fluctuations whose variances do not add over time. (b) Schematic of Hidden Markov Model describing frequency trajectories.  $f_t$  is the true frequency at time  $t$  (hidden states) and  $f_t^{obs}$  is the observed frequency at time  $t$  (observed states). The inferred parameters are  $\tilde{N}_e(t) \equiv N_e(t)\tau(t)$ , the effective population size scaled by the generation time, and  $c_t$ , the overdispersion in measurement noise ( $c_t = 1$  corresponds to uniform sampling of sequences from the population). (c-f) Validation of method using Wright-Fisher simulations of frequency trajectories with time-varying effective population size and measurement noise. (c) Simulated number of sequences. (d) Simulated lineage frequency trajectories. (e) Inferred scaled effective population size ( $\tilde{N}_e(t)$ ) on simulated data compared to true values. (f) Inferred measurement noise ( $c_t$ ) on simulated data compared to true values. In (e) the shaded region shows the 95% confidence interval calculated using the posterior, and in (f) the shaded region shows the 95% confidence interval calculated using bootstrapping (see Methods).

98 We first summarize the statistical inference method that we developed to infer time-varying effective  
 99 population sizes from neutral lineage frequency time series that are affected by overdispersed measurement  
 100 noise (more variable than uniform sampling). We explain the method more extensively in the Methods.

101 Briefly, we use a Hidden Markov Model (HMM) with continuous hidden and observed states (similar to  
102 a Kalman filter), where the hidden states are the true frequencies ( $f_t$ , where  $t$  is time), and the observed  
103 states are the observed frequencies ( $f_t^{obs}$ ) (Figure 1b) (see Methods). The transition probability between  
104 hidden states is set by genetic drift, where the mean true frequency is the true frequency at the previous  
105 time  $E(f_{t+1}|f_t) = f_t$ , and when the frequencies are rare the variance in frequency is proportional to the  
106 mean,  $\text{Var}(f_{t+1}|f_t) = \frac{f_t}{N_e(t)}$ .  $\tilde{N}_e(t) = N_e(t)\tau(t)$  where  $N_e(t)$  is the effective population size and  $\tau(t)$  is the  
107 generation time, and both quantities can vary over time; however, we are only able to infer the compound  
108 parameter  $N_e(t)\tau(t)$ . The emission probability between hidden and observed states is set by measurement  
109 noise, where the mean observed frequency is the true frequency  $E(f_t^{obs}|f_t) = f_t$  and when the frequencies  
110 are rare the variance in the observed frequency is proportional to the mean,  $\text{Var}(f_t^{obs}|f_t) = c_t \frac{f_t}{M_t}$ .  $c_t \geq 1$   
111 describes the time-varying deviation from uniform sampling ( $c_t = 1$ ), and  $M_t$  is the number of sequences  
112 at time  $t$ . Our model assumes that the number of individuals and frequency of a lineage is high enough  
113 such that the central limit theorem applies; to meet this condition, we created “superlineages” where we  
114 randomly and exclusively grouped lineages together such that the sum of their abundances and frequencies  
115 was above a threshold (see Methods).

116 Using the transition and emission probability distributions (see Methods) and the HMM structure, we  
117 determine the likelihood function (Equation 13 in Methods) describing the probability of observing a par-  
118 ticular set of lineage frequency time-series data given the unknown parameters, namely the scaled effective  
119 population size across time  $\tilde{N}_e(t)$  and the strength of measurement noise across time  $c_t$ . We then maximize  
120 the likelihood over the parameters to determine the most likely parameters that describe the data. Because  
121 we are relying on a time-series signature in the data for the inference, we need to use a sufficiently large  
122 number of timesteps of data, but on the other hand, the longer the time series, the more parameters would  
123 need to be inferred (since both  $\tilde{N}_e(t)$  and  $c_t$  are allowed to change over time). To balance these two factors,  
124 we assumed that the effective population size stays constant over a time period of 9 weeks (a form of “reg-  
125 ularization”). We then shift this window of 9 weeks across time to determine how  $\tilde{N}_e(t)$  changes over time  
126 (see Methods), but this effectively averages the inferred  $\tilde{N}_e(t)$  over time.  $c_t$  is still allowed to vary weekly.

127 To validate our model, we ran Wright-Fisher simulations with time-varying effective population size and  
128 time-varying measurement noise (Figure 1c-f). Because a substantial number of lineages would go extinct  
129 over the simulation timescale of 100 weeks, we introduced new lineages with a small rate (a rate of 0.01 per  
130 week per individual of “mutating” to start a new lineage) to prevent the number of lineages from becoming  
131 too low. We then did inference on the simulated time-series frequency trajectories (Figure 1d). The inferred  
132  $\tilde{N}_e(t)$  and  $c_t$  closely follow the true values (Figure 1e-f), and the 95% confidence intervals (see Methods  
133 for how they are calculated) include the true value in a median (across timepoints) of 95% of simulation  
134 realizations (Figure S5). The error in  $c_t$  is higher when the variance contributed to the frequency trajectories  
135 by measurement noise is lower than that of genetic drift, which occurs when the effective population size  
136 is low or number of sequences is high (more clearly seen in Figure S6, where the effective population size  
137 is held constant). However, the error on  $\tilde{N}_e(t)$  seems to be unchanged or even slightly decrease when the  
138 error on  $c_t$  is increased because the contribution to the variance due to genetic drift is higher. We also  
139 observe that the inferred  $\tilde{N}_e(t)$  is smoothed over time due to the assumption of constant  $\tilde{N}_e(t)$  over 9 weeks  
140 (Figure S7); this is a potential drawback when there are sharp changes in the effective population size over  
141 time. Importantly, we observed that the inferred  $\tilde{N}_e(t)$  will be underestimated if sampling is assumed to be  
142 uniform when it is actually overdispersed (Figure 1e). This is because variance in the frequency trajectories  
143 due to measurement noise is incorrectly being attributed to genetic drift. The underestimation is strongest  
144 when the variance contributed due to measurement noise is high, either due to high measurement noise  
145 overdispersion, a low number of sampled sequences, or a high effective population size. In this situation,  
146 joint inference of measurement noise and  $\tilde{N}_e(t)$  from the data is necessary for accurate inference of  $\tilde{N}_e(t)$ .

147 In summary, we developed a method to infer the strength of genetic drift and measurement noise from  
148 lineage frequency time series data and validated the accuracy of the method with simulations.

## 149 Application to COG-UK data in England

150 We next applied this method to study the effective population size and strength of measurement noise for  
151 SARS-CoV-2 in England. Because our method assumes that lineages are neutral with respect to one another  
152 (no selection), we performed separate analyses on groups of lineages that have been shown to exhibit fitness

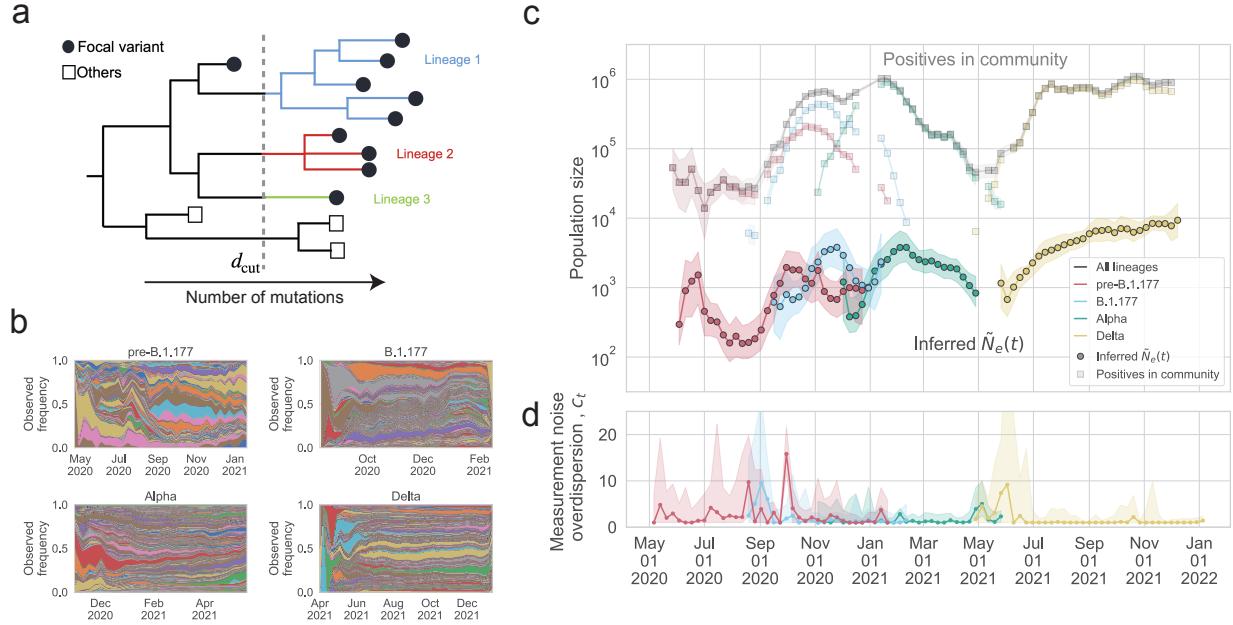


Figure 2: The inferred effective population size and overdispersion of measurement noise in England compared with the number of positive individuals. (a) Schematic of lineage construction for B.1.177, Alpha, and Delta from the COG-UK phylogenetic tree. The filled circles represent the sequences of a focal variant sampled in England, while the unfilled squares represent other variants or sampled in other countries. The phylogenetic tree is cut at a certain depth  $d = d_{\text{cut}}$ , and each branch cut by the line  $d = d_{\text{cut}}$  defines a lineage. Lineages pre-B.1.1.7 are defined using the pango nomenclature [29, 30]. (b) Muller plot of lineage frequency time series for lineages pre-B.1.177, of B.1.177, of Alpha, of Delta. (c) Inferred scaled effective population size ( $\tilde{N}_e(t) \equiv N_e(t)\tau(t)$ ) for pre-B.1.177 sequences, B.1.177, Alpha, and Delta, compared to the estimated number of people testing positive for SARS-CoV-2 in England at the community level, as measured by the COVID-19 Infection Survey [31], for all lineages and by variant or group of lineages. To simplify the plot, only data where the number of positive individuals for a given variant or group of lineages was higher than  $10^3$  in a week are shown. The inferred  $\tilde{N}_e(t)$  is considerably lower than the number of positive individuals for all times and for all variants or group of lineages. (d) Inferred measurement noise overdispersion ( $c_t$ ) for pre-B.1.177 sequences, B.1.177, Alpha, and Delta.

153 differences or deterministic changes in frequency: lineages pre-B.1.177, B.1.177, Alpha, and Delta [28, 17, 154 32, 33]. We did not find any studies in the literature claiming detectable fitness differences between lineages 155 within each of these groups; thus, we assumed that our neutral model should be valid when analyzing lineages 156 only within a single group. We checked that this assumption is valid and describe the results below.

157 To obtain lineage frequency time series data for SARS-CoV-2 in England, we downloaded genomic meta- 158 data from the COVID-19 Genomics UK Consortium (COG-UK) [34] (Figure 2b) and the associated phy- 159 logenetic trees that were created at different points in time. To minimize potential bias, we used only 160 surveillance data (labeled as “pillar 2”). For sequences pre-B.1.177, we used the pangolin lineages assign- 161 ments from COG-UK [29, 30]. However, B.1.177, Alpha, and Delta were subdivided into few or only one 162 pangolin lineage, since a new lineage is defined by sufficiently many mutations and evidence of geographic 163 importation. However, these requirements are not important for our purposes and instead we only need reso- 164 lution of neutral lineages within a variant. Thus, we created our own neutral lineages by grouping sequences 165 together based on phylogenetic distance in the tree (see Figure 2a and Methods), and cutting the tree at a 166 particular point. Most sequenced samples were included in the trees (Figure S8), and any downsampling was 167 done by preserving genetic diversity. Most sequences in the tree were assigned to lineages (see Methods), 168 and we corrected for the fraction of sequences that were not assigned to lineages in our inference of  $\tilde{N}_e(t)$

169 (see Methods). This yielded 486 lineages for pre-B.1.177, 4083 lineages for B.1.177, 6225 lineages for Alpha,  
 170 24867 lineages for Delta.

171 The inferred effective population size is shown in Figure 2c. The inferred effective population size was  
 172 lower than the number of positive individuals in the community by a factor of 16 to 1055 at different  
 173 points in time. The most notable differences between the changes over time in the number of positives  
 174 in the community and that of the effective population size were: the inferred effective population size of  
 175 lineages pre-B.1.177 peaked slightly before the number of pre-B.1.177 positives peaked, the inferred effective  
 176 population size of Alpha decreased slower than the number of positives decreased after January 2021, and the  
 177 shoulder for the inferred effective population size of Delta occurred earlier than in the number of positives.  
 178 We checked that the inferred effective population size is not sensitive to the depth at which the trees are cut  
 179 to create lineages (Figure S9), the threshold counts for creating superlineages (Figure S10), or the number  
 180 of weeks in the moving time window (Figure S11).

181 The inferred measurement noise for each group of lineages is shown in Figure 2d. The inferred measure-  
 182 ment noise overdispersion was mostly indistinguishable from 1 (uniform sampling), but at times was above 1  
 183 (sampling that is more variable than uniform sampling). There were also at times differences in the strength  
 184 of measurement noise between variants when they overlapped in time. In particular, measurement noise for  
 185 lineages pre-B.1.177 peaked in October 2020 despite measurement noise being low for B.1.177 at that time.

186 To better understand the observed levels of genetic drift, we compared the inferred  $\tilde{N}_e(t)$  to that of an  
 187 SIR null model, which includes a susceptible, infectious, and recovered class. The  $\tilde{N}_e(t)$  for an SIR model  
 188 was derived in Ref. [35, 36, 37] and is given by

$$\tilde{N}_e^{\text{SIR}}(t) = \frac{I(t)}{2R_t\gamma_I} \quad (1)$$

where  $I(t)$  is number of infectious individuals,  $R_t$  is the effective reproduction number, and  $\gamma_I$  is the rate at  
 which infectious individuals recover. For the number of infectious individuals, we used the number of positive  
 individuals estimated from the UK Office for National Statistics' COVID-19 Infection Survey [31], which is  
 a household surveillance study that reports positive PCR tests, regardless of symptom status. We used the  
 measured effective reproduction number in England reported by the UK Health Security Agency [38]. We  
 used  $\gamma_I^{-1} = 5.5$  days [39, 40], and our results are robust to varying  $\gamma_I$  within a realistic range of values  
 (Figure S12). We found that  $\tilde{N}_e^{\text{SIR}}(t)$  is very similar to the number of positives because the effective  
 reproduction number in England was very close to 1 across time and  $\gamma_I$  is also very close to 1 in units of  
 weeks<sup>-1</sup>. To calculate  $\tilde{N}_e^{\text{SIR}}(t)$  for each variant or group of lineages, we rescaled the population-level  $I(t)$  and  
 $R_t$  based on the fraction of each variant in the population and the relative differences in reproduction numbers  
 between variants (see Methods). We then calculated the scaled true population size,  $\tilde{N}(t) \equiv N(t)\tau(t)$ , for  
 the SIR model by multiplying by the variance in offspring number,  $\sigma^2$ , for the SIR model [41]

$$\tilde{N}^{\text{SIR}}(t) = \tilde{N}_e^{\text{SIR}}(t)\{\sigma^2\}^{\text{SIR}} \quad (2)$$

$$\{\sigma^2\}^{\text{SIR}} = 2. \quad (3)$$

189 Overall, the inferred  $\tilde{N}_e(t)$  is lower than  $\tilde{N}^{\text{SIR}}(t)$  by a time-dependent factor that varies between 16 and  
 190 589 (Figures 3c and S13), suggesting high levels of genetic drift in England across time. We find similar  
 191 results when using an SEIR rather than an SIR model which additionally includes an exposed class and  
 192 may be more realistic (Methods, Supplementary information, and Figure S14). The ratio of  $\tilde{N}^{\text{SIR}}(t)$  to the  
 193 inferred  $\tilde{N}_e(t)$  was similar across variants and across time, except that for Alpha the ratio initially peaked  
 194 and then decreased over time.

195 Because non-neutral lineages could potentially bias the inferred effective population size to be lower  
 196 in a model that assumes all lineages are neutral, we checked the assumption that lineages are neutral with  
 197 respect to one another within a group or variant (pre-B.1.177, B.1.177, Alpha, and Delta) using methods that  
 198 detect deterministic changes in lineage frequency. We used two methods: a more conservative, deterministic  
 199 method that ignores genetic drift, and a more accurate method that accounts for time-varying genetic drift.  
 200 To detect any lineages that could possibly be non-neutral, we used the conservative method, and found that  
 201 about 20% of lineages in each group were significantly non-neutral at a significance level of 5%. Very likely,  
 202 some of these lineages are detected as non-neutral simply because the model does not correctly account for

203 strong genetic drift. Excluding these lineages from the analysis of the inferred effective population size leads  
204 to slightly different values of the effective population size, but by less than an order of magnitude and mostly  
205 by less than the 95% confidence intervals (Figure S15). This result shows that conservatively excluding  
206 lineages that could be non-neutral does not change the result that the inferred effective population size is  
207 one to two order of magnitudes lower than the SIR or SEIR model effective population size. Additionally,  
208 when we used these new estimates of effective population size in the more accurate method for inferring  
209 fitness coefficients that accounts for time-varying effective population size, we find that only about 1% of  
210 lineages are non-neutral at a significance level of 5%, which further supports that lineages under constant  
211 selection are unlikely to explain our results.

212 We also probed the spatial structure of transmission by inferring the scaled effective population size  
213 separately for each region within England. We find that the scaled effective population size in the regions of  
214 England is substantially smaller than that in England as a whole for B.1.177, Alpha, and Delta (Figure S1),  
215 suggesting that the transmission was not well-mixed at that time. Additionally, the discrepancy between  
216 the inferred regional scaled effective population size and the observed number of positive individuals in a  
217 region was comparable to that seen in England as a whole (Figure S3), which is consistent with spatially  
218 segregated dynamics with similar levels of genetic drift in each region. We further describe these results in  
219 the Supplementary Information.

## 220 Potential mechanisms that can contribute to the high levels of genetic drift

221 Two potential mechanisms that can contribute to the observed high levels of genetic drift are: (1) variability  
222 at the individual level through superspreading (Figure 3a), and (2) host population structure (Figure 3b).  
223 We investigate each of these mechanisms in turn and compare it to our results. While in reality, both  
224 mechanisms (and others not explored here) are likely at play, it is challenging to tease them apart given our  
225 limited data. Therefore, we consider the extreme situations where only one mechanism at a time is driving  
226 the dynamics.

227 Superspreading occurs due to overdispersion in the number of secondary cases, which decreases the  
228 effective population size. If superspreading were the only mechanism at play, then the variance in offspring  
229 number that would explain our results would be the same as the ratio between the SIR null model  $\tilde{N}^{\text{SIR}}(t)$   
230 and the inferred  $\tilde{N}_e(t)$  (16-589) (Figure 3c). Current estimates of the variance in offspring number measured  
231 by contact tracing and modeling across a wide range of times and locations are from around 0.7 in one  
232 study to 65 in another (Table S1). We found two studies that apply to the UK, one which used a model  
233 that incorporated the empirical viral load trajectories and contact numbers to estimate superspreading [43]  
234 and another which used a branching process model of the number of imported and local cases [42]. The  
235 literature estimates whose time window overlapped with our time windows found substantially lower levels  
236 of superspreading than what we observe (Figure 3c and corresponding overdispersion parameter shown in  
237 Figure S16). It is possible that contact tracing and modeling over- or under-estimates overdispersion due to  
238 missed contacts [20]. However, on the other hand, it may be the case that superspreading is not the only  
239 mechanism at play.

240 Host deme structure is another mechanism that can lead to decreased effective population size. In  
241 such a model, individuals within a deme are very well-connected to one another (i.e. households or friend  
242 groups, also known as “communities” in network science [44]), but there are few connections between demes  
243 (Figure 3c). It is possible for deme structure to occur without superspreading. For instance, in the schematic  
244 in Figure 3c, the number of contacts is either 4 or 5; if every contact led to a transmission, this would be  
245 an extremely narrow offspring number distribution (i.e. no superspreading). Because individuals are very  
246 well-connected within a deme, once the pathogen spreads to a susceptible deme, it will spread rapidly in a  
247 deme until all individuals are infected (a jackpot event). In this way, deme structure can lower the effective  
248 population size by lowering the effective number of stochastic transmissions. For instance, in the example  
249 in Figure 3c, there are 20 individuals, but only 3 potential stochastic transmissions. Deme structure may  
250 also arise from correlations in the number of secondary infections over a series of hosts (i.e. a series of high  
251 numbers of secondary infections in a transmission chain, or conversely low numbers of secondary infections  
252 in a transmission chain) [45]. This may arise, for instance, if individuals in a transmission chain have similar  
253 behavior, due to geographical proximity, or similar value systems on risk aversion. A recent study has found  
254 that individuals infected by superspreading tend to be superspreaders themselves more than expected by

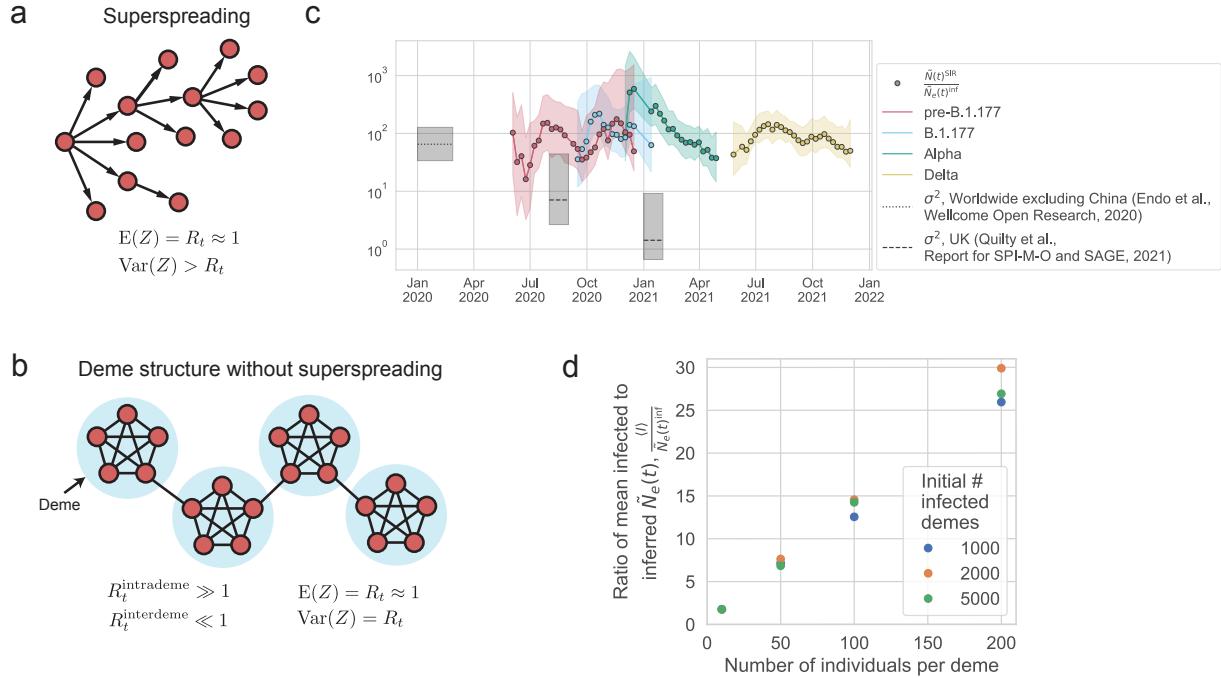


Figure 3: Potential mechanisms that can generate a low effective population size. (a) Superspreading, where the distribution of the number of secondary cases ( $Z$ ) from a single infected individual is broadly distributed (variance greater than mean). (b) Deme structure without superspreading, due to heterogeneity in the host network structure, where the distribution of the number of secondary cases is not broadly distributed (variance approximately equal to mean). (c) The ratio between the  $\tilde{N}^{\text{SIR}}(t)$  (the scaled population size calculated from an SIR model using the number of observed positive individuals and the observed effective reproduction number) and the inferred  $\tilde{N}_e(t)$  for each variant. Only data where the error in the SIR model  $\tilde{N}^{\text{SIR}}(t)$  is less than 3 times the value are shown, because larger error bars make it challenging to interpret the results. The inferred  $\tilde{N}_e(t)$  is lower than the  $\tilde{N}^{\text{SIR}}(t)$  (which assumes well-mixed dynamics and no superspreading) by a factor of 16 to 589, indicating high levels of genetic drift. The variance in offspring number from the literature [42, 43] does not entirely explain the discrepancy between the true and effective population sizes. (d) Simulations of deme structure without superspreading can generate high levels of genetic drift via jackpot events. SEIR dynamics are simulated within demes (with  $R_t = 10$ , i.e. deterministic transmission) and Poisson transmission is simulated between demes ( $R_t \ll 1$ , i.e. stochastic transmission) such that the population  $R_t \sim 1$  (see Methods). Simulation parameters are: mean transition rate from exposed to infected  $\gamma_E = (2.5 \text{ days})^{-1}$ , mean transition rate from infected to recovered  $\gamma_I = (6.5 \text{ days})^{-1}$ , total number of demes  $D_{\text{total}} = 5.6 \times 10^5$ . The ratio between the number of infected individuals and the inferred effective population size is found to scale linearly with the deme size and not with the number of infected demes. This scaling results because of jackpot events where a lineage that happens to infect a susceptible deme grows rapidly until all susceptible individuals in the deme are infected.

255 chance [46], which would be consistent with this phenomenon.

256 To check our intuition that deme structure can decrease the effective population size and increase genetic  
 257 drift, we ran simulations of a simplified deme model (see Methods): all demes have the same number of  
 258 individuals, and there is a sufficiently large enough number of demes that the total number of demes does  
 259 not matter. Initially some number of demes are infected, and transmission occurs such that the overall  
 260 effective reproduction number in the population is around 1. From our simulations, we find that when the  
 261 number of individuals in a deme increases, the ratio between the number of infected individuals and the  
 262 inferred effective population size increases (Figure 3d); in other words, the more individuals there are in a

263 deme, the higher the level of genetic drift we observe compared what is expected from the number of infected  
264 individuals. This is because while the number of infected individuals increases when the deme size increases  
265 (Figure S17a), the inferred effective population size (and thus the level of stochasticity) stays the same as a  
266 function of deme size (it is more dependent on the number of infected demes) (Figure S17b). However, the  
267 exact ratio of the number of infected individuals to the inferred effective size depends on the parameters of  
268 the model.

## 269 Discussion

270 Here, we systematic studied the strength of genetic drift of SARS-CoV-2 in England across time and spatial  
271 scales. To do this, we developed and validated a method for jointly inferring time-varying genetic drift  
272 and overdispersed measurement noise using lineage frequency time series data (Figure 1), allowing these  
273 two effects to be disentangled, which overcomes a major challenge in the ability to infer the strength of  
274 genetic drift from time-series data. We find that the effective population size of SARS-CoV-2 in England  
275 was lower than that of an SIR null model true population size (using the observed number of positives)  
276 by a time-dependent factor ranging from 16 to 589 (Figure 3c), suggesting that there were higher levels of  
277 genetic drift than expected from uniform transmission. We also find evidence for spatial structure in the  
278 transmission dynamics during the B.1.177, Alpha, and Delta waves, as the inferred  $\tilde{N}_e(t)$  was substantially  
279 lower in regions compared to that of all England (Figure S1). These findings are consistent with other studies  
280 that have found spatial structure in transmission of B.1.177 [47], Alpha [48], and Delta [49].

281 The levels of genetic drift that we observe are higher than literature values of superspreading; while this  
282 may be partly due to the challenges of accurately estimating the extent of superspreading from data, it also  
283 suggests that additional mechanisms may be leading to increased stochasticity. In particular, we explore a  
284 simplified deme model with groups of individuals that are well-connected to each other (demes) but not to  
285 individuals from other demes, and find that such a simple model can generate a low effective population  
286 size even in the absence of superspreading, due to jackpot events. In reality, both superspreading and host  
287 structure are likely at play. Additionally, they could interact with each other. For instance, there could be  
288 superspreading within a deme. Future work can try to tease apart the contribution of these two mechanisms,  
289 which for instance may be possible with better transmission network data, building on previous work on  
290 transmission networks [50], or with time-resolved contact tracing data [20]. This will be important because  
291 the relative contributions of the two mechanisms of superspreading and host population structure to genetic  
292 drift can affect the establishment of new variants in the population [3].

293 Accurately estimating the strength of genetic drift allows us to better understand disease spread and ex-  
294 tinction, as well as to better parameterize evolutionary models and understand how mutations will establish  
295 in the population. We observed that with a few exceptions, the amount by which genetic drift was elevated  
296 compared to the number of positives did not change much over time or across variants (Figure 3c), despite  
297 changes in lockdowns and restrictions (which we may expect to decrease behavior that leads to superspread-  
298 ing). On the other hand, this may not be so surprising given the findings that restrictions affect the mobility  
299 network structure in a complex way, decreasing some types of mobility while increasing others [51]. One  
300 exception was that Alpha had significantly higher genetic drift compared to Delta and the strength of genetic  
301 drift in Alpha first peaked then slowly decreased over time. This may be either due to differences in the  
302 properties of the virus or differences in host behavior. For instance, it may suggest that the stochasticity  
303 in the transmission of Alpha sharply increased then slowly decreased over time. Alternatively, this may  
304 be driven by Alpha's expanding geographic range combined with reimported cases of Alpha into the UK  
305 (observed from February 2021 onwards), which could both also decrease the effective population size [52].

306 We observe that measurement noise of SARS-CoV-2 is mostly indistinguishable from uniform sampling,  
307 but data from some variants at some times do exhibit more elevated measurement noise than uniform  
308 sampling. Thus, we expect that assuming uniform sampling, as many methods do, or constant overdispersion  
309 will lead to accurate estimates for this dataset [27, 22, 23, 28]. The number of SARS-CoV-2 sequences from  
310 England is extremely high and sampling biases are expected to be low, because of efforts to reduce sampling  
311 biases by sampling somewhat uniformly from the population through the COVID-19 Infection Survey [31]  
312 (from which a subset of positives are sequenced and included in the COG-UK surveillance sequencing data  
313 that we use). On the other hand, other countries may have higher sampling biases, so jointly estimating

314 measurement noise and genetic drift may be more crucial in those settings. It may also be interesting to use  
315 this method to test whether genomics data taken from wastewater has lower levels of measurement noise as  
316 compared to sequenced cases.

317 We find that constant selection is unlikely to explain our results, as excluding potentially non-neutral  
318 lineages does not significantly change the inferred effective population size. Our method allows for the  
319 inference of constant selection while accounting for time-varying effective population size, and thus may be  
320 interesting in future applications of tests for selection. However, the model ignores other processes that  
321 may lead to deterministic changes in frequency, such as migration and mutation (discussed below); thus, we  
322 recommend combining it with complementary methods (for instance, phylogenetic methods) that test for  
323 selection as somewhat independent tests for selection.

324 In summary, we find that the strength of genetic drift in SARS-CoV-2 transmission in England is higher  
325 than expected based on the number of positive individuals and the levels of superspreading reported in the  
326 literature. Our results suggest that additional models and methods will be needed to better understand the  
327 mechanisms behind the observed elevated levels of genetic drift.

## 328 Limitations of the study and opportunities for future directions

329 First, in our analysis, we focused on standing variation that existed at a particular depth in the phylogenetic  
330 tree and ignored de novo mutations subsequently arising during the time series. However, we don't think this  
331 should substantially affect our results because introducing mutations in the form of new lineages with a small  
332 rate in the simulations did not have a large effect on the method performance (Figure 1e). We also ignored  
333 importation of SARS-CoV-2 into England and exportation of SARS-CoV-2 out of England. Migration can  
334 substantially change frequencies that are locally rare, but we expect importations to only weakly influence  
335 the frequency fluctuations of abundant variants, on which we have focused in this work. Additionally, we  
336 did not test for more complex forms of selection, such as fluctuating selection. More generally, future work  
337 should explore joint inference of selection, migration, and/or mutation in the model, as is appropriate for  
338 the pathogen of interest, building on previous work in this area [53, 54, 26].

339 Second, there may be biases in the way that data are collected that are not captured in our model. While  
340 our method does account for sampling biases that are uncorrelated in time, sampling biases that remain over  
341 time cannot be identified as such (i.e. if one geographical region was dominated by a particular lineage and it  
342 consistently had higher sequencing rates compared to another geographical region), and this can potentially  
343 bias the inferred effective population size; although, this is also a problem in phylogenetic methods. One  
344 approach to this problem that was utilized by some early methods during the pandemic is to develop sample  
345 weights based on geography, time, and number of reported cases. Future work should study the effect of  
346 different sampling intensities between regions on uncorrelated and correlated sampling noise. Additionally,  
347 we assume that the measurement noise overdispersion is identical for all lineages within a variant; in reality,  
348 there may be differences in sampling between lineages. Future work should explore whether this is the case  
349 in the actual dataset, as well as the effect of lineage-specific measurement noise overdispersion on overall  
350 method performance.

351 Third, the quantity of effective population size is a summary statistic that is influenced by many factors,  
352 making its interpretation challenging. The effective population size describes the population size under a  
353 well-mixed Wright-Fisher model, whereas in reality, this assumption is broken by many effects, including  
354 host structure and broad offspring number distributions. Thus, in our study we are careful to interpret  
355 effective population size only in the broadest terms of genetic drift, without being able to determine what  
356 mechanisms lead to the inferred effective population size (although we do explore some possibilities above).  
357 While within-host dynamics may in principle impact the lineage frequency trajectories, this effect is likely  
358 small for our analysis because we focus on acute infections (infections in the community rather than in  
359 hospitals and nursing homes). This is because acute infections of SARS-CoV-2 are thought to generate little  
360 within-host diversity that is passed on due to the short infection duration and small bottleneck size between  
361 hosts [55, 56]; while new mutations arising within acute hosts have been observed to be transmitted, these  
362 events are rare [55].

363 Fourth, the use of a sliding window of 9 weeks on the lineage frequency data will lead to smoothing of  
364 sharp changes in effective population size. It may be interesting in future work to develop a continuous  
365 method that uses a prior to condition on changes in effective population size, similar to those that have been

366 developed for coalescence-based methods [1, 57]. This would allow us to infer continuous changes in effective  
367 population size without needing to use a sliding window.

368 While we have focused on SARS-CoV-2 in this study, the method developed here can be extended to  
369 study genetic drift in other natural populations that are influenced by measurement noise and where genomic  
370 frequency data are available, for instance other pathogens, field studies, and ancient DNA [58, 59, 60]. More  
371 generally, ongoing methods development that integrates genomics, epidemiological, and other data sources  
372 is crucial for being able to harness the large amounts of data that have been generated to better understand  
373 and predict evolutionary dynamics.

## 374 Materials and Methods

### 375 Data sources and processing

376 We downloaded sequence data from the COVID-19 Genomics UK Consortium (COG-UK) [34]. We only  
377 used surveillance data (labeled as “pillar 2”); this dataset is composed of a random sample of the positive  
378 cases from the COVID-19 Infection Survey, which is a surveillance study of positive individuals in the  
379 community administered by the Office for National Statistics (see below). For lineages that appeared before  
380 B.1.177, we downloaded the metadata from the COG-UK Microreact dashboard [61], which included the  
381 time and location of sample collection (at the UTLA level), as well as the lineage designation using the  
382 Pango nomenclature [29, 30]. For B.1.177, Alpha, and Delta sequences, because the Pango nomenclature  
383 classified them into very few lineages, we created our own lineages from the phylogenetic trees (see below).  
384 We downloaded the publicly available COG-UK tree on February 22, 2021 for B.1.177; June 20, 2021 for  
385 Alpha; and January 25, 2022 for Delta. We also downloaded the COG-UK metadata for all lineages on  
386 January 16, 2022, which included the time and location (at the UTLA level) of sample collection. For the  
387 data of B.1.177, Alpha, and Delta, the data was deduplicated to remove reinfections in the same individual  
388 by the same lineage, but reinfections in the same individual by a different lineage were allowed. This yielded  
389 a total of 490,291 sequences.

390 The lineage frequency time-series is calculated separately for each variant or group of lineages (pre-  
391 B.1.177, B.1.177, Alpha, and Delta). First, the sequence metadata are aggregated by epidemiological week  
392 (Epiweek) to average out measurement noise that may arise due to variations in reporting within a week.  
393 Then, the lineage frequency is calculated by dividing the number of sequences from that lineage in the  
394 respective tree by the total number of sequences of that variant (or group of lineages) that were assigned to  
395 any lineage in the respective tree.

396 Because our model describes birth-death processes when the central limit theorem can be applied, we need  
397 the lineage frequencies to be sufficiently high. Thus, we randomly combine rare lineages into “superlineages”  
398 that are above a threshold number of counts and threshold frequency in the first and last timepoint of each  
399 trajectory. For the threshold, we chose of 20 counts and frequency of 0.01. Sensitivity analyses showed  
400 that the choice of the superlineage count threshold does not substantially affect the results (Figure S10).  
401 Superlineages are non-overlapping (i.e. each sequence belongs to exactly one superlineage).

402 The estimated number of people testing positive for COVID-19 in England and each region of England  
403 was downloaded from the UK Office for National Statistics’ COVID-19 Infection Survey [31]. The COVID-19  
404 Infection Survey includes households that are semi-randomly chosen, and individuals are tested regardless  
405 of whether they are reporting symptoms. Infections reported in hospitals, care homes, and other communal  
406 establishments are excluded. Thus the dataset provides a representative number of positive individuals in  
407 the community setting. The reported date of positive cases is the date that the sample was taken. The error  
408 on the number of positive individuals from April 17, 2020 to July 5, 2020 is reported as the 95% confidence  
409 interval, and after July 5, 2020 is reported as the 95% credible interval. The regional data reported the  
410 positivity rate over two week intervals. To get the number of positives, we multiplied by the number of  
411 individuals in the community setting in the region (excluding hospitals, care homes, and other communal  
412 establishments). As the data was reported over two week intervals, we obtained the number of positives for  
413 each week using linear interpolation.

414 The observed effective reproduction numbers for England and each region of England were downloaded  
415 from the UK Health Security Agency [38]. Only times where the certainty criteria are met and the inference  
416 is not based on fewer days or lower quality data are kept. The error on the effective reproduction number is  
417 reported as the 90% confidence interval. Although not reported in the dataset, we choose the point estimate  
418 of the effective reproduction number to be the midpoint between the upper and lower bounds of the 90%  
419 confidence interval.

### 420 Creating lineages in B.1.177, Alpha, and Delta

421 For B.1.177, Alpha, and Delta, we divided each of them into neutral lineages based on phylogenetic distance.  
422 Specifically, for B.1.177 and Alpha, we cut a phylogenetic tree (in units of number of mutations) at a certain  
423 depth,  $d = d_{\text{cut}}$ . Each of the internal or external branches that are cut by the line  $d = d_{\text{cut}}$  defines a lineage  
424 (Figure 2a). The (observed) frequency of a lineage at a given time point in England was computed by

425 counting the number of England sequences (leaf nodes) belonging to the lineage and by normalizing it by the  
 426 total number of sequences in all assigned lineages of the focal variant in England at that time point. Lineage  
 427 frequencies at the regional level were similarly computed by counting the number of sequences separately for  
 428 each region.

429 The choice of  $d_{\text{cut}}$  is arbitrary to some extent. Because we wanted a sufficiently high resolution of lineages  
 430 from the early phase of spreading of a variant and because the evolutionary distance correlates with the actual  
 431 sample date (Figure S18), for each focal variant, we chose the depth  $d_{\text{cut}}$  that roughly corresponds to the  
 432 time point when it began to spread over England.

433 For the Delta variant, the sequences form two distinct groups along the depth direction, as seen from the  
 434 last panel of Figure S18. Therefore, to divide the Delta variant into lineages with small frequencies, we cut  
 435 the phylogenetic tree at two depths sequentially; we first cut the tree at  $d_{\text{cut}}^{(1)}$ , which resulted in lineages with  
 436 small frequencies plus a lineage with  $\mathcal{O}(1)$  frequency. Then, to divide the latter lineage further, we took the  
 437 subtree associated with this lineage and cut the subtree at  $d_{\text{cut}}^{(2)}$ .

438 For the results presented in the main text, we used (in units of substitutions per site, with the reference  
 439  $d=0$  being the most recent common ancestor)  $d_{\text{cut}} = 2.323 \cdot 10^{-2}$  for B.1.177,  $d_{\text{cut}} = 2.054 \cdot 10^{-3}$  for Alpha,  
 440 and  $d_{\text{cut}}^{(1)} = 1.687 \cdot 10^{-3}$  and  $d_{\text{cut}}^{(2)} = 1.954 \cdot 10^{-3}$  for Delta. We confirmed that our results are robust to the  
 441 choice of  $d_{\text{cut}}$  as well as the choice of the phylogenetic tree data we used (Figure S9).

## 442 Model for inferring effective population size from lineage frequency time series

443 We use a Hidden Markov Model with continuous hidden and observed states to describe the processes of  
 444 genetic drift and sampling of cases for sequencing (similar to a Kalman filter) (Figure 1A). The hidden states  
 445 describe the true frequencies of the lineages and the observed states describe the observed frequencies of the  
 446 lineages as measured via sequenced cases.

447 The transition probability between the true frequencies  $f_t$  (the hidden states) due to genetic drift when  
 448  $0 \ll f \ll 1$  has been shown in [62] to be well-described by the following expression, which we use as our  
 449 transition probability,

$$450 p(f_{t+1}|f_t, \tilde{N}_e(t)) = \frac{1}{2} \sqrt{\frac{2f_t^{1/2}}{\pi f_{t+1}^{3/2}(\tilde{N}_e(t))^{-1}}} \exp\left(-\frac{2(\sqrt{f_{t+1}} - \sqrt{f_t})^2}{(\tilde{N}_e(t))^{-1}}\right). \quad (4)$$

451  $\tilde{N}_e(t) \equiv N_e(t)\tau(t)$  where  $N_e(t)$  is the time-dependent effective population size and  $\tau(t)$  is the time-dependent  
 452 generation time, which is defined as the mean time between two subsequent infections per individual (i.e.  
 453 the time between when an individual becomes infected and infects another individual, or the time between  
 454 two subsequent infections caused by the same individual). This transition probability gives the correct first  
 455 and second moments describing genetic drift when  $f \ll 1$ ,  $E(f_{t+1}|f_t) = f_t$  and  $\text{Var}(f_{t+1}|f_t) = \frac{f_t}{\tilde{N}_e(t)}$ , and  
 456 is a good approximation when the central limit theorem can be applied, which is the case when  $f \gg 0$ .  
 457 By assuming that  $f_{t+1} \approx f_t$ , and defining  $\phi_t \equiv \sqrt{f_t}$ , Equation 4 can be approximated as a simple normal  
 458 distribution

$$459 p(\phi_{t+1}|\phi_t, \tilde{N}_e(t)) = \mathcal{N}\left(\phi_t, \frac{1}{4\tilde{N}_e(t)}\right). \quad (5)$$

460 We describe the emission probability from the true frequency  $f_t$  to the observed frequency  $f_t^{\text{obs}}$  (the  
 461 observed states), defining  $\phi_t^{\text{obs}} \equiv \sqrt{f_t^{\text{obs}}}$ , as

$$462 p(\phi_t^{\text{obs}}|\phi_t, c_t) = \mathcal{N}\left(\phi_t, \frac{c_t}{4M_t}\right) \quad (6)$$

463 where  $M_t$  is the number of input sequences. Again, this distribution is generically a good description when  
 464 the number of counts is sufficiently large, due to the central limit theorem. The first and second moments  
 465 of this emission probability are  $E(f_t^{\text{obs}}|f_t) = f_t$  and  $\text{Var}(f_t^{\text{obs}}|f_t) = \frac{c_t}{M_t} f_t$ , or equivalently considering the  
 466 number of sequences  $n_t^{\text{obs}} = f_t^{\text{obs}} M_t$  and the true number of positive individuals  $n_t$ ,  $E(n_t^{\text{obs}}|n_t) = n_t$  and  
 467  $\text{Var}(n_t^{\text{obs}}|n_t) = c_t n_t$ . Thus,  $c_t$  describes the strength of measurement noise at time  $t$ . When  $c_t = 1$ ,  
 468 the emission probability approaches that describing uniform sampling of sequences from the population

466 of positive individuals (i.e. can be described by a Poisson distribution in the limit of a large number of  
 467 sequences), namely  $\text{Var}(n_t^{obs}|n_t) = n_t$  or equivalently  $\text{Var}(f_t^{obs}|f_t) = \frac{f_t}{M_t}$ . This is the realistic minimum  
 468 amount of measurement noise. When  $c_t > 1$ , it describes a situation where there is bias (that is uncorrelated  
 469 in time) in the way that sequences are chosen from the positive population. The case of  $0 < c_t < 1$  describes  
 470 underdispersed measurement noise, or noise that is less random than uniform sampling. The case of  $c_t = 0$   
 471 describes no measurement noise (for instance, when all cases are sampled for sequencing). These last two  
 472 situations are unlikely in our data, and thus as we describe below, we constrain  $c_t \geq 1$  in the inference  
 473 procedure. In addition to being a good description of measurement noise, defining the emission probability  
 474 in the same normal distribution form as the transmission probability allows us to easily derive an analytical  
 475 likelihood function, described below (Note: see Ref. [26] for a method to derive an analytical likelihood  
 476 function for arbitrary forms of the transition and emission probabilities).

We derive the likelihood function (up to a constant) for the Hidden Markov Model using the forward algorithm, although it can alternatively be derived by marginalizing over all hidden states. We assume an (improper) uniform prior on  $\phi_0$  (i.e. no information about the initial true frequency of the lineage).

$$p(\phi_0, \phi_0^{obs}, \theta_0) = p(\phi_0^{obs}|\phi_0, c_0)p(\phi_0) \quad (7)$$

$$p(\phi_0) \propto 1 \quad (8)$$

$$p(\phi_t, \phi_{0:t}^{obs}, \theta_{0:t}) = p(\phi_t^{obs}|\phi_t, c_t) \int_{-\infty}^{\infty} p(\phi_t|\phi_{t-1}, \tilde{N}_e(t))p(\phi_{t-1}, \phi_{0:t-1}^{obs}, \theta_{0:t-1})d\phi_{t-1}, \quad 0 < t \leq T \quad (9)$$

$$p(\phi_{0:T}^{obs}, \theta_{0:T}) = \int_{-\infty}^{\infty} p(\phi_T, \phi_{0:T}^{obs}, \theta_{0:T})d\phi_T \quad (10)$$

$$\mathcal{L}(\vec{\phi}_{0:T}^{obs}|\theta_{0:T}) = \prod_{\alpha} p(\{\phi_{0:T}^{obs}\}_{\alpha}, \theta_{0:T})p(\theta_{0:T}) \quad (11)$$

$$p(\theta_{0:T}) \propto 1 \quad (12)$$

$$\mathcal{L}(\vec{\phi}_{0:T}^{obs}|\theta_{0:T}) = \prod_{\alpha} p(\{\phi_{0:T}^{obs}\}_{\alpha}, \theta_{0:T}). \quad (13)$$

477 where  $\phi_{0:t}^{obs} \equiv \{\phi_0^{obs}, \dots, \phi_t^{obs}\}$ ,  $\theta_{0:t} \equiv \{\tilde{N}_e(0), \dots, \tilde{N}_e(t), c_0, \dots, c_t\}$ , and the subscript  $\alpha$  indicates a particular  
 478 lineage. We use a uniform prior on the parameters. The parameters  $\theta_{0:T}$  are inferred by maximizing the  
 479 likelihood (described below).

The forward algorithm has an analytical form for the simple case of Gaussian transition and emission probabilities. We use the identity for the product of two normal distributions  $N(x, \mu, v)$ , where  $\mu$  is the mean and  $v$  is the variance:

$$N(x, \mu_1, v_1)N(x, \mu_2, v_2) = N(\mu_1, \mu_2, v_1 + v_2)N(x, \mu_{12}, v_{12}) \quad (14)$$

$$\mu_{12}(\mu_1, \mu_2, v_1, v_2) = \frac{\mu_1 v_2 + \mu_2 v_1}{v_1 + v_2} \quad (15)$$

$$v_{12}(v_1, v_2) = \frac{1}{\frac{1}{v_1} + \frac{1}{v_2}}. \quad (16)$$

Solving the forward algorithm recursively, we have

$$p(\phi_{0:T}^{obs}, \theta_{0:T}) = \prod_{i=1}^T N(\phi_i^{obs}, \mu_i, \frac{c_i}{4M_i} + v_i) \quad (17)$$

$$(18)$$

where

$$\mu_1 = \phi_0^{obs} \quad (19)$$

$$v_1 = \frac{\frac{1}{\tilde{N}_e(t)} + \frac{c_0}{M_0}}{4} \quad (20)$$

$$\mu_{i+1} = \mu_{12}(\mu_i, \phi_i^{obs}, v_i, \frac{c_i}{4M_i}) \quad (21)$$

$$v_{i+1} = v_{12}(\frac{c_i}{4M_i}, v_i) + \frac{1}{4\tilde{N}_e(t)}. \quad (22)$$

(23)

480 Equation 17 can be substituted into Equation 13 to obtain the full analytical likelihood function.

## 481 Fitting the model to data

482 We split the time series data into overlapping periods of 9 Epiweeks, over which the effective population size is  
483 assumed to be constant. We first use the moments of the probability distributions combined with least squares  
484 minimization to get an initial guess for the parameters. Then, we perform maximum likelihood estimation  
485 using the full likelihood function. To capture uncertainties that arise from the formation of superlineages  
486 from lineages, we create superlineages randomly 100 times (except where indicated otherwise). We infer the  
487 strength of measurement noise and the effective population size for each superlineage combination (described  
488 below).

## 489 Determining the initial guess for the parameters using method of moments approach

Combining the transition and emission probabilities, and marginalizing over the hidden states we have

$$p(f_j^{obs} | f_i^{obs}) \propto \sqrt{\frac{1}{(f_j^{obs})^{3/2}}} \exp\left(-\frac{2\left(\sqrt{f_j^{obs}} - \sqrt{f_i^{obs}}\right)^2}{\kappa_{i,j}}\right) \quad (24)$$

$$p(\phi_j^{obs} | \phi_i^{obs}) = \mathcal{N}(\phi_i^{obs}, \kappa_{i,j}) \quad (25)$$

$$\kappa_{i,j} \equiv \frac{c_i}{4M_i} + \frac{c_j}{4M_j} + \frac{(j-i)}{4\tilde{N}_e(t)}. \quad (26)$$

490 The first two terms of  $\kappa_{i,j}$  are the contribution to the variance from measurement noise at times  $i$  and  $j$ , and  
491 the third term is the contribution to the variance from genetic drift.

492 We calculate the maximum likelihood estimate of  $\kappa_{i,j}$ ,  $\hat{\kappa}_{i,j}$ , which is simply the mean squared displacement

$$\hat{\kappa}_{i,j} = \left\langle (\phi_j^{obs} - \phi_i^{obs})^2 \right\rangle. \quad (27)$$

493 The standard error is given by

$$\Delta\hat{\kappa}_{i,j} = \sqrt{\frac{\left\langle \left[ (\phi_j^{obs} - \phi_i^{obs})^2 - \hat{\kappa}_{i,j} \right]^2 \right\rangle}{Z}} \quad (28)$$

494 where  $Z$  is the number of superlineages.

495 By looking across all pairs of timepoints  $i$  and  $j$ , we get a system of linear equations in  $\kappa_{i,j}$  that depend  
496 on the parameters  $c_t$  and  $\tilde{N}_e(t)$ . To determine the most likely values of the parameters, we minimize

$$\ln \sum_{i,j} \frac{(\hat{\kappa}_{i,j} - Ac)^2}{\Delta\hat{\kappa}_{i,j}} \quad (29)$$

497 using `scipy.optimize.minimize` with the L-BFGS-B method and the bounds  $1 \leq c_t \leq 100$  and  $1 \leq \tilde{N}_e(t) \leq 10^7$ .

498 While underdispersed measurement noise ( $c_t < 1$ ) is in principle possible, we constrain  $c_t \geq 1$  because

499 realistically, the lowest amount of measurement noise will be from uniform sampling of sequences. An  
500 example of inferred parameters using the methods of moments approach on simulated data is shown in  
501 Figure S19.

502 **Maximum likelihood estimation of the parameters**

For each set of superlineages, we use the inferred measurement noise values ( $c_t$ ) and inferred scaled effective population size from above ( $\tilde{N}_e(t)$ ) as initial guesses in the maximization the likelihood function in Equation 13 over the parameters. For the optimization, we use `scipy.optimize.minimize_scalar` with the Bounded method and the bounds  $1 \leq c_t \leq 100$  and  $1 \leq \tilde{N}_e(t) \leq 10^{11}$ . The time  $t$  in the inferred  $\tilde{N}_e(t)$  is taken to be the midpoint of the 9 Epiweek period. The reported  $\tilde{N}_e(t)$  is the median inferred  $\tilde{N}_e(t)$  across all superlineage combinations where  $\tilde{N}_e(t) < 10^5$  (values above  $10^5$  likely indicate non-convergence of the optimization). The reported errors on  $\tilde{N}_e(t)$  are the 95% confidence intervals (again taking the median across all superlineage combinations where  $\tilde{N}_e(t) < 10^5$ ) which are calculated by using the likelihood ratio to get a p-value [63, 64]. We replace the likelihood with the profile likelihood, which has the nuisance parameters  $c_{0:T}$  profiled out:

$$p > 0.05 \quad (30)$$

$$p = \int I \left[ \frac{\mathcal{L}_{\tilde{N}_e}(\hat{c}_{0:T} | \vec{\phi}_{0:T}^{obs})}{\mathcal{L}_{\tilde{N}'_e}(\hat{c}_{0:T} | \vec{\phi}_{0:T}^{obs})} > 1 \right] P_{\tilde{N}'_e}(\hat{c}_{0:T} | \vec{\phi}_{0:T}^{obs}) d\tilde{N}'_e \quad (31)$$

$$\hat{c}_{0:T} = \arg \max_{c_{0:T}} \mathcal{L}_{\tilde{N}_e}(c_{0:T} | \vec{\phi}_{0:T}^{obs}) \quad (32)$$

$$P_{\tilde{N}'_e}(\hat{c}_{0:T} | \vec{\phi}_{0:T}^{obs}) \propto \mathcal{L}_{\tilde{N}'_e}(\hat{c}_{0:T} | \vec{\phi}_{0:T}^{obs}) p(\tilde{N}'_e) \quad (33)$$

$$p(\tilde{N}_e) \propto 1 \quad (34)$$

503 where  $I$  is an indicator function that equals one when the argument is true and zero otherwise,  $\mathcal{L}_{\tilde{N}_e}(\hat{c}_{0:T} | \vec{\phi}_{0:T}^{obs})$   
504 is the profile likelihood with the nuisance parameters (in this case)  $c_{0:T}$  profiled out,  $P_{\tilde{N}'_e}(\hat{c}_{0:T} | \vec{\phi}_{0:T}^{obs})$  is the  
505 posterior where we have used a uniform prior. We also tried a Jeffreys prior which is used for variance  
506 parameters, but it gave similar results on simulated data because it looked relatively flat over the values of  
507  $\tilde{N}_e(t)$  of interest. As the Jeffreys prior was more computationally expensive than the uniform prior and the  
508 two priors gave similar results, we used the uniform prior for the analyses.

509 The reported values of  $c_t$  are the median across all superlineage combinations and across all time series  
510 segments where the timepoint appears. The reported errors on  $c_t$  are the 95% confidence intervals as  
511 calculated by the middle 95% of values across superlineage combinations and time series segments.

512 An example of inferred parameters on simulated data using the maximum likelihood estimation approach,  
513 compared to the initial guesses of the parameters from the methods of moments approach, is shown in  
514 Figure S19.

515 **Correcting for the number of sequences assigned to lineages**

516 Because some sequences occur before the cut point in the tree that is used for creating lineages, they are  
517 not included in any lineages. As a result, the number of sequences assigned to lineages is lower than the  
518 number of sequences in the tree. To correct for the bias in inferred effective population size that results  
519 from leaving out sequences from parts of the tree, we divide the inferred effective population size by the  
520 fraction of sequences in the tree that are assigned to a lineage. We note that while the number of sequences  
521 in the tree is less than the total number of sampled sequences, the sequences in the tree were chosen to be a  
522 representative fraction of the total sampled sequences. Thus, we do not need to additionally correct for the  
523 downsampling of sequences that were included in the tree. To test that randomly subsampling sequences for  
524 the analysis does not affect the results, we randomly subsampled half of the Delta sequences, and reran the  
525 analyses; the inferred effective population size was very similar to that from the full number of sequences  
526 (Figure S20).

## 527 Simulations for validating method

For the model validation, we perform simulations of the lineage trajectories using a discrete Wright-Fisher model. 500 lineages are seeded initially, and the initial frequency of lineages is taken to be the same across all lineages. In each subsequent Epiweek, the true number of counts for a lineage is drawn from a multinomial distribution where the probabilities of different outcomes are the true frequencies of the lineages in the previous Epiweek and the number of experiments is the effective population size. The true frequency is calculated by dividing the true number of counts by  $N$ . The observed counts are drawn from a negative binomial distribution,

$$p(n_t^{obs}|f_t) = NB(r, q) \equiv \binom{n_t^{obs} + r - 1}{r - 1} q^r (1 - q)^{n_t^{obs}} \quad (35)$$

$$r = \frac{f_t M_t}{c_t - 1} \quad (36)$$

$$q = \frac{1}{c_t} \quad (37)$$

528 which has the same mean and variance as the emission probability in Equation 6. The total number of  
529 observed sequences in each timepoint is calculated empirically after the simulation is completed, as it may  
530 not be exactly  $M_t$ . The simulation is run for 10 weeks of “burn-in” time before recording to allow for  
531 equilibration. Superlineages are created in the same way as described above.

532 For long time series simulations, some lineages will go extinct due to genetic drift, making it challenging  
533 to have sufficient data for the analysis. To be able to have a high enough number of lineages for the entire  
534 time series, we introduce mutations to a new lineage with a small rate  $\mu = 0.01$  per generation per individual.

## 535 Calculating the effective population size for an SIR or SEIR model

536 The effective population size times the generation time in an SIR model is given by Refs. [41, 35]

$$\tilde{N}_e^{\text{SIR}}(t) \equiv N_e^{\text{SIR}}(t)\tau(t) = \frac{I(t)}{2R_t\gamma_I}. \quad (38)$$

537 The variance in offspring number for an SIR model is approximately 2.

538 For an SEIR model, we calculated  $\tilde{N}_e(t)$  following the framework from Ref. [36]. Using this framework,  
539 we were only able to consider a situation where the epidemic is in equilibrium. We test how well this  
540 approximates the situation out of equilibrium using simulations (see Supplementary Information).

541 We first considered how the mean number of lineages,  $A$ , changes going backwards in time,  $s$ , which is  
542 given by

$$\frac{dA}{ds} = -fp_c \quad (39)$$

543 where  $f$  is the number of transmissions per unit time and  $p_c$  is the probability that a transmission results  
544 in a coalescence being observed in our sample.  $p_c$  is given by the number of ways of choosing two lineages  
545 divided by the number of ways of choosing two infectious individuals

$$p_c = \frac{\binom{A(s)}{2}}{\binom{N(s)}{2}} \underset{\lim_{N(s) \rightarrow \infty}}{=} \binom{A(s)}{2} \frac{2}{N(s)^2}. \quad (40)$$

546 where the limit assumes that the number of infectious individuals,  $N(s)$ , is large. In the Kingman coalescent  
547 we also have

$$\frac{dA}{ds} = -\binom{A(s)}{2} \frac{1}{\tilde{N}_e(t)}. \quad (41)$$

548 Combining Equations 39, 40, and 41, we have

$$\tilde{N}_e(t) = \frac{N(s)^2}{2f}. \quad (42)$$

549 Thus by determining the number of transmissions per unit time,  $f$ , and the number of infectious individuals,  
 550  $N(s)$ , in an SEIR model, we can find an expression for  $\tilde{N}_e(t)$ .

These quantities can be derived from the equations describing the number of susceptible ( $S$ ), exposed ( $E$ ), infectious ( $I$ ), and recovered ( $R$ ) individuals in an SEIR model

$$\frac{dS}{dt} = -\beta I \frac{S}{N_H} \quad (43)$$

$$\frac{dE}{dt} = \frac{\beta IS}{N_H} - \gamma_E E - \delta_E E \quad (44)$$

$$\frac{dI}{dt} = \gamma_E E - \gamma_I I - \delta_I I \quad (45)$$

$$\frac{dR}{dt} = \gamma_I I \quad (46)$$

551 where  $\beta$  is the number of transmissions per infectious individual per unit time (the number of contacts  
 552 made by an infectious individual per unit time multiplied by the probability that a contact results in a  
 553 transmission),  $N_H$  is the total population size ( $N_H = S + E + I + R$ ),  $\gamma_E$  is the rate that an exposed  
 554 individual becomes infectious,  $\delta_E$  is the rate of death for an exposed individual,  $\gamma_I$  is the rate than an  
 555 infectious individual recovers, and  $\delta_I$  is the rate of death for an infectious individual.

556 The number of infectious individuals in a generation,  $N(s)$ , is given by the instantaneous number of infec-  
 557 tious individuals plus the number of exposed individuals that will become infectious in that generation [41].  
 558 Thus,

$$N(s) = \frac{\gamma_E}{\gamma_E + \delta_E} E + I. \quad (47)$$

559 The number of transmissions per unit time is given by

$$f = \beta I \frac{S}{N_H}. \quad (48)$$

560 We rewrite  $f$  in terms of the effective reproduction number (for which data are available) which is given by  
 561 the number of transmissions per unit time ( $f$ ) divided by the number of recoveries and deaths per unit time

$$R_t = \frac{f}{(\gamma_I + \delta_I)I + \delta_E E}. \quad (49)$$

562 Putting everything together, we have that  $\tilde{N}_e(t)$  for an SEIR model in equilibrium is given by

$$\tilde{N}_e^{\text{SEIR,eq}}(t) = \frac{\left[ \left( \frac{\gamma_E}{\gamma_E + \gamma_I} \right) E + I \right]^2}{2R_t[(\gamma_I + \delta_I)I + \delta_E E]}. \quad (50)$$

563 For SARS-CoV-2, the death rates are much lower than the rate at which exposed individuals become in-  
 564 fectious and the rate at which infectious individuals recover ( $\delta_E, \delta_I \ll \gamma_E, \gamma_I$ ). In this limit, Equation 50  
 565 simplifies to

$$\tilde{N}_e^{\text{SEIR,eq}}(t) = \frac{(E + I)^2}{2R_t \gamma_I I}. \quad (51)$$

566 To calculate the  $\tilde{N}_e$  for an SIR or SEIR model, we use the estimated number of positives from the  
 567 COVID-19 Infection Survey for  $I(t)$ . This number is an estimate of the number of positive individuals in  
 568 the community as measured by surveillance and includes both symptomatic and asymptomatic individuals.  
 569 While the estimated number of positives does not include cases from hospitals, care homes, and other com-  
 570 munal establishments, community cases likely contribute the most to transmission. We used the measured  
 571 effective reproduction number from the UK Health Security Agency for  $R_t$ .

572 To calculate the number of exposed individuals for the SEIR model, we solved for  $E$  in Equation 45  
 573 (taking  $\delta_E \ll \gamma_E$ )

$$E = \frac{1}{\gamma_E} \left( \frac{dI}{dt} + \gamma_I I \right). \quad (52)$$

574  $\frac{dI}{dt}$  was calculated numerically as  $\frac{I(t+\Delta t) - I(t-\Delta t)}{2\Delta t}$  where  $\Delta t = 1$  week. The parameter values used were  $\gamma_E^{-1} = 3$  days and  $\gamma_I^{-1} = 5.5$  days [39, 40]. We checked that varying the value used for  $\gamma_I$  does not substantially affect the results (Figure S12). The error on  $E$  was calculated by taking the minimum and maximum possible values from the combined error intervals of  $I(t + \Delta t)$  and  $I(t - \Delta t)$  (note that this does not correspond to a specific confidence interval size).

579 The error on  $\tilde{N}_e(t)$  for the SIR or SEIR model was calculated similarly by taking the minimum and  
580 maximum possible values from the combined error intervals of  $E$ ,  $I$ , and  $R_t$ . Only time points where the  
581 error interval of  $\tilde{N}_e(t)$  was less than 3 times the point estimate were kept.

## 582 Calculating the effective population size for an SIR or SEIR model by variant

583 To calculate the effective population size for an SIR or SEIR model by variant, we needed to determine  
584 the variant-specific: number of infectious individuals  $I(t)$ , number of exposed individuals  $E(t)$ , effective  
585 reproduction number  $R_t$ , and rate than an infectious individual recovers  $\gamma_I$ . We assumed that  $\gamma_I$  is constant  
586 between variants. We calculated the number of infectious individuals  $I(t)$  by multiplying the total number  
587 of positives by the fraction of each variant in the reported sequences. This should be a good representation  
588 of the fraction of the variant in the population as the sequences are a random sample of cases detected  
589 via surveillance. We calculated the number of variant-specific exposed individuals  $E(t)$  in the same way as  
590 described above using the variant-specific number of infectious individuals. We assumed that the rate an  
591 exposed individual becomes infectious  $\gamma_E$  is constant between variants.

592 We calculated the variant-specific effective reproduction number by rescaling the measured effective  
593 reproduction number for the whole population

$$R_t^v = R_t \frac{R_0^v}{\sum_w R_0^w f^w} \quad (53)$$

594 where  $R_0^w$  is the basic reproduction number of the variant  $w$  and  $f^w$  is the fraction of the infectious population  
595 with variant  $w$ . The values of  $R_0$  when rescaled to  $R_0^{\text{pre-B.1.1.7}}$  that are used for the data presented in the  
596 main text are  $\frac{R_0^{\text{pre-B.1.1.7}}}{R_0^{\text{pre-B.1.1.7}}} = 1$ ,  $\frac{R_0^{\text{Alpha}}}{R_0^{\text{pre-B.1.1.7}}} = 1.7$  (Ref. [17]),  $\frac{R_0^{\text{Delta}}}{R_0^{\text{pre-B.1.1.7}}} = 1.97$  (Ref. [65]). Varying the variant  
597  $R_0$  within the ranges reported in the literature does not substantially affect the results (Figure S21).

## 598 Inference of fitness from lineage frequency time series

599 We sought to infer the fitness effects of individual lineages, so that we could then determine if putatively  
600 selected lineages are influencing the estimation of the time-varying effective population sizes. We first used  
601 a deterministic method to estimate lineage fitness effects, similar to the method described in [66].

602 On average, when the frequency of lineage  $i$  is sufficiently small  $f_{t,i} \ll 1$ , the frequency dynamics will  
603 exponentially grow/decay according to the lineage fitness effect,  $s_i$ ,

$$\langle f_{t,i} \rangle = f_{0,i} e^{s_i t}$$

604 The two sources of noise—genetic drift and measurement noise—both arise from counting processes, so the  
605 combined noise will follow  $\text{var}(f_{t,i}) \propto \langle f_{t,i} \rangle$ . To account for the inherent discreteness of the number of cases  
606 in a lineage—especially important to accurately model lineages at low frequencies—we modeled the observed  
607 counts at Epiweek  $t$  of lineage  $i$ ,  $r_{t,i}$ , as a negative binomial random variable,

$$r_{t,i} | s_i, f_{0,i} \sim \text{NB}(\mu_{t,i}, \zeta_t) \quad (54)$$

$$\langle r_{t,i} \rangle = \mu_{t,i} \quad (55)$$

$$\text{var}(r_{t,i}) = \zeta_t \langle r_{t,i} \rangle \quad (56)$$

$$\mu_{t,i} = M_t f_{0,i} e^{s_i t} \quad (57)$$

608 Where  $M_t$  is the total number of sequences, and  $\zeta_t$  is a dispersion parameter. We took  $\zeta_t$  as the total  
609 marginal variance at a given time-point, i.e.  $\zeta_t = c_t + M_t/N_e(t)$ , where we computed estimates of  $c_t$  and  $N_e$

610 as previously described (section “Maximum likelihood estimation of the parameters”). The final likelihood  
 611 for the fitness,  $s_i$ , of lineage  $i$  is obtained by combining the data from all the relevant time-points,

$$P(\mathbf{r}_i|s_i, f_{0,i}) = \prod_t \frac{\Gamma\left(r_{t,i} + \frac{\mu_{t,i}}{\zeta_t - 1}\right)}{\Gamma\left(\frac{\mu_{t,i}}{\zeta_t - 1}\right) \Gamma(r_{t,i} + 1)} \frac{(\zeta_t - 1)^{r_{t,i}}}{\zeta_t^{r_{t,i} + \frac{\mu_{t,i}}{\zeta_t - 1}}} \quad (58)$$

612 The point estimate of the lineage fitness,  $\hat{s}_i$ , is then numerically computed as the maximum likelihood,

$$\hat{s}_i = \operatorname{argmax}_{s_i} \log P(\mathbf{r}_i|s_i, f_{0,i}) \quad (59)$$

613 For ease of computation and generality we compute the p-value as the posterior probability that the  
 614 likelihood ratio between null and alternative hypotheses is greater than 1, i.e. the probability that the  
 615 data more strongly support the null hypothesis (the lineage is neutral) over the alternative (the lineage is  
 616 non-neutral),

$$p_i = P_{s_i|\mathbf{r}} \left( \frac{\mathcal{L}(0|\mathbf{r}_i)}{\mathcal{L}(s_i|\mathbf{r}_i)} > 1 \right)$$

$$P(s_i|\mathbf{r}) \propto \mathcal{L}(s_i|\mathbf{r}_i)$$

617 Where  $\mathcal{L}(s_i|\mathbf{r}_i)$  is the profile likelihood. This convenient definition has been shown to be equivalent to  
 618 the frequentist definition of the p-value using a likelihood ratio test statistic (if the distribution is invariant  
 619 under transformation) [63, 64], and does not require asymptotic approximations. After obtaining p-values  
 620 for all the lineages, we performed a standard Benjamini-Hochberg FDR on all the p-values. As mentioned in  
 621 the main text, the  $N_e$  inference method was then re-run using all lineages, except those that were significant  
 622 at  $\text{FDR} < 0.05$  using the above method.

623 Subsequently, we sought to use a more accurate method to detect if lineages have constant, non-neutral  
 624 fitness effects. So, we turned back to our HMM method and modified it slightly. We wanted to measure the  
 625 fitness effects of lineages that may be very small, with counts at some time points close to or at 0, which  
 626 then would not satisfy the central limit theorem, upon which our main HMM method rests. Thus, we used a  
 627 different transition [62] and emission probability for the HMM, which are more valid for small population  
 628 sizes and counts:

$$p(f_{t+1}|f_t, N_e(t), s_i) = \frac{1}{2} \sqrt{\frac{2(e^{s_i t} f_t)^{1/2}}{\pi f_{t+1}^{3/2} (\tilde{N}_e(t))^{-1}}} \exp\left(-\frac{2(\sqrt{f_{t+1}} - \sqrt{e^{s_i t} f_t})^2}{(\tilde{N}_e(t))^{-1}}\right) \quad (60)$$

$$f_t^{obs}|f_t, c_t \sim \text{NB}(f_t, c_t) \quad (61)$$

$$\langle f_t^{obs} \rangle = f_t \quad (62)$$

$$\text{var}(f_t^{obs}) = c_t f_t. \quad (63)$$

629 We used the point estimates of  $N_e$  and  $c_t$  that were calculated after we had excluded the selected lineages  
 630 detected by the conservative method. To get the maximum likelihood estimate of  $s$  for each lineage, we used  
 631 the standard HMM forward algorithm, numerically integrating over the intermediates. We then calculated  
 632 the p-value for each lineage by calculating the log-likelihood ratio  $\text{LLR}_i = 2(\mathcal{L}_i(\hat{s}_i) - \mathcal{L}_i(0))$ , and then  
 633 comparing it to a chi-squared ratio with d.f.=1. After obtaining p-values for all the lineages, we performed  
 634 a standard Benjamini-Hochberg FDR on all the p-values.

## 635 Stochastic simulations of SEIR model

636 The stochastic simulations of an SEIR model were performed using a Gillespie simulation with 4 states:  
 637 susceptible, exposed, infectious, and recovered, where the number of individuals in each state are denoted  
 638 by  $S(t)$ ,  $E(t)$ ,  $I(t)$ , and  $R(t)$  respectively. There are 3 types of events that lead to the following changes in  
 639 the number of individuals in each state

1. Infection of an susceptible individual with probability  $\frac{\beta I(t)S(t)}{N(t)}$

$$S(t) = S(t) - 1 \quad (64)$$

$$E(t) = E(t) + 1 \quad (65)$$

2. Transition of an exposed individual to being infectious with probability  $\gamma_E E(t)$

$$E(t) = E(t) - 1 \quad (66)$$

$$I(t) = I(t) + 1 \quad (67)$$

3. Recovery of an infectious individual with probability  $\gamma_I I(t)$

$$I(t) = I(t) - 1 \quad (68)$$

$$R(t) = R(t) + 1 \quad (69)$$

640 where  $\beta \equiv R_0\gamma_I$ ,  $R_0$  is the basic reproduction number,  $\gamma_E$  is the rate that exposed individuals become  
 641 infectious, and  $\gamma_I$  is the rate that infectious individuals recover. As in the rest of this work, we assume that  
 642 the birth rate of susceptible individuals, background death rate, and the death rate due to disease are much  
 643 slower compared to the rates of the above processes and thus can be neglected from the dynamics.

644 The time until the next event is drawn from an exponential distribution with rate given by the inverse  
 645 of the sum of the above probabilities, and the type of event is randomly drawn weighted by the respective  
 646 probabilities.

647 Because the time of the events occurs in continuous time, but the inference method of the effective  
 648 population size works in discrete time, we must convert from continuous to discrete time. To perform this  
 649 conversion, we calculate the net number of events of each type in each chosen unit of discrete time (1 week)  
 650 and perform the changes in the number of individuals of each state as described above. Thus, for example, if  
 651 within the same week an individual becomes exposed and then becomes infectious, it will cause the number  
 652 of susceptible individuals to decrease by 1, no change in the number of exposed individuals, and the number  
 653 of infectious individuals to increase by 1.

654 The infected (or infected and exposed) individuals are randomly assigned a lineage at a given time after  
 655 the start of the epidemic. For our simulations, we chose the lineage labeling time as 75 days or 10.7 weeks  
 656 since the approximate number of infectious individuals was high enough at that time to generate sufficient  
 657 diversity in lineages, and we chose the number of different types of lineages as 100. The other parameters that  
 658 we used for the simulations were  $R_0 = 2$ ,  $\gamma_E^{-1} = 3$  days,  $\gamma_I^{-1} = 5.5$  days,  $N(t) = S(t) + E(t) + I(t) + R(t) = 10^6$ .  
 659 The initial condition of the simulation is  $S(t) = N(t) - 1$ ,  $E(t) = 1$ , and  $I(t) = R(t) = 0$ .

660 To test the sensitivity of the results to whether the reported PCR positive individuals are infectious or  
 661 whether they can also be from the exposed class, we recorded the results in two ways. In the first case, only  
 662 the infectious individuals we recorded as positive (Figure S22), and in the second case both the exposed and  
 663 infectious individuals were recorded as positive (Figure S23). Inference of  $\tilde{N}_e(t)$  was subsequently done on  
 664 the lineage frequency trajectories of the recorded positive individuals. The SIR or SEIR model  $\tilde{N}_e(t)$  were  
 665 calculated analytically using the true numbers of infectious and exposed individuals and numerically using  
 666 the number of positive individuals as described above in “Calculating the effective population size for an  
 667 SIR or SEIR model”.

## 668 Deme simulations

669 To better understand the effect of host population structure on the effective population size, we simulated  
 670 a simple situation where there are “demes”, or groups, of individuals with very high rates of transmission  
 671 between individuals in that deme, but the rate of transmission between individuals from different demes  
 672 is very low. In a given simulation, all demes have the same number of individuals (10, 50, 100, or 200).  
 673 The total number of demes is chosen to be very high ( $5.6 \times 10^6$ ). Initially, a certain number of demes  
 674 (100, 1000, 2000, or 5000) are each seeded by a single infectious individual infected by a randomly chosen  
 675 lineage (200 different lineages). We simulated deterministic SEIR dynamics within demes with  $R_0 = 10$ ,  
 676  $\gamma_E = (2.5 \text{ days})^{-1}$ ,  $\gamma_I = (6.5 \text{ days})^{-1}$ . We simulated Poisson transmission dynamics between demes. In

677 order to calibrate the overall population dynamics to be roughly in equilibrium (the number of infectious  
678 individuals is not deterministically growing or shrinking), we draw the number of between-deme infections  
679 caused by a given deme from a Poisson distribution with mean 1. The time of the between-deme infection  
680 event is randomly chosen, weighted by the number of infected individuals within a deme at a given time. The  
681 number of infectious individuals in each lineage is recorded every 1 week, and the frequency of the lineage  
682 is calculated by dividing by the total number of infectious individuals from all lineages in that week. The  
683 lineage frequency data from a period of 9 weeks starting in week 42 is used for the inference of effective  
684 population size. The effective population size inference is performed as above except in the absence of  
685 measurement noise, so there is no emission step in the HMM.

## 686 Data and code availability

687 Data and code to reproduce the analyses in this manuscript are available at [https://github.com/qinqin-yu/sars-cov-2\\_genetic\\_drift](https://github.com/qinqin-yu/sars-cov-2_genetic_drift).

## 689 Acknowledgements

690 We are grateful to the Hallatschek lab for helpful discussions, feedback, and comments on earlier versions  
691 of this manuscript, particularly Giulio Issachini and Valentin Slepukhin. We are grateful to Aditya Prasad  
692 for advice on computing. We thank Mike Boots, Vince Buffalo, Katia Koelle, Priya Moorjani, Rasmus  
693 Nielsen, Daniel Reeves, and Daniel Weissman for helpful discussions and feedback. We thank Hernan G.  
694 Garcia and Yun S. Song for helpful comments on earlier versions of this manuscript. This material is based  
695 upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No.  
696 DGE 1106400 (to QY), and JSPS KAKENHI (#22K03453, #JP22K06347, to TO). JAA acknowledges  
697 support from an NSF Graduate Research Fellowship and a Berkeley Fellowship. OH acknowledges support  
698 by a Humboldt Professorship of the Alexander von Humboldt Foundation. This research used resources  
699 of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy  
700 Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract  
701 No. DE-AC02-05CH11231 using NERSC BER-ERCAP0019907. COG-UK is supported by funding from  
702 the Medical Research Council (MRC) part of UK Research & Innovation (UKRI), the National Institute  
703 of Health Research (NIHR) [grant code: MC\_PC\_19027], and Genome Research Limited, operating as the  
704 Wellcome Sanger Institute. The authors acknowledge use of data generated through the COVID-19 Genomics  
705 Programme funded by the Department of Health and Social Care. The views expressed are those of the author  
706 and not necessarily those of the Department of Health and Social Care or UKHSA. We thank the COG-UK  
707 consortium and all partners and contributors who are listed at <https://www.cogconsortium.uk/about/>.

## 708 References

- 709 1. Volz EM and Didelot X. Modeling the growth and decline of pathogen effective population size provides  
710 insight into epidemic dynamics and drivers of antimicrobial resistance. *Systematic Biology* 2018; 67:719–  
711 28
- 712 2. Lloyd-Smith JO, Schreiber SJ, Kopp PE, and Getz WM. Superspreading and the effect of individual  
713 variation on disease emergence. *Nature* 2005; 438:355–9
- 714 3. Alexander H and Day T. Risk factors for the evolutionary emergence of pathogens. *Journal of The  
715 Royal Society Interface* 2010; 7:1455–74
- 716 4. Kucharski A and Althaus CL. The role of superspreading in Middle East respiratory syndrome coron-  
717 avirus (MERS-CoV) transmission. *Eurosurveillance* 2015; 20:21167
- 718 5. Melsew YA, Gambhir M, Cheng AC, McBryde ES, Denholm JT, Tay EL, and Trauer JM. The role  
719 of super-spreading events in *Mycobacterium tuberculosis* transmission: evidence from contact tracing.  
720 *BMC Infectious Diseases* 2019; 19:1–9

721 6. Pastor-Satorras R, Castellano C, Van Mieghem P, and Vespignani A. Epidemic processes in complex  
722 networks. *Reviews of Modern Physics* 2015; 87:925

723 7. Otto SP, Day T, Arino J, Colijn C, Dushoff J, Li M, Mechai S, Van Domselaar G, Wu J, Earn DJ,  
724 et al. The origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19  
725 pandemic. *Current Biology* 2021; 31:R918–R929

726 8. Goyal A, Reeves DB, and Schiffer JT. Early super-spreader events are a likely determinant of novel  
727 SARS-CoV-2 variant predominance. *medRxiv* 2021

728 9. Day T, Gandon S, Lion S, and Otto SP. On the evolutionary epidemiology of SARS-CoV-2. *Current  
729 Biology* 2020; 30:R849–R857

730 10. Tasakis RN, Samaras G, Jamison A, Lee M, Paulus A, Whitehouse G, Verkoczy L, Papavasiliou FN,  
731 and Diaz M. SARS-CoV-2 variant evolution in the United States: High accumulation of viral mutations  
732 over time likely through serial Founder Events and mutational bursts. *PloS ONE* 2021; 16:e0255169

733 11. Ghafari M, Liu Q, Dhillon A, Katzourakis A, and Weissman DB. Investigating the evolutionary origins  
734 of the first three SARS-CoV-2 variants of concern. *Frontiers in Virology* 2022 :76

735 12. Sneppen K, Nielsen BF, Taylor RJ, and Simonsen L. Overdispersion in COVID-19 increases the ef-  
736 fectiveness of limiting nonrepetitive contacts for transmission control. *Proceedings of the National  
737 Academy of Sciences* 2021; 118

738 13. Charlesworth B. Effective population size and patterns of molecular evolution and variation. *Nature  
739 Reviews Genetics* 2009; 10:195–205

740 14. Lakdawala SS and Menachery VD. Catch me if you can: superspreading of COVID-19. *Trends in  
741 Microbiology* 2021; 29:919–29

742 15. Althouse BM, Wenger EA, Miller JC, Scarpino SV, Allard A, Hébert-Dufresne L, and Hu H. Super-  
743 spreading events in the transmission dynamics of SARS-CoV-2: Opportunities for interventions and  
744 control. *PLoS Biology* 2020; 18:e3000897

745 16. Goyal A, Reeves DB, Cardozo-Ojeda EF, Schiffer JT, and Mayer BT. Viral load and contact hetero-  
746 geneity predict SARS-CoV-2 transmission and super-spreading events. *eLife* 2021; 10:e63537

747 17. Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, Hinsley WR, Laydon DJ, Dabrera G,  
748 O'Toole Á, et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* 2021;  
749 593:266–9

750 18. Ragonnet-Cronin M, Boyd O, Geidelberg L, Jorgensen D, Nascimento FF, Siveroni I, Johnson RA,  
751 Baguelin M, Cucunubá ZM, Jaunekaitė E, et al. Genetic evidence for the association between COVID-  
752 19 epidemic severity and timing of non-pharmaceutical interventions. *Nature Communications* 2021;  
753 12:1–7

754 19. Du Plessis L, McCrone JT, Zarebski AE, Hill V, Ruis C, Gutierrez B, Raghwani J, Ashworth J,  
755 Colquhoun R, Connor TR, et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in  
756 the UK. *Science* 2021; 371:708–12

757 20. Adam D, Gostic K, Tsang T, Wu P, Lim WW, Yeung A, Wong J, Lau E, Du Z, Chen D, et al.  
758 Time-varying transmission heterogeneity of SARS and COVID-19 in Hong Kong (preprint). 2022

759 21. Frost SD, Pybus OG, Gog JR, Viboud C, Bonhoeffer S, and Bedford T. Eight challenges in phylody-  
760 namic inference. *Epidemics* 2015; 10:88–92

761 22. Bollback JP, York TL, and Nielsen R. Estimation of  $2N_e s$  from temporal allele frequency data. *Genetics*  
762 2008; 179:497–502

763 23. Buffalo V and Coop G. The linked selection signature of rapid adaptation in temporal genomic data.  
764 *Genetics* 2019; 213:1007–45

765 24. Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS, and Sherlock G. Quantitative evolutionary  
766 dynamics using high-resolution lineage tracking. *Nature* 2015; 519:181–6

767 25. Williamson EG and Slatkin M. Using maximum likelihood to estimate population size from temporal  
768 changes in allele frequencies. *Genetics* 1999; 152:755–61

769 26. Steinrücken M, Bhaskar A, and Song YS. A novel spectral method for inferring general diploid selection  
770 from time series genetic data. *The Annals of Applied Statistics* 2014; 8:2203

771 27. Stadler T. On incomplete sampling under birth–death models and connections to the sampling-based  
772 coalescent. *Journal of Theoretical Biology* 2009; 261:58–66

773 28. Dorp CHv, Goldberg EE, Hengartner N, Ke R, and Romero-Severson EO. Estimating the strength of  
774 selection for new SARS-CoV-2 variants. *Nature Communications* 2021; 12:1–13

775 29. O’Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, Colquhoun R, Ruis C, Abu-Dahab  
776 K, Taylor B, et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin  
777 tool. *Virus Evolution* 2021; 7:veab064

778 30. Rambaut A, Holmes EC, O’Toole Á, Hill V, McCrone JT, Ruis C, Plessis L du, and Pybus OG.  
779 A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature  
780 microbiology* 2020; 5:1403–7

781 31. UK Office for National Statistics. Coronavirus (COVID-19) Infection Survey: England. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/coronaviruscovid19infectionssurveydata>. Accessed: 2021-12-10

782 32. Challen R, Dyson L, Overton CE, Guzman-Rincon LM, Hill EM, Stage HB, Brooks-Pollock E, Pel-  
783 lis L, Scarabel F, Pascall DJ, et al. Early epidemiological signatures of novel SARS-CoV-2 variants:  
784 establishment of B.1.617.2 in England. *medRxiv* 2021

785 33. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, Pearson CA, Russell TW,  
786 Tully DC, Washburne AD, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7  
787 in England. *Science* 2021; 372:eaabg3055

788 34. The COVID-19 Genomics UK (COG-UK) consortium. An integrated national scale SARS-CoV-2 ge-  
789 nomic surveillance network. *The Lancet Microbe* 2020; 1:e99

790 35. Volz EM, Kosakovsky Pond SL, Ward MJ, Leigh Brown AJ, and Frost SD. Phylodynamics of infectious  
791 disease epidemics. *Genetics* 2009; 183:1421–30

792 36. Frost SD and Volz EM. Viral phylodynamics and the search for an ‘effective number of infections’.  
793 *Philosophical Transactions of the Royal Society B: Biological Sciences* 2010; 365:1879–90

794 37. Volz EM. Complex population dynamics and the coalescent under neutrality. *Genetics* 2012; 190:187–  
795 201

796 38. UK Health Security Agency. The R value and growth rate. <https://www.gov.uk/guidance/the-r-value-and-growth-rate#:~:text=The%20R%20range%20for%20the,as%20of%2019%20March%202021.&text=The%20R%20range%20for%20the,as%20of%2012%20March%202021..> Accessed: 2021-  
797 12-10

798 39. Miller D, Martin MA, Harel N, Tirosh O, Kustin T, Meir M, Sorek N, Gefen-Halevi S, Amit S, Vorontsov  
799 O, et al. Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel.  
800 *Nature Communications* 2020; 11:1–10

801 40. He X, Lau EH, Wu P, Deng X, Wang J, Hao X, Lau YC, Wong JY, Guan Y, Tan X, et al. Temporal  
802 dynamics in viral shedding and transmissibility of COVID-19. *Nature medicine* 2020; 26:672–5

803 41. Koelle K and Rasmussen DA. Rates of coalescence for common epidemiological models at equilibrium.  
804 *Journal of the Royal Society Interface* 2012; 9:997–1007

805 42. Endo A et al. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside  
806 China. *Wellcome Open Research* 2020; 5

807 43. Quilty BJ, Chapman LA, Wong KL, Gimma A, Pickering S, JD S, Neil RPG, Jarvis CI, and Kucharski  
808 AJ. Reconstructing the secondary case distribution of SARS-CoV-2 from heterogeneity in viral load  
809 trajectories and social contacts. *Report for SPI-M-O and SAGE* 2021

810 44. Obadia T, Silhol R, Opatowski L, Temime L, Legrand J, Thiébaut AC, Herrmann JL, Fleury E,  
811 Guillemot D, Boelle PY, et al. Detailed contact data and the dissemination of *Staphylococcus aureus*  
812 in hospitals. *PLoS Computational Biology* 2015; 11:e1004170

817 45. Steiner MC and Novembre J. Population genetic models for the spatial spread of adaptive variants: A  
818 review in light of SARS-CoV-2 evolution. *PLoS Genetics* 2022; 18

819 46. Taube JC, Miller PB, and Drake JM. An open-access database of infectious disease transmission trees  
820 to explore superspreader epidemiology. *PLoS Biology* 2022; 20:e3001685

821 47. Hodcroft EB, Zuber M, Nadeau S, Vaughan TG, Crawford KH, Althaus CL, Reichmuth ML, Bowen  
822 JE, Walls AC, Corti D, et al. Spread of a SARS-CoV-2 variant through Europe in the summer of 2020.  
823 *Nature* 2021; 595:707–12

824 48. Kraemer MU, Hill V, Ruis C, Dellicour S, Bajaj S, McCrone JT, Baele G, Parag KV, Battle AL,  
825 Gutierrez B, et al. Spatiotemporal invasion dynamics of SARS-CoV-2 lineage B.1.1.7 emergence. *Science*  
826 2021; 373:889–95

827 49. McCrone JT, Hill V, Bajaj S, Pena RE, Lambert BC, Inward R, Bhatt S, Volz E, Ruis C, Dellicour S,  
828 et al. Context-specific emergence and growth of the SARS-CoV-2 Delta variant. *Nature* 2022; 610:154–  
829 60

830 50. Nande A, Adlam B, Sheen J, Levy MZ, and Hill AL. Dynamics of COVID-19 under social distancing  
831 measures are driven by transmission network structure. *PLoS computational biology* 2021; 17:e1008684

832 51. Edsberg Møllgaard P, Lehmann S, and Alessandretti L. Understanding components of mobility during  
833 the COVID-19 pandemic. *Philosophical Transactions of the Royal Society A* 2022; 380:20210118

834 52. Wakeley J. Coalescent theory. Roberts & Company 2009

835 53. Feder AF, Kryazhimskiy S, and Plotkin JB. Identifying signatures of selection in genetic time series.  
836 *Genetics* 2014; 196:509–22

837 54. Mathieson I and McVean G. Estimating selection coefficients in spatially structured populations from  
838 time series data of allele frequencies. *Genetics* 2013; 193:973–84

839 55. Lythgoe KA, Hall M, Ferretti L, Cesare M de, MacIntyre-Cockett G, Trebes A, Andersson M, Otecko  
840 N, Wise EL, Moore N, et al. SARS-CoV-2 within-host diversity and transmission. *Science* 2021;  
841 372:eabg0821

842 56. Martin MA and Koelle K. Comment on “Genomic epidemiology of superspreading events in Austria re-  
843 veals mutational dynamics and transmission properties of SARS-CoV-2”. *Science translational medicine*  
844 2021; 13:eabh1803

845 57. Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, and Suchard MA. Improving Bayesian population  
846 dynamics inference: a coalescent-based model for multiple loci. *Molecular biology and evolution* 2013;  
847 30:713–24

848 58. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz  
849 MHY, et al. A draft sequence of the Neandertal genome. *Science* 2010; 328:710–22

850 59. Dehasque M, Ávila-Arcos MC, Diez-del-Molino D, Fumagalli M, Guschanski K, Lorenzen ED, Malaspinas  
851 AS, Marques-Bonet T, Martin MD, Murray GG, et al. Inference of natural selection from ancient DNA.  
852 *Evolution Letters* 2020; 4:94–108

853 60. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U,  
854 Johnson PL, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*  
855 2010; 468:1053–60

856 61. COG-UK and Microreact. UK SARS-CoV-2. <https://microreact.org/project/mxgBucHEZCZgsSWNrnQQRo-%20uk-sars-cov-2-2020-02-052021-04-20>. Accessed: 2021-04-20

857 62. Ba ANN, Cvijović I, Echenique JIR, Lawrence KR, Rego-Costa A, Liu X, Levy SF, and Desai MM.  
858 High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature* 2019;  
859 575:494–9

860 63. Aitkin M, Boys RJ, and Chadwick T. Bayesian point null hypothesis testing via the posterior likelihood  
861 ratio. *Statistics and Computing* 2005; 15:217–30

862 64. Smith I and Ferrari A. Equivalence between the posterior distribution of the likelihood ratio and a  
863 p-value in an invariant frame. *Bayesian Analysis* 2014; 9:939–62

865 65. Campbell F, Archer B, Laurenson-Schafer H, Jinnai Y, Konings F, Batra N, Pavlin B, Vandemaele  
866 K, Van Kerkhove MD, Jombart T, et al. Increased transmissibility and global spread of SARS-CoV-2  
867 variants of concern as at June 2021. *Eurosurveillance* 2021; 26:2100509

868 66. Ascensao J, Wetmore K, Good B, Arkin A, and Hallatschek O. Quantifying the Local Adaptive Land-  
869 scape of a Nascent Bacterial Community. 2022

870 67. Pybus OG, Rambaut A, and Harvey PH. An integrated framework for the inference of viral population  
871 history from reconstructed genealogies. *Genetics* 2000; 155:1429–37

872 68. Ho SY and Shapiro B. Skyline-plot methods for estimating demographic history from nucleotide se-  
873 quences. *Molecular Ecology Resources* 2011; 11:423–34

874 69. Zinger T, Gelbart M, Miller D, Pennings PS, and Stern A. Inferring population genetics parameters of  
875 evolving viruses using time-series data. *Virus Evolution* 2019; 5:vez011

876 70. Ferrer-Admetlla A, Leuenberger C, Jensen JD, and Wegmann D. An approximate Markov model for  
877 the Wright–Fisher diffusion and its application to time series data. *Genetics* 2016; 203:831–46

878 71. Lumby CK, Zhao L, Breuer J, and Illingworth CJ. A large effective population size for established  
879 within-host influenza virus infection. *eLife* 2020; 9:e56915

880 72. Sobel Leonard A, Weissman DB, Greenbaum B, Ghedin E, and Koelle K. Transmission bottleneck  
881 size estimation from pathogen deep-sequencing data, with an application to human influenza A virus.  
882 *Journal of Virology* 2017; 91:e00171–17

883 73. Lau MS, Grenfell B, Thomas M, Bryan M, Nelson K, and Lopman B. Characterizing superspreading  
884 events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA. *Proceedings of the  
885 National Academy of Sciences* 2020; 117:22430–5

886 74. Kirkegaard JB and Sneppen K. Variability of individual infectiousness derived from aggregate statistics  
887 of COVID-19. *medRxiv* 2021

888 75. Riou J and Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019 novel coro-  
889 navirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance* 2020; 25:2000058

890 76. Laxminarayan R, Wahl B, Dudala SR, Gopal K, Mohan B C, Neelima S, Jawahar Reddy K, Rad-  
891 hakrishnan J, and Lewnard JA. Epidemiology and transmission dynamics of COVID-19 in two Indian  
892 states. *Science* 2020; 370:691–7

893 77. Adam DC, Wu P, Wong JY, Lau EH, Tsang TK, Cauchemez S, Leung GM, and Cowling BJ. Clustering  
894 and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nature Medicine* 2020; 26:1714–  
895 9

896 78. Sun K, Wang W, Gao L, Wang Y, Luo K, Ren L, Zhan Z, Chen X, Zhao S, Huang Y, et al. Transmission  
897 heterogeneities, kinetics, and controllability of SARS-CoV-2. *Science* 2021; 371:eabe2424

898 79. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, Liu X, Wei L, Truelove SA, Zhang T, et al. Epidemiology  
899 and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a  
900 retrospective cohort study. *The Lancet Infectious Diseases* 2020; 20:911–9

## 901 Supplementary Information

### 902 Summary of existing methods for inferring the strength of genetic drift

903 There are currently four main types of methods for estimating the strength of genetic drift in pathogen  
904 transmission, which we summarize here for giving context to this study.

- 905 1. **Contact tracing** can directly measure superspreading by following the close contacts of infected  
906 individuals to measure the distribution of the number of secondary cases (the offspring number distribution) [2]. However, some secondary cases may be missed which can lead to measurement bias [20].  
907 Additionally, it is challenging to trace multiple generations of transmission, so we miss important  
908 information on host contact network structure.
- 909 2. Another type of method fits disease prevalence over time to **branching process models** [42]. These  
910 models assume a particular distribution for the offspring number distribution (often a negative binomial  
911 distribution) and estimate the combination of parameters of the offspring number distribution along  
912 with growth rate that best fit the observed disease prevalence. External information about the growth  
913 rate can be used to constrain the parameters of the offspring number distribution.
- 914 3. **Phylogenetics** methods arrange genomics sequences into a tree based on genomic distance and either  
915 measure the distribution of lineage sizes (number of sequences in different parts of the tree) [19] or fit  
916 the rate at which branches in the tree coalescence to determine the effective population size [27, 67,  
917 1, 68]. The effective population size is the population size that would reproduce the observed popu-  
918 lation dynamics under the idealized conditions of Wright-Fisher dynamics (discrete non-overlapping  
919 generations, a constant population size, and offspring determined by sampling with replacement from  
920 the previous generation). A lower effective population size indicates a higher level of genetic drift.
- 921 4. **Time series frequency methods** make use of a signature that genetic drift leaves in time series data,  
922 which is that it causes fluctuations in the lineage abundances. Higher amounts of genetic drift (lower  
923 effective population size) lead to larger fluctuations, and the magnitude of the fluctuations can be fit  
924 to determine the effective population size [69, 24] (Figure 1a). Time series methods have also been  
925 used extensively in population genetics [22, 70, 53, 23, 26, 25] and to estimate within-host effective  
926 population size [71] and between-host transmission bottleneck sizes [72].

### 928 Comparison to SEIR null model

929 In the main text, we compared the inferred  $\tilde{N}_e(t)$  to an SIR model. However, there are likely more complex  
930 epidemiological dynamics describing SARS-CoV-2. Here we check the results for an SEIR model which  
931 includes a susceptible, exposed, infectious, and recovered class. The SEIR model is a good representation of  
932 the epidemiology of SARS-CoV-2 when PCR test positivity is closely associated with an infected host being  
933 infectious; the literature suggests that this is a good assumption for SARS-CoV-2 [16], but we also test this  
934 assumption below. The exposed class thus represents individuals before they are infectious and test positive.  
935  $\tilde{N}_e(t)$  for an SEIR model in equilibrium (number of infectious individuals is constant over time) is given by  
936 (see Methods for derivation):

$$\tilde{N}_e^{\text{SEIR,eq}}(t) \equiv \{N_e(t)\tau(t)\}^{\text{SEIR,eq}} = \frac{(E(t) + I(t))^2}{2R_t\gamma_I(t)}. \quad (70)$$

937 where  $E(t)$  is the number of exposed individuals,  $I(t)$  is the number of infectious individuals,  $R_t$  is the  
938 effective reproduction number, and  $\gamma_I$  is the rate at which infectious individuals stop being infectious.  
939 While this equation is derived under equilibrium conditions, we show using simulations that this equation  
940 accurately estimates  $\tilde{N}_e(t)$  in non-equilibrium conditions after the peak of the pandemic (Figure S22); before  
941 the pandemic peak, this equation overestimates  $\tilde{N}_e(t)$  but by less than one order of magnitude. Additionally,  
942 we show that calculating the  $\tilde{N}_e(t)$  using the equation for an SIR model (Equation 1) when the dynamics are  
943 actually described by an SEIR model provides a lower bound on the actual  $\tilde{N}_e(t)$ . Thus, if the true dynamics  
944 of SARS-CoV-2 in England are actually SEIR dynamics, then the inference results shown in Figure 3c using

945 the SIR model should be an underestimate of the level of genetic drift; thus our main result that the literature  
946 values of superspreading do not sufficiently explain our results should still hold.

947 In reality, it may also be the case that some people test positive in a PCR test before they become  
948 infectious. To test the impact of this possibility on our results, in our simulations we recorded both exposed  
949 and infectious individuals as testing positive. We then calculated the SEIR model  $\tilde{N}_e(t)$  numerically as  
950 described in “Calculating the effective population size for an SIR or SEIR model” assuming that  $I(t)$  includes  
951 both infectious and exposed individuals (Figure S23). We find that the numerical solutions give slightly higher  
952  $\tilde{N}_e(t)$  as compared with the true analytical solutions; however, the numerical solutions to the SEIR and SIR  
953 models bound the inferred  $\tilde{N}_e(t)$ . Thus we also expect that our main result that the literature values of  
954 superspreading do not sufficiently explain our results should still hold in this scenario.

To calculate the SEIR model  $\tilde{N}_e(t)$  for the actual data, for the number of infectious individuals, we used the number of positive individuals estimated from the UK Office for National Statistics’ COVID-19 Infection Survey [31], which is a household surveillance study that reports positive PCR tests, regardless of symptom status. We used the measured effective reproduction number in England reported by the UK Health Security Agency [38]. We found that  $\tilde{N}_e^{\text{SEIR}}(t)$  is very similar to the number of positives because the effective reproduction number in England was very close to 1 across time. To calculate  $\tilde{N}_e^{\text{SEIR}}(t)$  for each variant or group of lineages, we rescaled the population-level  $I(t)$  and  $R_t$  based on the fraction of each variant in the population and the relative differences in reproduction numbers between variants (see Methods). We then calculated the scaled true population size,  $\tilde{N}(t) \equiv N(t)\tau(t)$ , for the SEIR model by multiplying by the variance in offspring number,  $\sigma^2$ , for the SEIR model [41]

$$\tilde{N}^{\text{SEIR}}(t) = \tilde{N}_e^{\text{SEIR}}(t)\{\sigma^2\}^{\text{SEIR}} \quad (71)$$

$$\{\sigma^2\}^{\text{SEIR}} = 2. \quad (72)$$

955 Overall, the inferred  $\tilde{N}_e(t)$  is lower than  $\tilde{N}^{\text{SEIR}}(t)$  by a time-dependent factor that varies between 70 and  
956 2000 (Figure S14), suggesting high levels of genetic drift in England across time, which is consistent with  
957 what we find with an SIR model (Figures 2 and S13). Also similarly to in the case with an SIR model, the  
958 ratio of  $\tilde{N}^{\text{SEIR}}(t)$  to the inferred  $\tilde{N}_e(t)$  for Alpha decreased over time, suggesting that the stochasticity in  
959 the transmission of Alpha decreased over time.

## 960 Application to COG-UK data by regions in England

961 The inference of effective population size can also reveal information about the well-mixed or spatially-  
962 structured nature of transmission dynamics within England. This can be done by inferring effective popu-  
963 lation size at smaller geographical scales within England. If the transmission dynamics were completely  
964 well-mixed, then we would expect  $\tilde{N}_e(t)$  to be the same across regions and compared to England. On the  
965 other hand, if the transmission dynamics were completely spatially segregated (i.e. transmission only occurs  
966 within the defined geographical areas, but not between them) and the dynamics were the same in each region,  
967 we would expect that the ratio  $\tilde{N}_e^{\text{SIR}}(t)/\tilde{N}_e^{\text{inf}}(t)$  to be the same across regions.

968 The geographical areas that we used were the 9 regions of England: East Midlands, East of England,  
969 London, North East, North West, South East, South West, West Midlands, and Yorkshire and The Humber.  
970 We looked at sequences from each region, repeating the analysis described above, and inferred the scaled  
971 effective population size (Figure S1). We observe a lower  $\tilde{N}_e(t)$  for in the region than in England for Delta  
972 in all regions, for Alpha in all regions except North East (where there was not enough data), and for B.1.177  
973 in all regions except North East. For lineages pre-B.1.177, the inferred  $\tilde{N}_e(t)$  is not significantly differnt in  
974 the region than in England. These results suggest that the dynamics are not well-mixed during the B.1.177,  
975 Alpha, and Delta waves.

976 The calculated SIR model  $\tilde{N}_e^{\text{SIR}}(t)$  (Figure S2) and the number of positive individuals in each region  
977 (Figure S3) were 1-2 orders of magnitude higher than the inferred  $\tilde{N}_e(t)$ , suggesting high levels of genetic  
978 drift. The ratios of the the SIR model  $\tilde{N}_e(t)$  and the number of positives to the inferred  $\tilde{N}_e(t)$  in the regions  
979 were similar to one another and to that seen in England as a whole, consistent with a scenario where the  
980 dynamics are spatially-structured and the extent of stochasticity in transmission is similar across regions.

981 Similarly to in England as a whole, the inferred measurement noise in each region was mostly indistin-  
982 guishable from uniform sampling except for in a few timepoints (Figure S4).

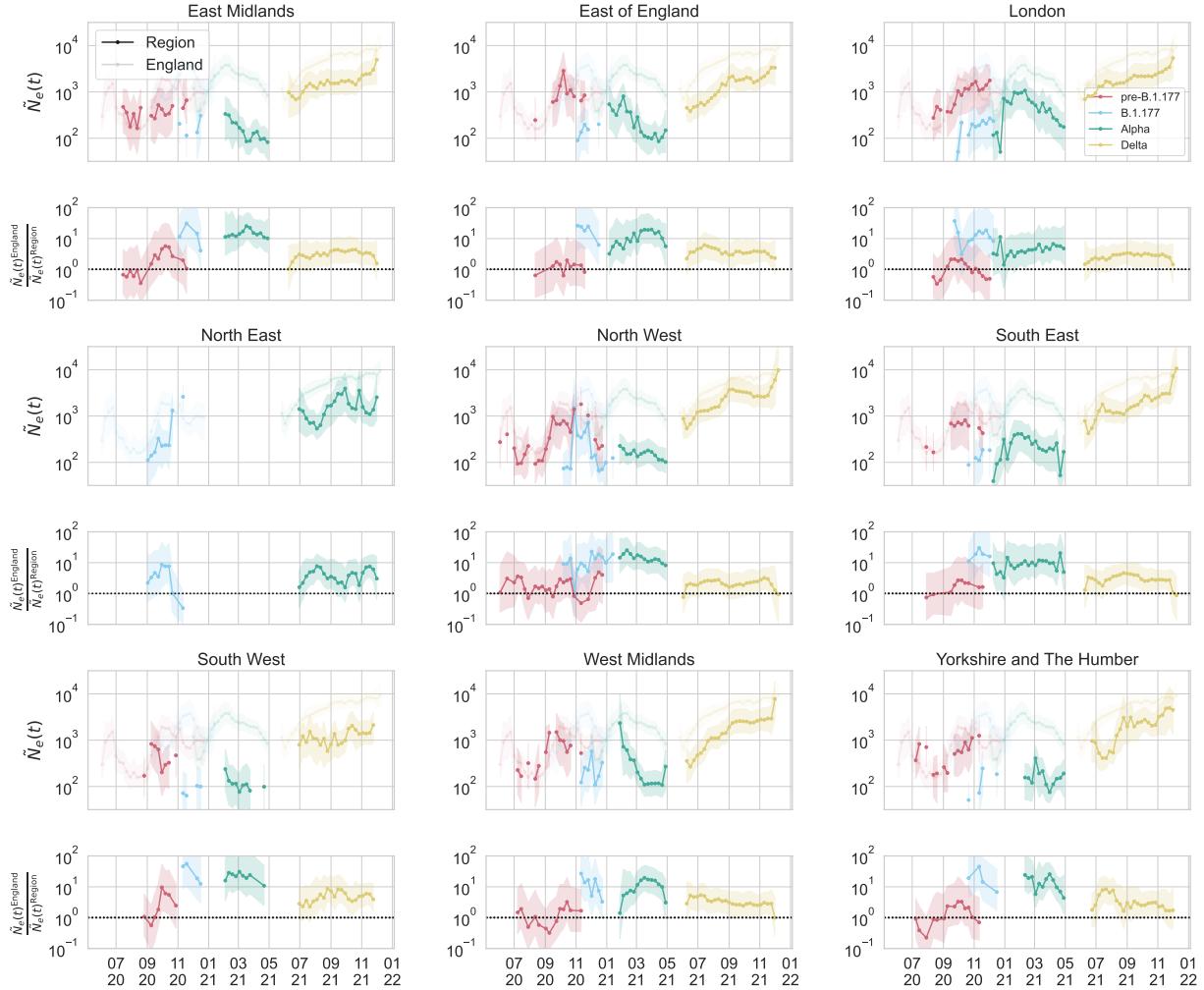


Figure S1: Inferred effective population size in regions of England. (Top panels) Inferred  $\tilde{N}_e(t)$  of pre-B.1.177 lineages, B.1.177, Alpha, and Delta for each region of England. The inferred  $\tilde{N}_e(t)$  for England as a whole is shown for reference. Shaded regions show 95% confidence intervals (see Methods). (Bottom panels) The ratio between the inferred  $\tilde{N}_e(t)$  of England and that of the region for each variant. A horizontal dashed line indicates a ratio of 1 (i.e.  $N_e(t)$  is the same in that region of England and England as a whole). Shared regions show the minimum and maximum possible values of the ratio from the combined error intervals of the numerator and denominator (thus, not corresponding to a specific confidence interval range).

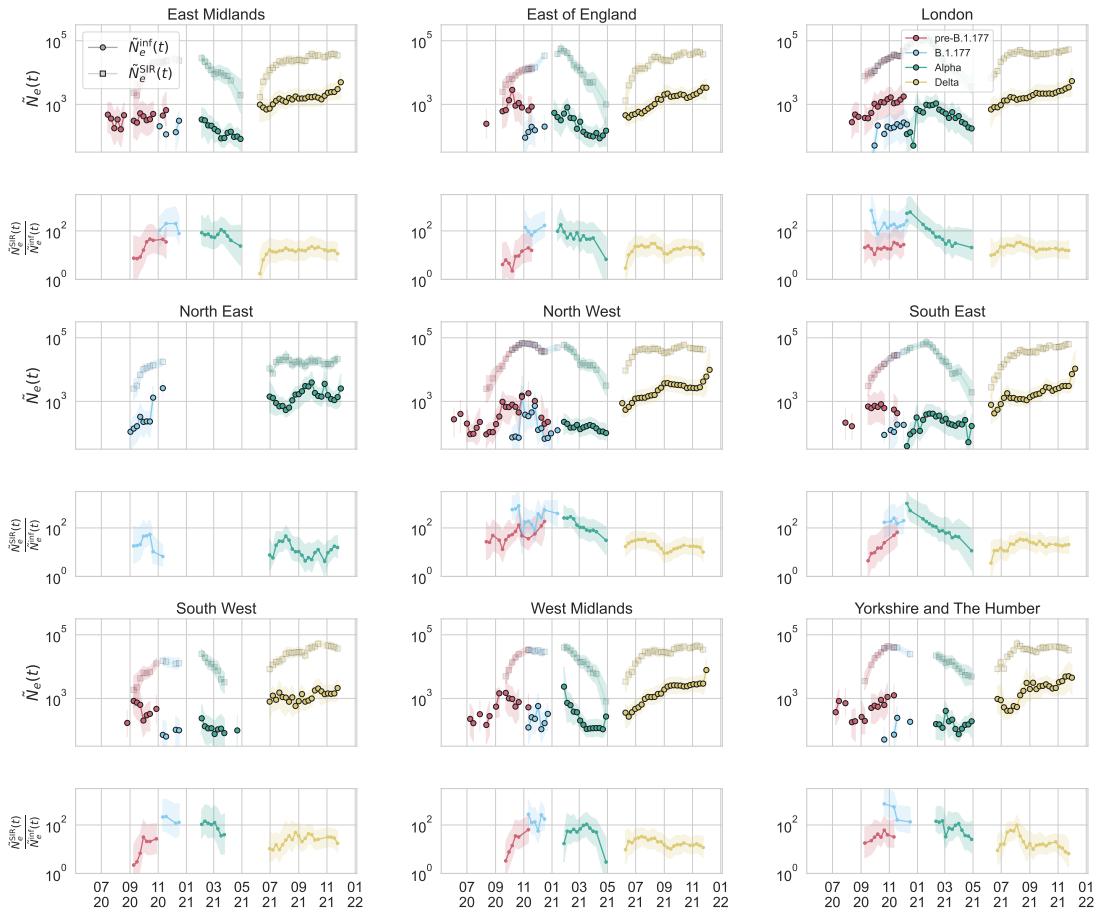


Figure S2: Inferred scaled effective population size by region in England, compared to that of an SIR model as calculated using the observed number of positives at the community level in that region reported by the COVID-19 Infection Survey [31] and the observed effective reproduction number in that region reported by the UK Health Security Agency [38].



Figure S3: Inferred scaled effective population size by region in England, compared to number of positives at the community level in that region reported by the COVID-19 Infection Survey [31].

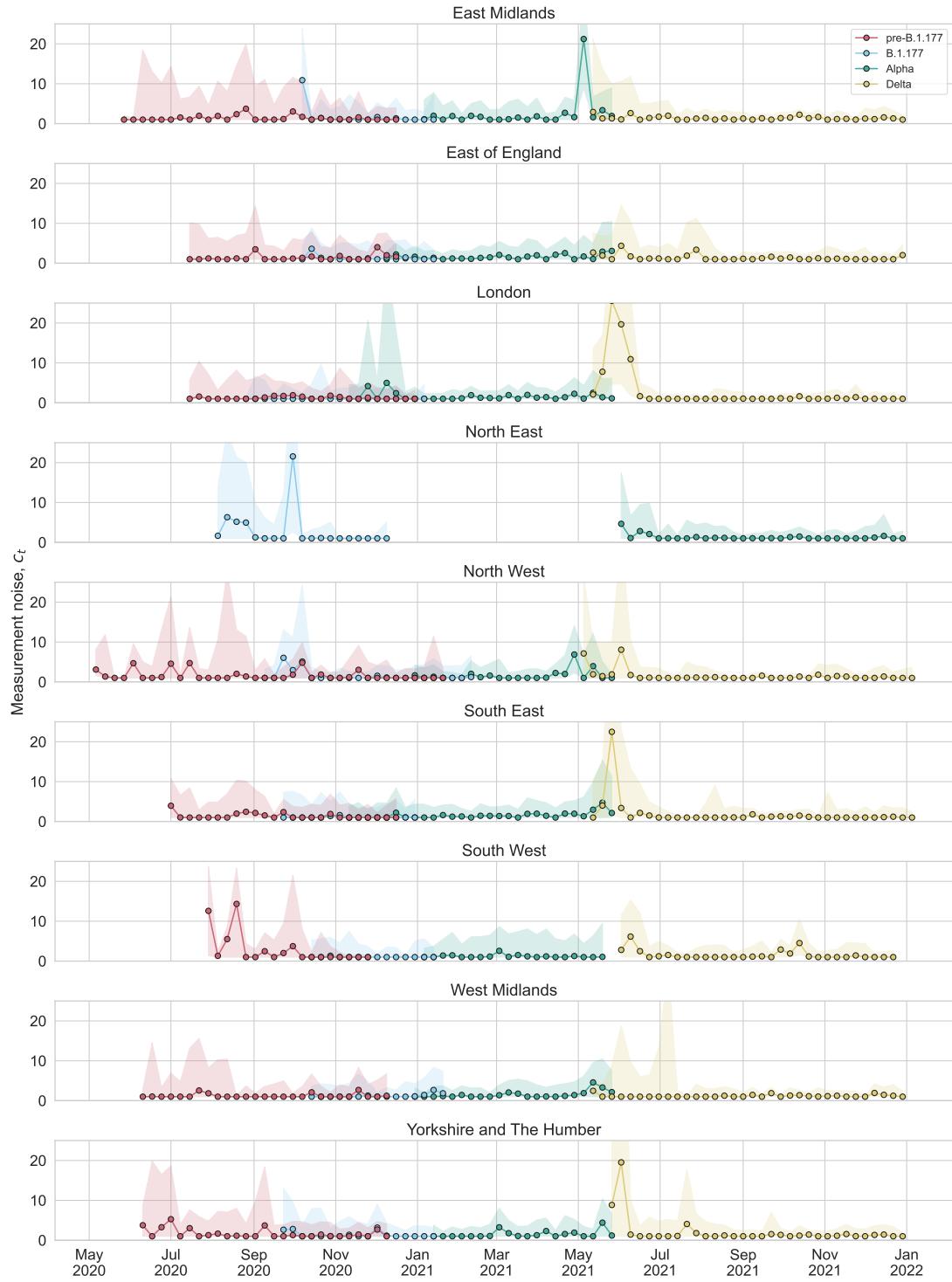


Figure S4: Inferred measurement noise by region in England.

983 Additional supplementary tables and figures

Date	Location	Method	$\langle Z \rangle$	$\text{Var}(Z)$	$k$	Reference
Beginning of pandemic to February 27 2020	Worldwide excluding China	Branching process model of number of imported and local cases	2.5 (,)	65 (33.75, 127.5)	0.1 (0.05, 0.2)	[42]
March 1 to May 3 2020	Georgia (USA)	Spatiotemporal transmission model fit to multiple data sources	2 (0.5, 3.5)	12.26 (0.88, 101.5)	0.39 (0.125, 0.65)	[73]
March 1 to November 1 2020	Denmark	Model fitting the case numbers across multiple regions	1.1 (0.8, 1.4)	12.1 (4.36, 25.9)	0.11 (0.08, 0.18)	[74]
Beginning of pandemic until January 18 2020	China (Wuhan)	Stochastic simulations fit to infected cases	2.2 (1.4, 3.8)	11.16 (1.68, 1035.2)	0.54 (0.014, 6.95)	[75]
August to September 2020	UK	Model using empirical viral load trajectories and contact numbers	1.21 (0.84, 2.51)	7.07 (2.65, 44.51)	0.25 (0.15, 0.39)	[43]
May 15 to August 1 2020	Tamil Nadu and Andhra Pradesh (India)	Contact tracing and incidence	1.25 (1.1, 1.4)	4.31 (3.43, 5.4)	0.51 (0.49, 0.52)	[76]
January to February 2021	UK	Model using empirical viral load trajectories and contact numbers	0.54 (0.4, 1.03)	1.42 (0.66, 9.19)	0.33 (0.13, 0.61)	[43]
January 23 to April 28 2020	Hong Kong	Contact tracing	0.58 (,)	1.36 (,)	0.43 (,)	[77]
January 16 to April 3 2020	Human (China)	Contact tracing	0.4 (0.35, 0.47)	0.93 (0.66, 1.43)	0.3 (0.23, 0.39)	[78]
January 14 to February 12 2020	Shenzhen (China)	Contact tracing	0.4 (0.3, 0.5)	0.68 (0.38, 1.21)	0.58 (0.35, 1.18)	[79]

Table S1: Overdispersion values from the literature ordered from highest to lowest variance in offspring number. Any error intervals that are reported are taken from the reference (sometimes defined differently). The estimate taken from Ref. [43] assumes no self-isolation upon symptom onset and no testing; lifting these assumptions leads to similar or lower overdispersion.

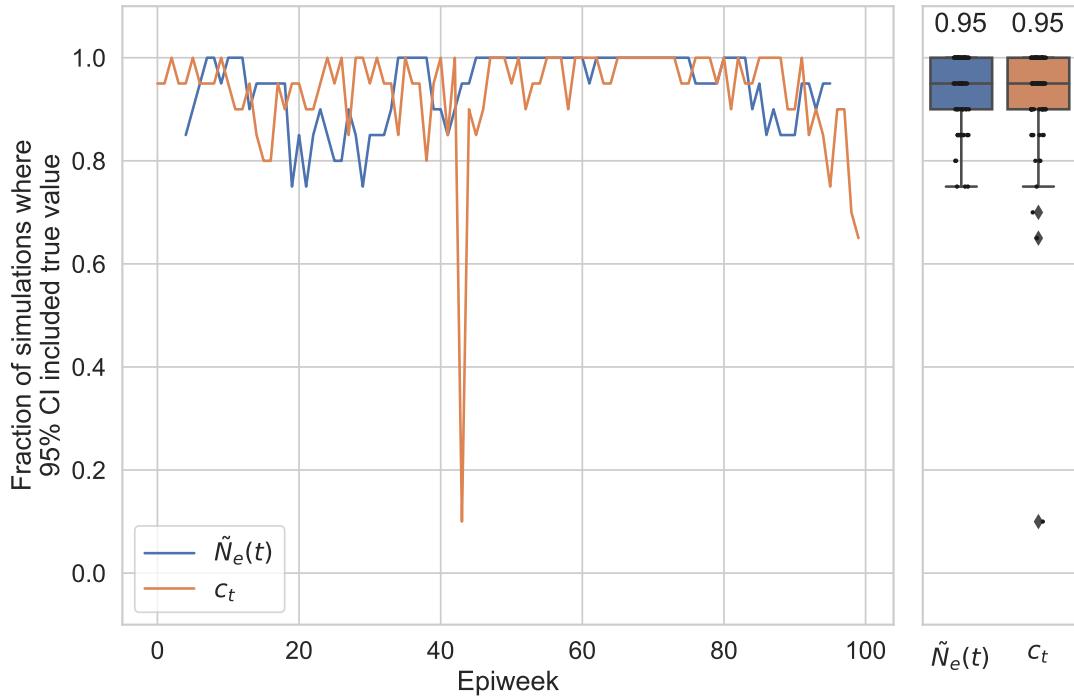


Figure S5: The fraction of simulations (20 total) where the inferred 95% confidence interval for  $\tilde{N}_e(t)$  or  $c$  included the true value (left) by timepoint and (right) for all timepoints. (Right) Boxes indicate the quartiles and the line inside the box (and number above) indicates the median. Whiskers indicate the extreme values excluding outliers. Simulation parameters are specified in the Methods and Figure 1, which shows a single simulation instance. For the inference, we created superlineages randomly 20 times.

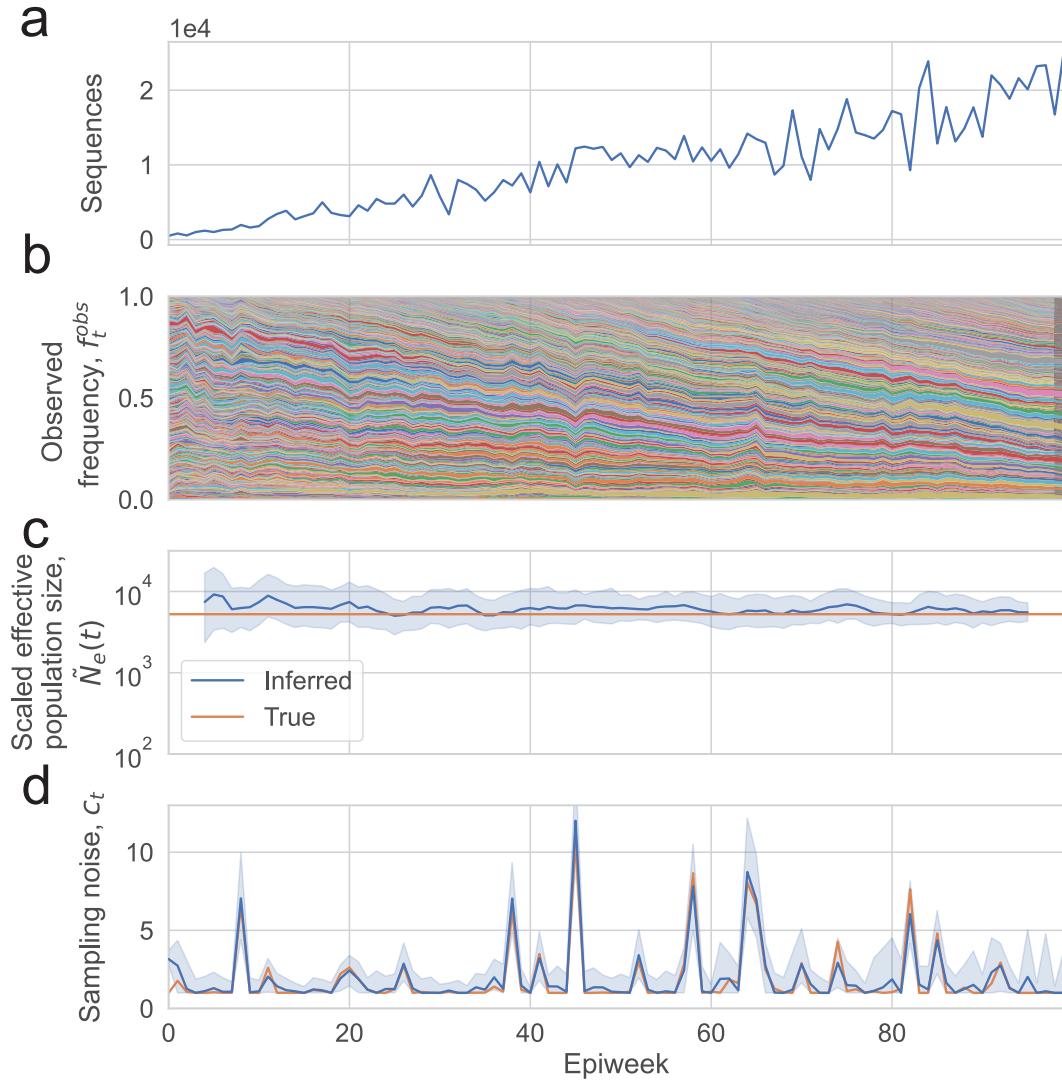


Figure S6: Wright-Fisher simulations where  $\tilde{N}_e(t)$  is constant over time, and the inferred  $\tilde{N}_e(t)$  and  $c_t$ . (a) Number of sequences sampled. (b) Simulated lineage frequency trajectories. (c) Inferred effective population size ( $\tilde{N}_e(t)$ ) on simulated data compared to true values. (d) Inferred measurement noise ( $c_t$ ) on simulated data compared to true values. In (c) the shaded region shows the 95% confidence interval calculated using the posterior, and in (d) the shaded region shows the 95% confidence interval calculated using bootstrapping (see Methods).

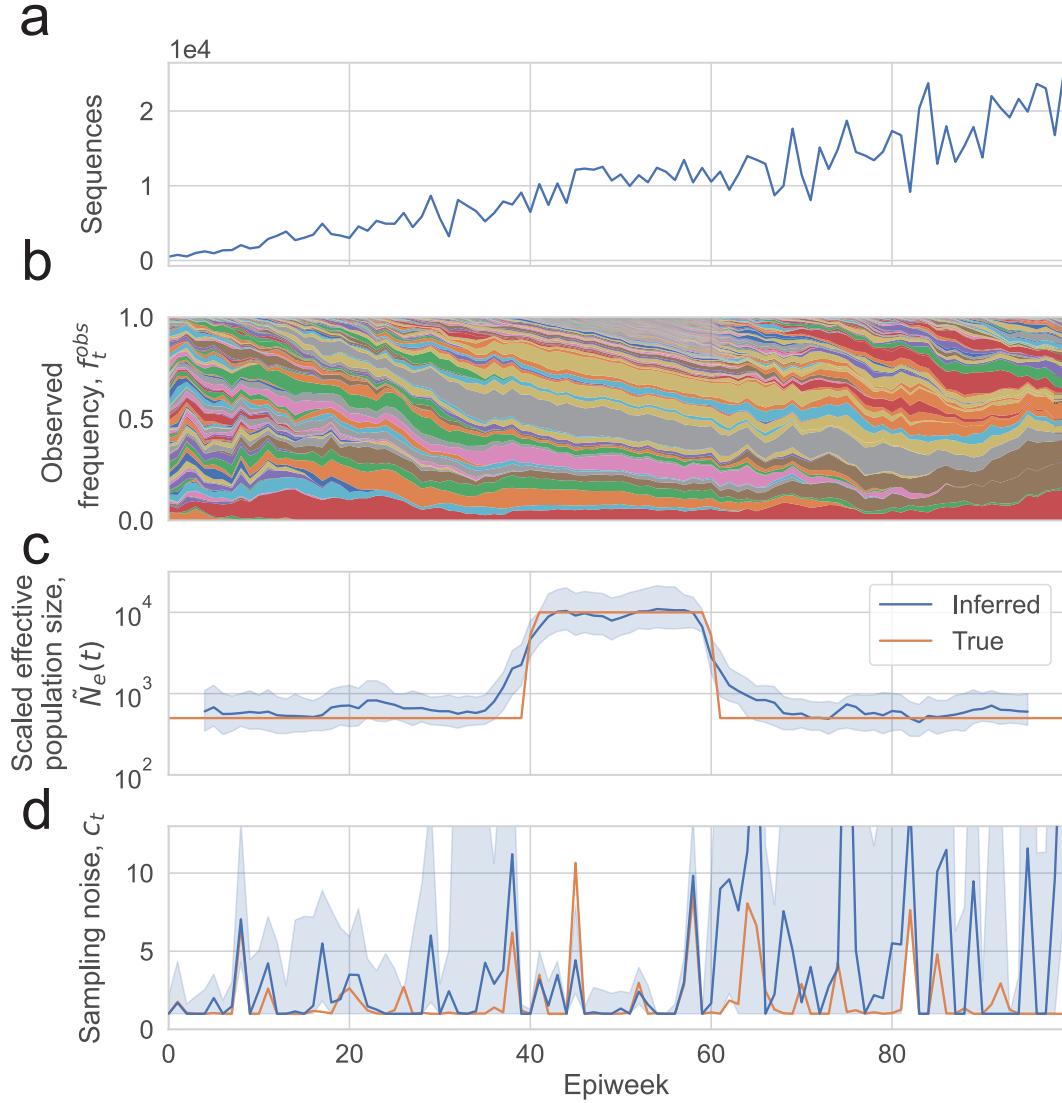


Figure S7: Wright-Fisher simulations where  $\tilde{N}_e(t)$  changes over time according to a rectangular function, and the inferred  $\tilde{N}_e(t)$  and  $c_t$ . (a) Number of sequences sampled. (b) Simulated lineage frequency trajectories. (c) Inferred effective population size ( $\tilde{N}_e(t)$ ) on simulated data compared to true values when  $c_t$  is jointly inferred and when  $c_t$  is fixed at 1 (uniform sampling). (d) Inferred measurement noise ( $c_t$ ) on simulated data compared to true values. In (c) the shaded region shows the 95% confidence interval calculated using the posterior, and in (d) the shaded region shows the 95% confidence interval calculated using bootstrapping (see Methods).

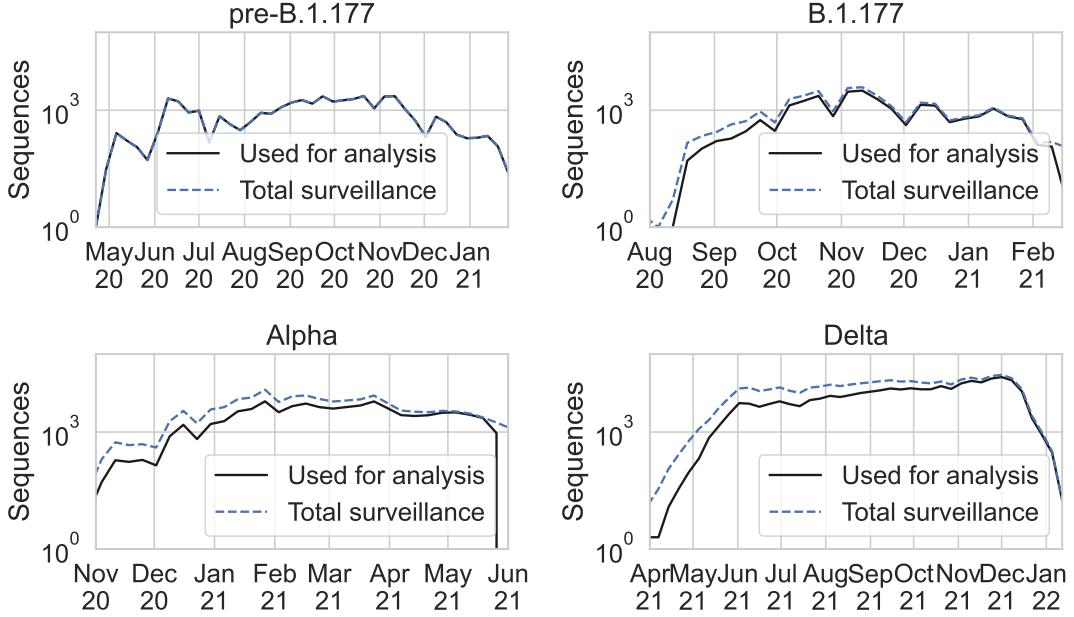


Figure S8: Total number of surveillance sequences of each variant in the metadata from COG-UK downloaded on January 16, 2022 and the number of sequences used in the analysis for each variant or group of lineages (determined by the number of sequences included in the tree, and the number of sequences which could be grouped into sublineages based on the procedure described in the Methods).

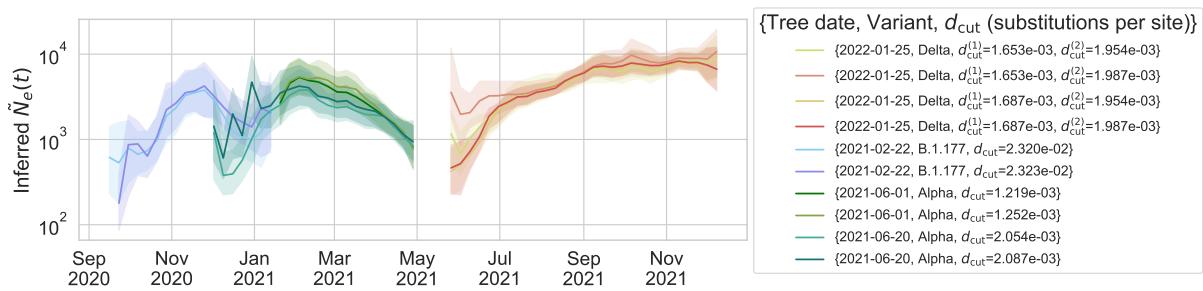


Figure S9: Varying the date of the tree downloaded from COG-UK and the depth at which the tree is cut for creating lineages ( $d_{cut}$ , see Methods) does not substantially change the inferred scaled effective population size. The the tree date and depth used in the main text are  $\{2021-02-22, B.1.177, d_{cut} = 2.323 \cdot 10^{-2}\}$ ,  $\{2021-06-20, \text{Alpha}, d_{cut} = 2.054 \cdot 10^{-3}\}$ ,  $\{2022-01-25, \text{Delta}, d_{cut}^{(1)} = 1.687 \cdot 10^{-3}, d_{cut}^{(2)} = 1.954 \cdot 10^{-3}\}$ . The color of the lines for the parameters that were used in the main text are the same as those shown in Figure 2.

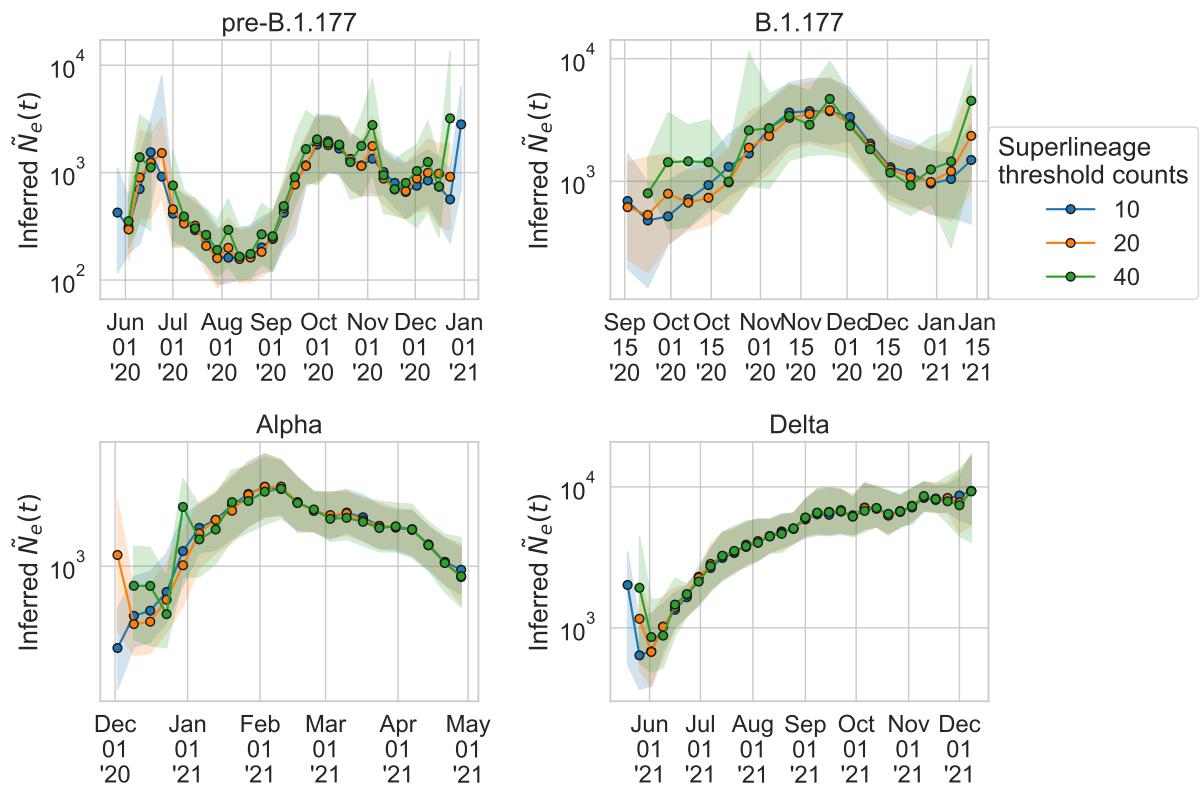


Figure S10: Varying the threshold counts for forming superlineages (see Methods) does not substantially change the inferred scaled effective population size. The superlineage threshold counts used in the main text is 20.

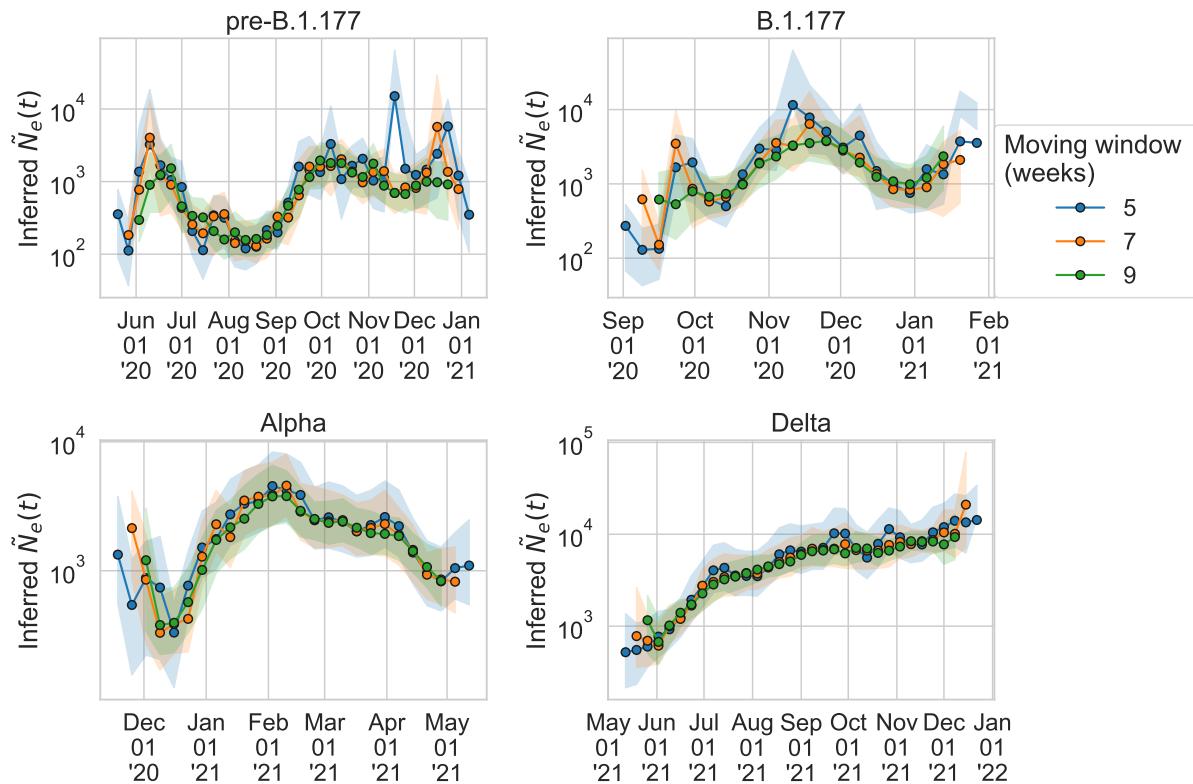


Figure S11: Varying the number of weeks in the moving window does not substantially change the inferred scaled effective population size. The size of the moving window used in the main text is 9 weeks.

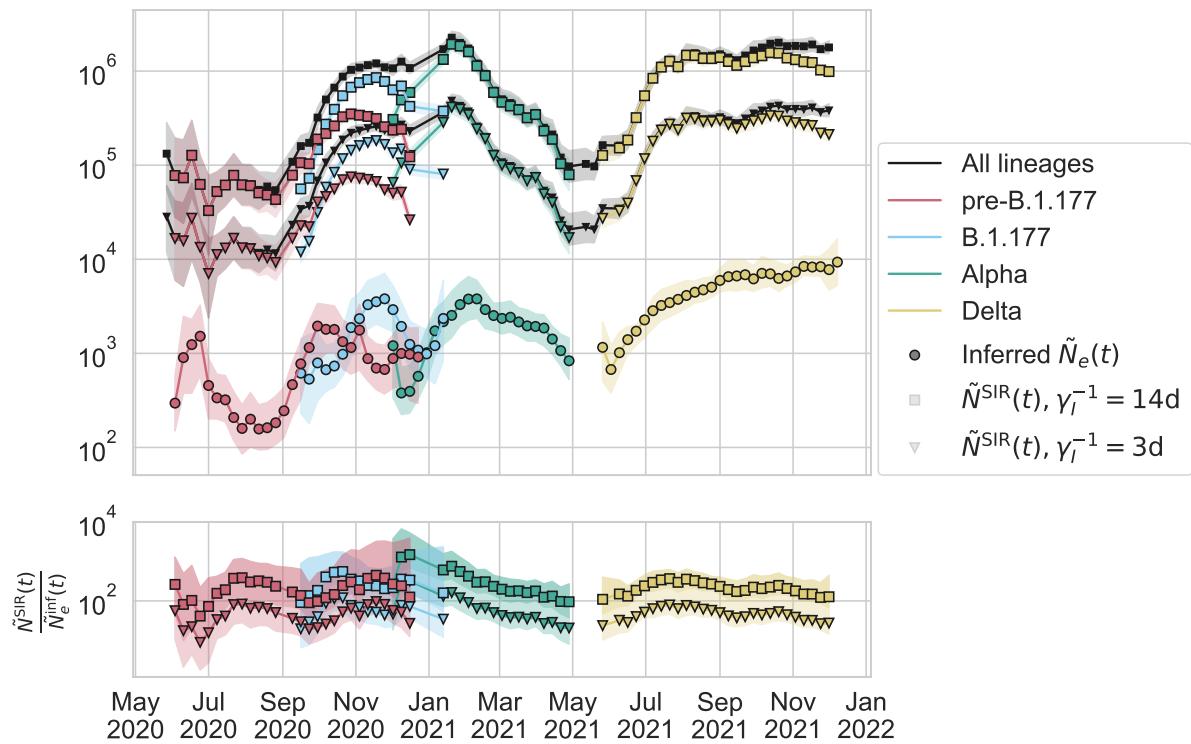


Figure S12: Varying the rate of transitioning from infected to recovered within literature ranges ( $\gamma_I = 3$  to 14 days) used for calculation of the SIR model  $\tilde{N}_e(t)$  (Methods) does not substantially decrease the observed ratio  $\tilde{N}_e^{\text{SIR}}(t)/\tilde{N}_e^{\text{inf}}(t)$ .

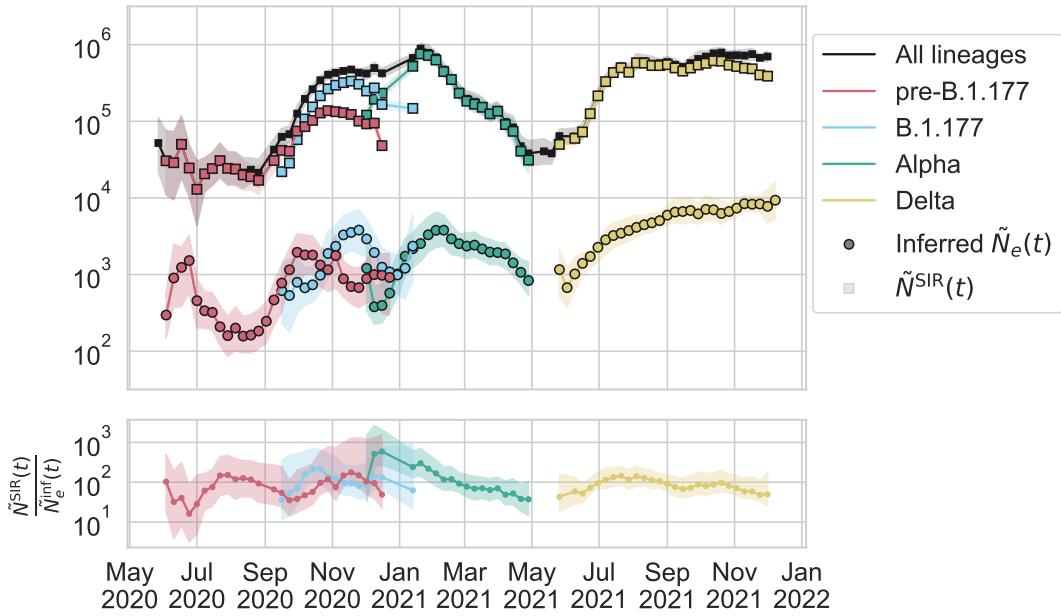


Figure S13: Inferred scaled effective population size compared to the SIR model scaled population size calculated using the observed number of positive individuals in England (see Methods).

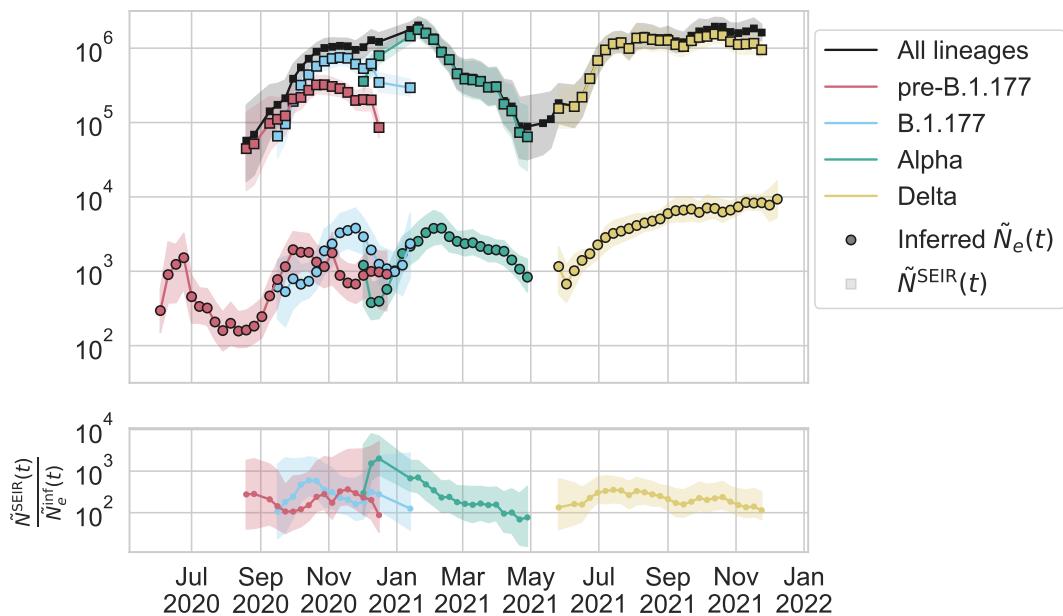


Figure S14: Inferred scaled effective population size compared to the SEIR model scaled population size calculated using the observed number of positive individuals in England (see Methods).

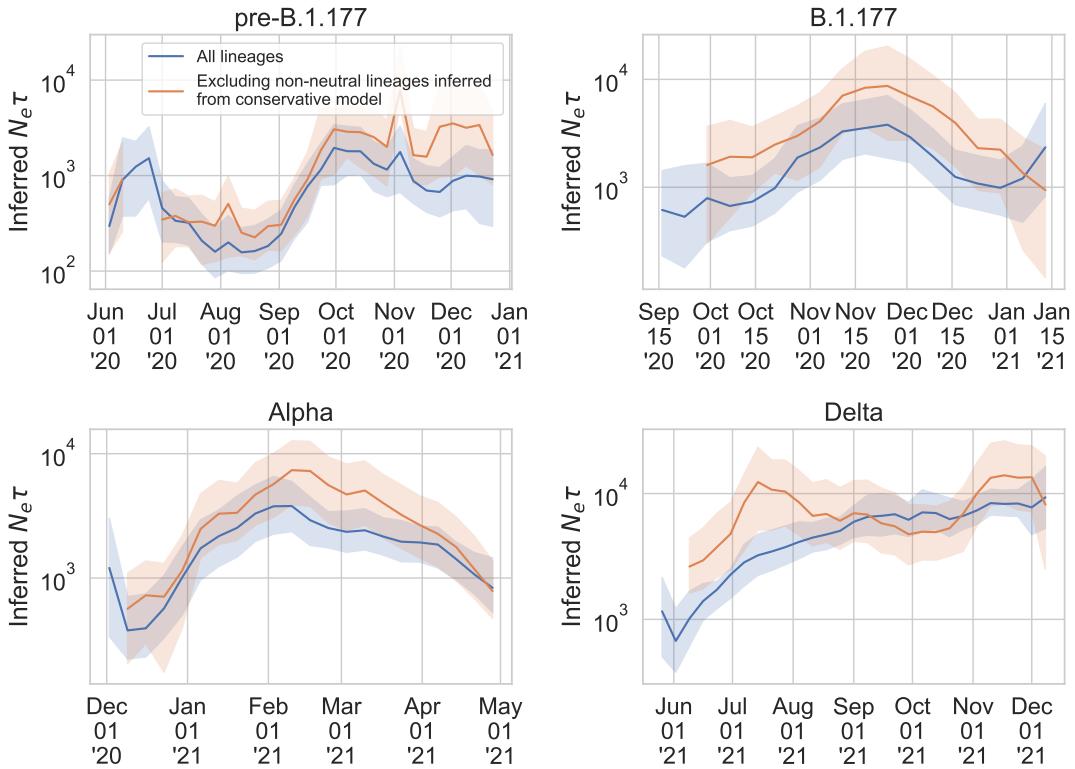


Figure S15: Inferred scaled effective population size of all lineages compared to that when excluding non-neutral lineages detected by conservative method (assumes no genetic drift).

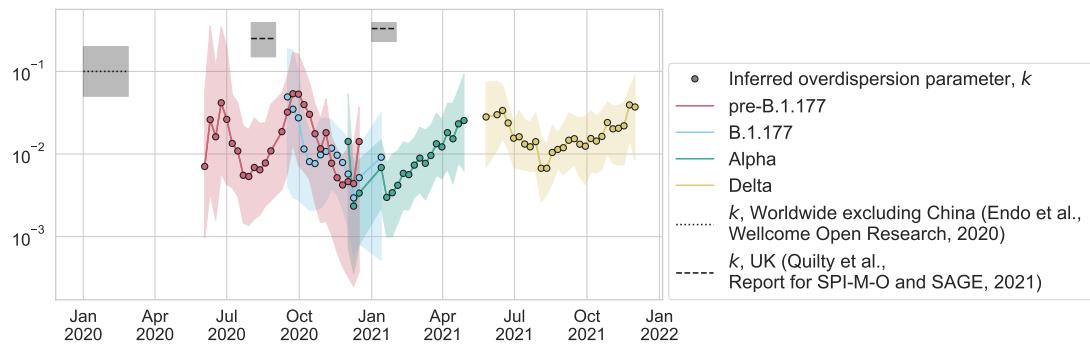


Figure S16: Same as Figure 3c, but plotting the overdispersion parameter,  $k = \frac{R_t}{\frac{\sigma^2}{R_t} - 1}$ , where  $R_t$  is the effective reproduction number and  $\sigma^2$  is the variance in offspring number. The circles show the inferred overdispersion parameter if we assume there is only superspreading and no deme structure. For the inferred overdispersion parameter, the estimated effective reproduction number in England by variant (see Methods) is used for  $R_t$ , and the ratio between the SIR model population size and the inferred effective population size is used for  $\sigma^2$ . The shaded area for the inferred overdispersion parameter  $k$  gives an estimate of the error and is calculated by combining minimum or maximum values of the individual parameters; note that this does not correspond to a particular confidence interval.

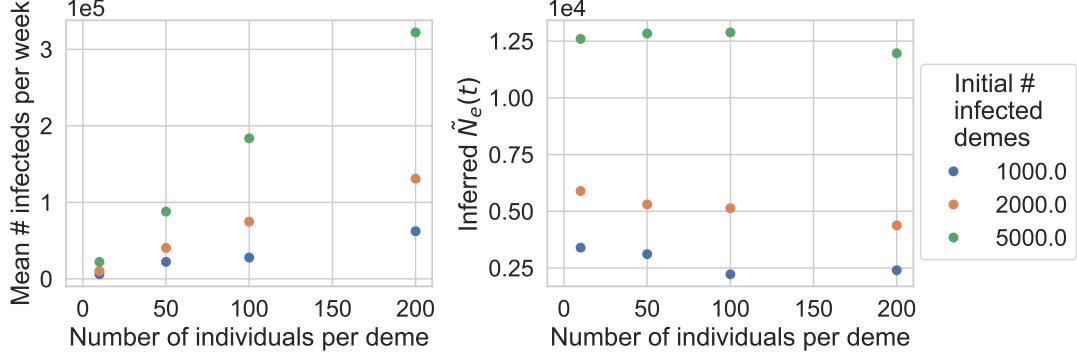


Figure S17: Simulations of deme structure (described in main text and Methods). (a) The mean number of infected individuals per week from Weeks 42 to 50. (b) The inferred  $\tilde{N}_e(t)$  using lineage trajectories from Weeks 42 to 50.

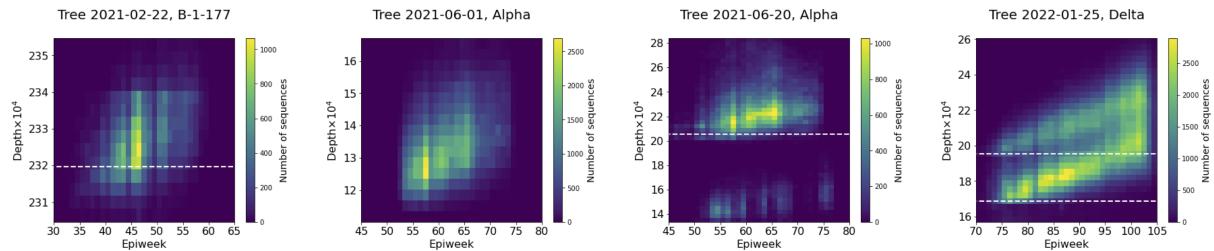


Figure S18: Sample epiweeks versus tree depths. In a phylogenetic tree, the number of sequences (leaf nodes) of a focal variant that fall within specific epiweek and tree depth ranges is counted and summarized as a two-dimensional histogram. The tree depth is the substitution rate measured in units of substitutions per site, with respect to the most recent common ancestor. From left to right, the phylogenetic tree (specified by date created by COG-UK, using the sequences available at the time) and focal variant are {2021-02-22, B-1-177}, {2021-06-01, Alpha}, {2021-06-20, Alpha}, and {2022-01-25, Delta}. Weeks are counted from 2019-12-29. The dashed horizontal lines indicate the values of  $d_{\text{cut}}$  ( $d_{\text{cut}}^{(1)}$  and  $d_{\text{cut}}^{(2)}$  for the Delta variant) used for the results presented in the main text.

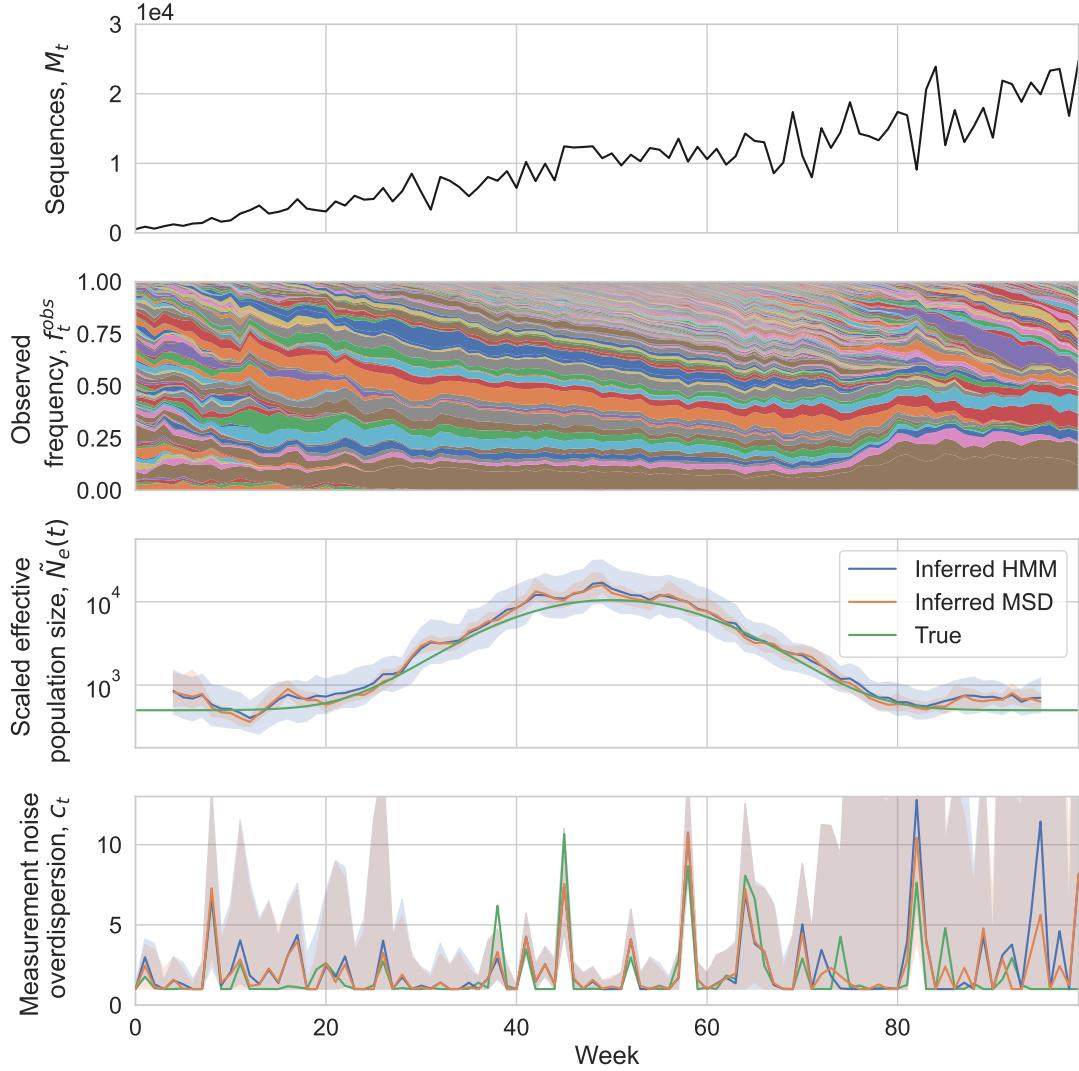


Figure S19: Comparing the inferred  $\tilde{N}_e(t)$  and  $c_t$  in Wright-Fisher simulations using the method of moments and maximum likelihood estimation approaches (see Methods). (a) Number of sequences sampled. (b) Simulated lineage frequency trajectories. (c) Inferred effective population size ( $\tilde{N}_e(t)$ ) on simulated data using the method of moments (MSD, for mean squared displacement) and maximum likelihood (HMM, for Hidden Markov Model) estimation approaches compared to true values. The shaded region shows the 95% confidence interval of the inferred values. The confidence interval using the method of moments approach was calculated by taking the middle 95% of values when bootstrapping over the superlineages. The confidence interval using the maximum likelihood estimation approach was determined using the posterior (see Methods) and takes into account joint errors in  $c_t$  and  $\tilde{N}_e(t)$ . (d) Inferred measurement noise ( $c_t$ ) on simulated data using the method of moments and maximum likelihood estimation approaches compared to true values. The shaded region shows the 95% confidence interval calculated using bootstrapping (see Methods).

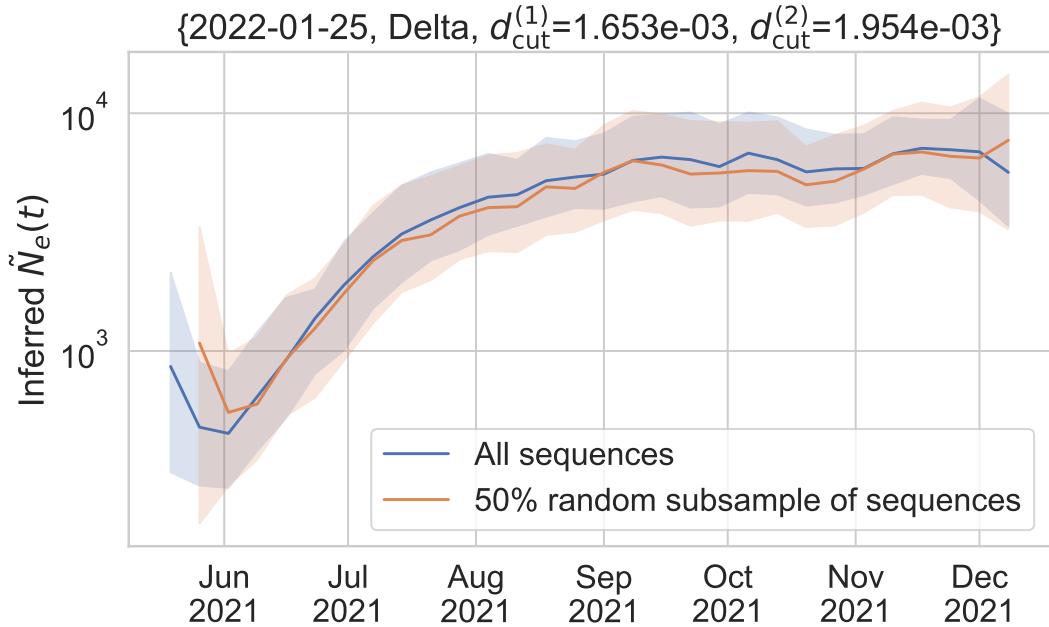


Figure S20: Randomly subsampling half of the Delta sequences used for the analysis does not substantially change the inferred scaled effective population size.

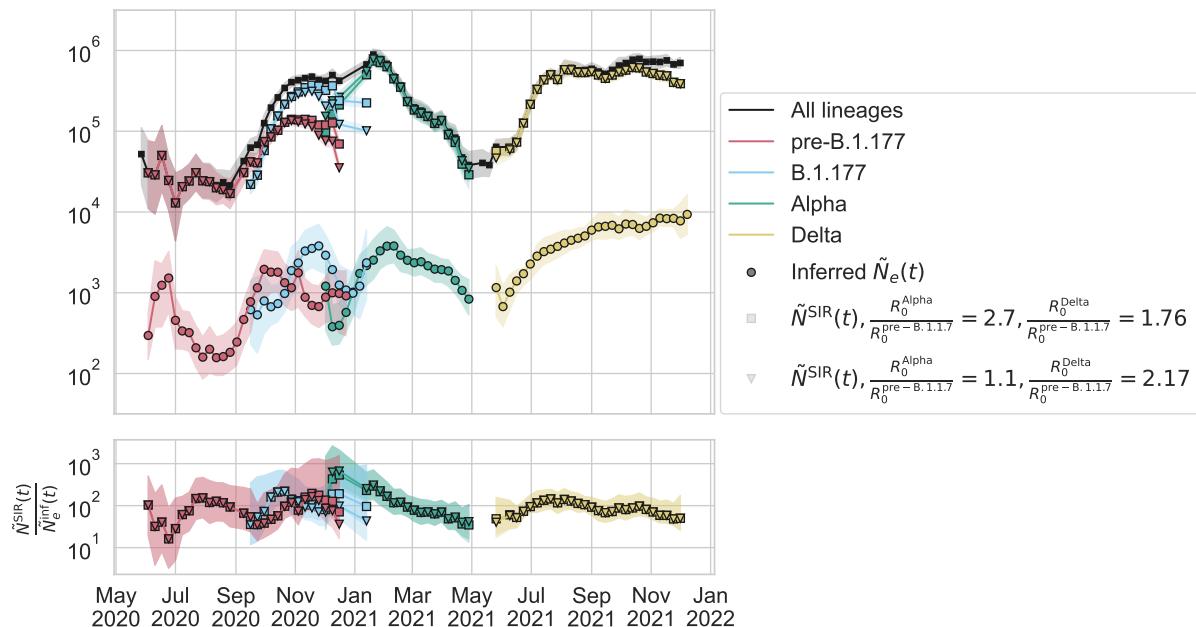


Figure S21: Varying the values of the basic reproduction number within literature ranges ( $\frac{R_0^{\text{Alpha}}}{R_0^{\text{pre-B.1.177}}} = 1.1 - 2.7$  [17],  $\frac{R_0^{\text{Delta}}}{R_0^{\text{pre-B.1.177}}} = 1.76 - 2.17$  [65]) used for calculation of the SIR model  $\tilde{N}_e(t)$  by variant (Methods) does not substantially affect the calculated  $\tilde{N}^{\text{SIR}}(t)$ .

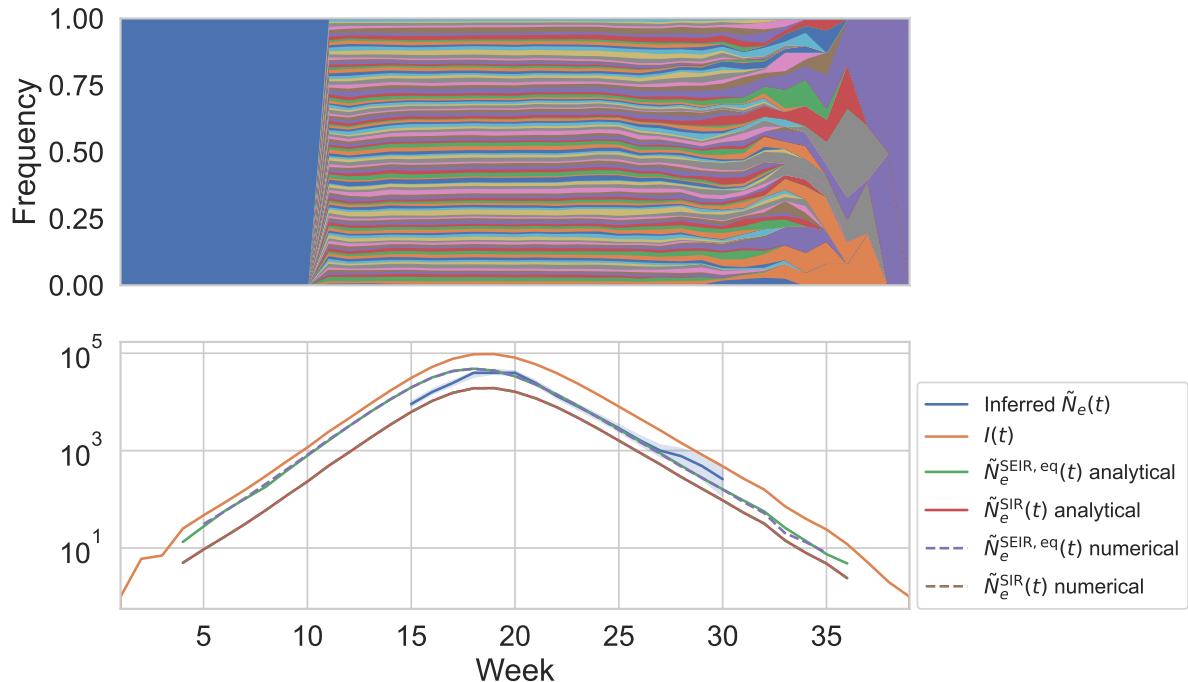


Figure S22: Simulations of stochastic SEIR dynamics without measurement noise, and comparison of the inferred  $\tilde{N}_e(t)$  to Equations 1 and 51 when the reported positive individuals include only the infectious individuals. (Top) Muller plot of simulated infectious individuals' lineage trajectories (simulations described in Methods). Infectious individuals are randomly assigned a lineage in week 11, and individuals that they transmit to are infected with the same lineage. The blue lineage before week 11 indicates the infectious individuals that existed before lineages were assigned. (Bottom) Comparison of the inferred  $\tilde{N}_e(t)$  using the lineage trajectories shown in the top panel to the number of infectious individuals  $I(t)$ , Equation 51 (SEIR model  $\tilde{N}_e(t)$  at equilibrium), and Equation 1 (SIR model  $\tilde{N}_e(t)$ ) calculated analytically or numerically as described in the Methods. The numerical solutions give the same results as the analytical solutions.

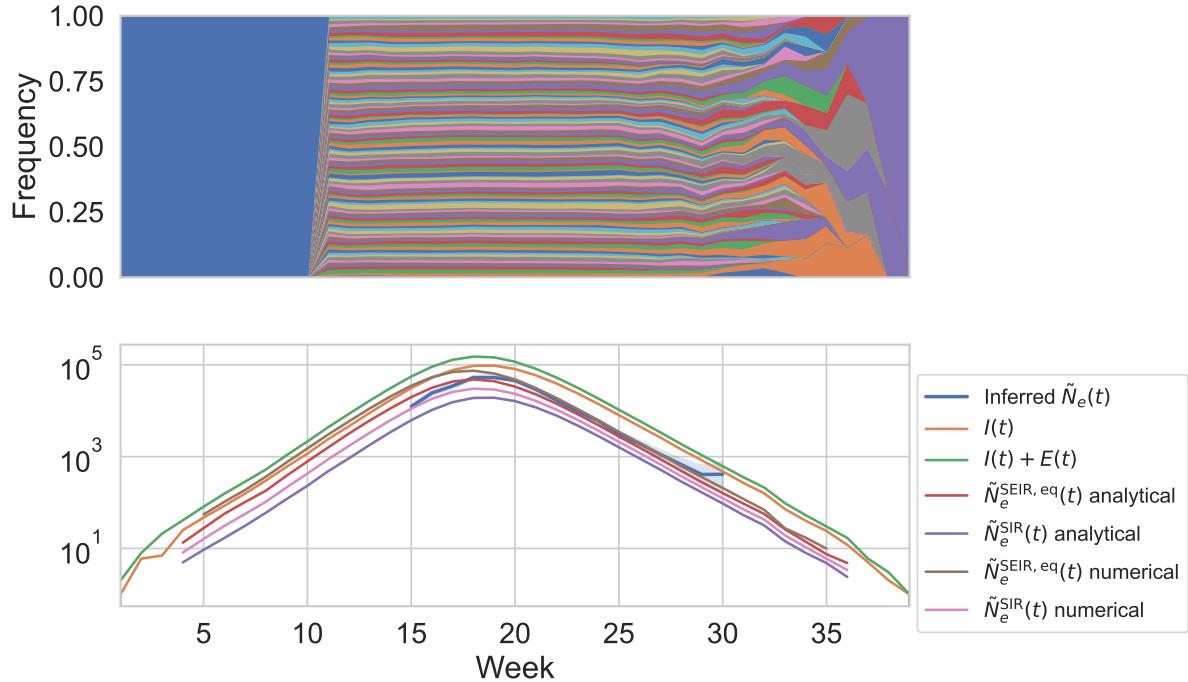


Figure S23: Simulations of stochastic SEIR dynamics without measurement noise, and comparison of the inferred  $\tilde{N}_e(t)$  to Equations 1 and 51 when the reported positive individuals include both infectious and exposed individuals. (Top) Muller plot of simulated infectious and exposed individuals' lineage trajectories (simulations described in Methods). Infectious and exposed individuals are randomly assigned a lineage in week 11, and individuals that they transmit to are infected with the same lineage. The blue lineage before week 11 indicates the infectious and exposed individuals that existed before lineages were assigned. (Bottom) Comparison of the inferred  $\tilde{N}_e(t)$  using the lineage trajectories shown in the top panel to the number of infectious individuals  $I(t)$ , the sum of the number of infectious and exposed individuals  $I(t) + E(t)$ , Equation 51 (SEIR model  $\tilde{N}_e(t)$ ), and Equation 1 (SIR model  $\tilde{N}_e(t)$ ) calculated analytically or numerically as described in the Methods. The numerical solutions give slightly higher  $\tilde{N}_e(t)$  as compared with the analytical solutions; however, the numerical solutions to the SEIR and SIR models bound the inferred  $\tilde{N}_e(t)$ .