Northumbria Research Link

Citation: du Plessis, Louis, McCrone, John T., Zarebski, Alexander E., Hill, Verity, Ruis, Christopher, Gutierrez, Bernardo, Raghwani, Jayna, Ashworth, Jordan, Colquhoun, Rachel, Connor, Thomas R., Faria, Nuno R., Jackson, Ben, Loman, Nicholas J., O'Toole, Áine, Nicholls, Samuel M., Parag, Kris V., Scher, Emily, Vasylyeva, Tetyana I., Volz, Erik M., Watts, Alexander, Bogoch, Isaac I., Khan, Kamran, Aanensen, David M., Kraemer, Moritz U. G., Rambaut, Andrew, Pybus, Oliver G., COVID-19 Genomics UK (COG-UK) Consortium, , Bashton, Matthew, Smith, Darren, Young, Greg and Nelson, Andrew (2021) Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. Science, 371 (6530). pp. 708-712. ISSN 0036-8075

Published by: American Association for the Advancement of Science

URL: https://doi.org/10.1126/science.abf2946 <https://doi.org/10.1126/science.abf2946>

This version was downloaded from Northumbria Research Link: http://nrl.northumbria.ac.uk/id/eprint/45196/

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: http://nrl.northumbria.ac.uk/policies.html

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

Northumbria University NEWCASTLE

CUSTOMER SERVICE EXCELLENCE CSE

University Library

science.sciencemag.org/cgi/content/full/science.abf2946/DC1

Supplementary Materials for

Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK

Louis du Plessis, John T. McCrone, Alexander E. Zarebski, Verity Hill, Christopher Ruis, Bernardo Gutierrez, Jayna Raghwani, Jordan Ashworth, Rachel Colquhoun, Thomas R. Connor, Nuno R. Faria, Ben Jackson, Nicholas J. Loman, Áine O'Toole, Samuel M. Nicholls, Kris V. Parag, Emily Scher, Tetyana I. Vasylyeva, Erik M. Volz, Alexander Watts, Isaac I. Bogoch, Kamran Khan, COVID-19 Genomics UK (COG-UK) Consortium, David M. Aanensen, Moritz U. G. Kraemer, Andrew Rambaut*, Oliver G. Pybus*

*Corresponding author. Email: a.rambaut@ed.ac.uk (A.R.); oliver.pybus@zoo.ox.ac.uk (O.G.P.)

Published 8 January 2021 on *Science* First Release
DOI: 10.1126/science.abf2946

This PDF file includes:

Materials and Methods
List of COG-UK Consortium members
Figs. S1 to S20
Tables S1 to S5
Captions for data S1 to S5
References

Other supplementary material for this manuscript includes:

Data S1, S2, S4, and S5 (Excel format)
Data S3 (.tsv format)
MDAR Reproducibility Checklist (PDF)

Materials and Methods

Genomic data

All SARS-CoV-2 genomes available on GISAID (*14*) on 23 June 2020 were downloaded and combined with all SARS-CoV-2 genomes sequenced by the COG-UK consortium (*15*) by 26 June 2020 (available at https://www.cogconsortium.uk/data/). The pipeline used to collect and process raw SARS-CoV-2 sequence data and sample-associated metadata across the national COG-UK network is described in (*39*). Among the genomes sequenced by the COG-UK consortium, approximately 59.2% were sequenced using Illumina sequencing technology and 27.5% using Oxford Nanopore Technologies (ONT). The sequencing platform was not recorded for 13.3% of samples. Of the samples sequenced on the Illumina platform, approximately 50.8% were sequenced on the NovaSeq, 31.6% on NextSeq, 13.5% on MiSeq and 4.1% on HiSeq. Of the samples sequenced by ONT approximately 89% were sequenced on GridION and 11% on MinION. All sequencing sites except Oxford used the ARTIC protocol (*40*) for amplification of SARS-CoV-2 samples. Raw sequence data of all SARS-CoV-2 genomes sequenced by the COG-UK consortium are available from the European Nucleotide Archive (ENA) at EMBL-EBI under the accession number PRJEB37886 (https://www.ebi.ac.uk/ena/browser/view/PRJEB37886). Mutations, insertions and deletions among genomes sequenced by the COG-UK consortium can be visualized using *CoV-GLUE* (http://cov-glue.cvr.gla.ac.uk; *41*). We removed sequences that were from duplicate or environmental samples, those without exact collection dates, and those with large clusters of substitutions or large indels. Each genome sequence was aligned to the reference (Wuhan-Hu-1, GenBank: MN908947.3) using *minimap v2.17* (*42*) and the resulting SAM alignment was converted to a FASTA alignment, with the 5' and 3' UTRs of each genome masked by Ns. Insertions relative to the reference were discarded and site 11,083 (site position relative to MN908947), which is globally homoplasic, was also masked. Genomes that contained >5% Ns after mapping and those with a genetic distance to WH04 (GISAID: EPI_ISL_406801) more than 4 standard deviations from the epi-week mean genetic distance to WH04 were discarded. The final dataset consisted of 50,887 genomes sampled between 24 December 2019 and 22 June 2020, of which 26,181 (~51%) were from the UK (see **Fig. 1A**). Accession numbers for all 50,887 genomes are provided on the GitHub repository (*38*; https://github.com/COG-UK/uk-intros-analyses).

Geographical metadata

Administrative level 2 (admin2) metadata for the sampling location of UK virus genome sequences in the dataset (roughly equivalent to counties in the UK) required cleaning in order to be mapped to official admin2 regions, as found in the Global Administrative Database (GADM, https://gadm.org).

Some sampling locations in the metadata could not be unambiguously mapped to a known location (e.g. "City Centre"), while others were for locations in overseas territories (e.g. Falklands and Gibraltar). Yet other genome sequences had uninformative spatial records (e.g. Yorkshire or Wales), or no admin2 level data at all. For these (3431 of 26,181) the admin2 region was not mapped. We carried out a simple one-to-one mapping where possible, which included correcting spelling mistakes and alternative entries for the same county (e.g. Durham versus County Durham). Locations recorded at a higher spatial resolution were mapped to the corresponding admin2 region (e.g. Solihull was mapped to Birmingham). Where the recorded

locations were larger than the admin2 regions (e.g. "West Midlands"), and most of the sequences in the area were from this larger conglomeration as opposed to its higher-resolution components, these admin2 regions were combined. When creating the map figures, we also merged some city authorities with no reported sequences with their surrounding county, on the assumption that the larger county was used to represent the location of city samples (e.g. for Leicester and Leicestershire). Finally, genome sequences from Northern Ireland reported locations as historical counties, rather than the official admin2 designations, and so these historical counties were used instead.

The lookup table showing the metadata-to-GADM location mapping is provided in **Data S1** and a Jupyter notebook containing the cleaning code is provided on the GitHub repository (*38*; https://github.com/COG-UK/uk-intros-analyses).

Phylogenetic analysis and molecular clock dating

We developed a new Bayesian molecular clock phylogenetic analysis pipeline in order to reconstruct a posterior set of time-scaled phylogenetic trees for our exceptionally large virus genome dataset. Using the standard Bayesian approach it is currently impractical to estimate time-scaled trees directly from genome sequence data for more than a few thousand sequences. Therefore, we employed a number of extensions to make the analysis tractable.

First, we divided the full genome sequence dataset (n=50,887) into five smaller datasets. Genomes were assigned SARS-CoV-2 lineages according to the nomenclature defined in *Rambaut et al. (43)* using *Pangolin* (*44;* github.com/cov-lineages/pangolin). Each lineage (and its sublineages) represents a monophyletic clade in the global SARS-CoV-2 phylogeny and can thus be analysed independently. For each lineage in A (n=3591), B (n=8821), B.1 (n=22,861), B.1.1 (n=15,616), we estimated an approximately maximum-likelihood tree using the Jukes-Cantor model in *FastTree v2.1.10* (*45*), then collapsed branch lengths shorter than 5×10^{-6} substitutions per site, which corresponded to distances smaller than one substitution across the whole virus genome, and likely result from nucleotide ambiguity codes in the genome sequences. By pruning out a large monophyletic clade the maximum-likelihood tree for B.1 was further divided into two trees, B.1.pruned (n=12,275) and B.1.X (n=10,586).

Prior to analysing the full dataset, an initial analysis was performed on a subset of genomes to obtain estimates of the molecular clock rate and of the TMRCA (time of the most recent common ancestor) of each large-scale phylogenetic tree defined above. The full dataset was subsampled as evenly as possible across epi-weeks and countries with a slight enrichment for samples immediately descended from five large polytomies in the global phylogeny. For each of these nodes, we always included the five oldest genomes, the most recent genome sequence and five other immediate descendants that were randomly chosen. The remaining genomes were sampled by allocating an even number of sequences per epi-week while maintaining a dataset size of <1,000 genomes. For each epi-week, genomes were sampled evenly by country until either its allocation was exhausted or there were no remaining genomes available. This subsampled dataset was analysed in *BEAST 1.10* (*46*) using a GTR+G+F substitution model, with a strict molecular clock model using a non-informative continuous-time Markov chain (CTMC) prior (*47*) and a Skygrid coalescent tree prior (*48*) with 40 grid points, roughly corresponding to weeks between 1 October 2019 and 2 July 2020. In the analysis, monophyly

constraints were used to ensure that the clades corresponding to the large-scale phylogenetic trees identified in the previous step were monophyletic. We combined four independent Markov Chain Monte Carlo (MCMC) chains that were each run for 40 million steps, discarding the first 4 million steps of each chain as burn-in and resampling states every 4000 steps. Convergence was assessed using *Tracer* (*49*).

Next, we applied a commonly used approach, recently implemented in *BEAST 1.10*, to convert branches of the large-scale phylogenetic trees from units of substitutions per site to time. This model takes the place of the nucleotide substitution model in a traditional Bayesian molecular clock dating analysis. Briefly, each branch of a maximum-likelihood tree is first scaled to represent the number of substitutions that occurred along that branch. Polytomies are resolved by inserting branches of length 0 substitutions. The likelihood of a branch b_i of length s_i substitutions is defined by a Poisson distribution with mean $t_i m$ where t_i is the length of the branch in years and m is the clock rate. The log-likelihood of the whole tree is then the sum of the log-likelihoods of each branch, which represents a fixed, strict-clock model and follows a commonly implemented approach for scaling phylogenies into time-calibrated trees (e.g. *50*-*52*).

Each large-scale phylogenetic tree was analysed under a strict clock model, with the clock rate fixed to the median estimate from the preliminary analysis (7.5×10^{-4} substitutions/site/year) and a Laplace root-height prior with mean equal to the median TMRCA estimate of the corresponding subtree in the preliminary analysis and scale equal to the average distance from the median. Trees were sampled using MCMC under the model described above with a Skygrid coalescent tree prior (*48*) using the same grid-points as in the preliminary analysis. A randomly resolved time-calibrated tree estimated in *TreeTime* (*53*) was used as the starting tree. To maintain a mapping between the topology in the estimated time-calibrated tree and the input genetic distance tree, we constrained the topologies such that any tree-move that broke a clade present in the input tree was rejected. The resulting MCMC chain, therefore, only samples different polytomy resolutions and branch durations. This approach allowed us to incorporate uncertainty in the polytomy resolutions and branch durations into our molecular clock analysis.

We ran between 8 and 24 chains for 60 to 100 million MCMC steps for each large-scale phylogeny. Upon completion, we discarded 15 million states as burn-in from each chain. Chains that did not converge or pass the burn-in in less than 15 million states were re-run. Chains were combined and resampled every 100,000 states using custom R-scripts, leaving between 6808 and 17,020 posterior samples of each large-scale phylogenetic tree. Convergence was assessed using *Tracer* (*49*) and the R-package *coda* (*54*).

<u>Identifying transmission lineages</u>

We define a "UK transmission lineage" as two or more UK infection cases that (i) descend from a shared, single importation of the virus into the UK from elsewhere, (ii) are the result of subsequent local transmission within the UK, and (iii) were present in our virus genome sequence dataset. This concept is illustrated in **Figure 2A** and is distinct from a transmission cluster, an epidemiological term commonly referring to a group of cases that occur close to each other in space and time (e.g. in a hospital or care home). Therefore, a large UK transmission lineage may comprise many different individual transmission clusters. Finally, if a UK transmission lineage is exported to another country, any infections occurring outside of the UK

are determined to not belong to the transmission lineage, and any new importations descending from the same lineage will be classified as new UK transmission lineages.

[It is important to note that the "UK transmission lineage" definition employed here is distinct from the lineage/phylotype designations used by other parts of the COG-UK consortium and that are displayed at https://microreact.org/project/cogconsortium. Those latter designations (which have the format "UK…") are defined on the basis of shared sets of mutations, rather than shared descent from an inferred single introduction event.]

We can identify UK transmission lineages in the time-calibrated trees estimated in the previous step as clades of two or more genomes sampled in the UK. The TMRCA of all genome sequences in a UK transmission lineage represents the earliest transmission event in the lineage revealed by the data; however, it does not necessarily represent the first transmission event in the lineage as a whole, nor does it represent the importation date (i.e. the arrival date of the index patient in the UK). The relationship between the TMRCA of a UK transmission lineage in our dataset and the importation date is illustrated in **Figure 2B**. Specifically, if the transmission lineage is well-sampled, then the TMRCA represents the date of the first transmission event in the lineage (TMRCA A in **Fig. 2B**). However, if the transmission lineage is sparsely sampled then the TMRCA may represent a later transmission event (TMRCA B in **Fig. 2B**). The "importation date" of each UK transmission lineage is the date that an infected inbound traveller entered the UK.

We used a two-state asymmetric discrete trait analysis (DTA) model (*55*) implemented in *BEAST 1.10* (*46*) to infer ancestral node locations (UK, non-UK) on empirical distributions of 500 time-calibrated trees sampled from each of the posterior tree distributions estimated above. Additionally, we used a robust counting approach (*56*) to estimate the expected number of location state transitions into and out of the UK. For each large-scale subtree, we combined 2 independent chains, each run for 5 million MCMC steps and sampled every 4500 states. The first 10% of each run was discarded as burn-in, resulting in 2000 trees with estimates of the ancestral location for each internal node. Finally, *TreeAnnotator 1.10* was used to generate maximum clade credibility (MCC) trees for each subtree, where each internal node is assigned a posterior probability of representing a transmission event in the UK.

Transmission lineages were identified by first labelling each node in the MCC trees as UK or non-UK and then initiating a depth-first search from each UK genome in the MCC trees. All nodes with a median age after 23 January 2020 and posterior probability >0.5 of the ancestral location being located in the UK were labelled as UK nodes. The depth-first search is continued until a non-UK node is encountered or there are no nodes left to explore. At the end of the depth-first search, all nodes visited by the search are added to the same UK transmission lineage. If only one tip is visited, the UK genome at the tip is marked as a singleton. This procedure is repeated iteratively until every UK genome in the tree has been assigned a transmission lineage or marked as a singleton. Transmission lineage names start with the dataset used to construct the MCC tree, followed by "_DTA_MCC_", and an arbitrary number e.g. "B.1.1_DTA_MCC_42." For the 8 largest transmission lineages we simplify the name as follows:

1. B.1.1_DTA_MCC_47 = DTA_47
2. B_DTA_MCC_1 = DTA_1

3. B_DTA_MCC_13 = DTA_13
4. B.1.pruned_DTA_MCC_17 = DTA_17
5. B.1.pruned_DTA_MCC_62 = DTA_62
6. B.1.pruned_DTA_MCC_234 = DTA_234
7. B_DTA_MCC_172 = DTA_172
8. B.1.pruned_DTA_MCC_290 = DTA_290

The above procedure was repeated on each of the 2000 posterior trees, for each subtree, from the DTA analyses described above to examine statistical uncertainty in the number, size and duration of UK transmission lineages and their TMRCA distribution. Transmission lineages identified on each posterior tree follow the same naming convention, but without "MCC" in the name.

Our methodology is likely to underestimate the true number of transmission lineages and singletons. Since only a small fraction of UK infections have been sequenced (**Fig. 2A**), many lineages will have gone undetected. Furthermore, the power to detect a transmission lineage in our sparsely-sampled dataset is dependent on its size (i.e. the frequency of a lineage being sampled from a small random sample of infections), making it more likely for larger lineages to be detected. The low sampling fraction means that some singletons detected in our dataset likely belong to observed and unobserved UK transmission lineages. Nonetheless, the true number of singletons (importations not resulting in onward transmission) is likely to be significantly more than our estimate, because their small size makes them difficult to detect with a low sampling fraction. Finally, under-sampling of genomes from other countries could result in mistaken aggregation of separate importations, reducing the number of detected lineages. This mistaken aggregation will result in larger, older lineages being estimated. This was the motivation for placing an age limit on UK nodes in the tree. We chose 23 January 2020 as the oldest possible date for a transmission event in the UK as this represents the date that the first patient who tested positive for SARS-CoV-2 in the UK entered the country (*57*) (tested positive on 30 January 2020). Although older importations into the UK could in theory be possible, if they had resulted in large autochthonous outbreaks we would have observed this in both epidemiological and genomic data.

We estimate a median of 2968 (95% HPD 2829-3103) non-UK to UK state transitions and an additional 1468 (95% HPD 1362-1566) UK to non-UK state transitions **(Fig. S1, Table S1)** using the robust counting approach (*56*). The former slightly exceeds the sum of transmission lineages and singletons as identified on the MCC trees (=2918) and across the 2000 posterior trees (median=2829, 95% HPD=2773-3048; **Table S1**). This result is expected, since multiple location state changes along long branches contribute to the total number of state transitions, but do not add to the total number of UK transmission lineages or singletons. The largest number of location state transitions occur on the B.1.1 phylogeny, with the fewest occurring on lineage A, which are the largest and smallest of the subtrees, respectively. Proportional to the number of tips, fewer state changes are inferred on the two B.1 phylogenies than other subtrees, while the number of UK to non-UK transitions on the B phylogeny exceeds that inferred on other lineages. We caution that UK to non-UK transitions are likely to be underestimated because of under-sampling in other countries and differences in the proportion of infections sequenced between countries.

The transmission lineage size distribution from the MCC trees falls within the HPD interval taken across the 2000 posterior trees (**Fig. 2C**). Although the sizes of the largest transmission lineages vary substantially across posterior trees, the cumulative size distributions are similar across all trees (**Fig. 2C**, inset). Similarly, the transmission lineage duration distribution on the MCC trees falls within the variation of the HPD interval taken across the 2000 posterior trees (**Fig. S6**).

We used the Jaccard index to compare the classification of UK genome sequences into transmission lineages and singletons between posterior trees and the MCC trees. **Figure S13A** shows the mean, median and 95% HPD interval of the Jaccard index for each posterior tree compared to the 1999 other posterior trees, across all subtrees. While most Jaccard indices are between 0.7 and 0.8, there is a noteworthy minority of trees with mean Jaccard indices <0.6 (n=100). Comparing the 2000 posterior trees to the classification on the MCC trees (**Fig. S13B**), results in a similar distribution of Jaccard indices, with most indices between 0.7 and 0.8 and minorities below 0.6 and above 0.8 (n=68, n=170 respectively).

We undertook a similar analysis of the sensitivity to phylogenetic uncertainty of the distribution of UK transmission lineage TMRCAs. We computed the median and 95% HPD interval of the number of transmission lineage TMRCAs on each date across the 2000 sampled posterior trees. **Figure S14** shows that the TMRCA distribution computed from the MCC trees falls within the comparatively narrow HPD limits, and oscillates around the median estimate for each date.

<u>UK epidemiological data</u>
The number of reported COVID-19 cases in the UK, by specimen date, were downloaded from https://coronavirus.data.gov.uk/cases (date accessed: 1 September 2020). The number of reported COVID-19 cases for each Upper Tier Local Authority (UTLA) in England, Local Health Board (LHB) in Wales and regional NHS Board in Scotland, by specimen date, were downloaded from https://coronavirus.data.gov.uk/downloads/csv/coronavirus-cases_latest.csv, http://www2.nphs.wales.nhs.uk:8080/CommunitySurveillanceDocs.nsf (file: *"Rapid COVID-19 surveillance data.xlsx"*) and https://github.com/DataScienceScotland/COVID-19-Management-Information (file: *"COVID19 - Daily Management Information - Scottish Health Boards - Cumulative cases.csv"*), respectively (date accessed: 15 October 2020).

To enable comparison of case and sequence data, locations used to report case data were combined to correspond to those used for sequence data and vice-versa (see the *Geographical metadata* section). Northern Ireland was not included due to inconsistencies between the locations used for case and sequence data reporting that could not be easily resolved.

<u>Global deaths due to COVID-19</u>
The cumulative number of daily COVID-19 deaths for each country were downloaded from the JHU CSSE COVID-19 Database (date accessed: 19 August 2020) (*58*). We removed data pertaining to cruise ships, and aggregated data to the country level where data were reported for subnational divisions (e.g. Australia). For countries with overseas territories included in the dataset (e.g. United Kingdom), we excluded the cumulative death counts in those overseas territories. For each country we computed a time series of the daily number of deaths by taking the difference in the cumulative number on consecutive days. When this difference was negative,

for example when corrections in the cumulative number were not propagated backwards, we set the value to zero. A relevant outlier in these time series is the addition of 1290 deaths in China on 17 April 2020, while on the days before and after no deaths were recorded. To account for these deaths, we uniformly distributed these deaths over the previous 85 days described by the epidemiological data.

Population data

Country population size estimates were downloaded from the UN Department of Economic and Social Affairs website (https://population.un.org/wpp/Download/Standard/Population/), using the *Medium* fertility projection for 2020 (*59*).

Travel and mobility data

To investigate temporal trends in SARS-CoV-2 importation intensity we sought information on the number of travellers entering the UK from each other country for the period from 1 January to 30 April 2020. Incoming travellers comprised both British nationals and resident and visiting citizens of other countries. Estimates were obtained by combining multiple data sources. First, the UK Home Office has provided statistics that describe the number of inbound travellers arriving in the UK by air on each day during this period (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/887655/statistics-relating-to-covid-19-and-the-immigration-system-tables-may-2020-arrivals.ods). This data set provides the daily number of incoming air passengers but not their source country. Second, we obtained the number of tickets sold for inbound flight journeys to the UK along with their origin location from the IATA (for passengers that transfer, the source location is the country from where the whole journey started). We used these numbers to calculate the percentage of arrivals from each country on a monthly basis from January to April 2020. We multiplied the monthly distribution of source destination by the total number of air passenger arrivals in the UK each day to estimate the number of arrivals from each country. Third, we augmented the above air passenger numbers with estimated numbers of incoming travellers arriving per day by short-sea ferry and through the Channel Tunnel (French: *Le tunnel sous la Manche*). Numbers of short-sea ferry passengers from France, Netherlands and the Republic of Ireland were estimated from monthly statistics obtained from the UK Department of Transport (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/908445/spas0107.ods). Within that data set, values are provided for the Republic of Ireland and for "Other EU countries". The latter total was broken down by country using data from 2019 showing that 72.7% of UK short-sea journeys are with France, 13.6% with the Republic of Ireland, 10.4% with the Netherlands, and 3.3% with other countries (https://www.gov.uk/government/statistics/sea-passenger-statistics-2019-short-sea-routes). Eurotunnel Shuttle vehicle movements from France were obtained from publicly available monthly records (https://www.eurotunnelfreight.com/uk/2020/02/shuttle-traffic-for-january-2020). In the absence of other information we assumed (i) inbound and outbound vehicle movements via the Eurotunnel Shuttle services were equally frequent and (ii) one passenger per truck and 1.5 passengers per passenger vehicle. Inbound Eurostar rail passenger numbers from France and Belgium were estimated from available data and adjusted as far as possible for post-pandemic reduction in travel. Specifically, ~2m passengers travelled by Eurostar in the first quarter of 2020 (https://www.breakingtravelnews.com/news/article/eurostar-passenger-count-

slips-by-a-fifth-in-early-2020). Monthly Eurostar passenger numbers were then calculated by assuming (i) inbound and outbound journeys were equally frequent, (ii) two thirds of inbound Eurostar journeys originated in France and one third in Belgium, in approximate proportion to the ratio of services, and (iii) the proportional decrease in Eurostar travel volumes during March and April 2020 was equal to that observed for vehicle movements via the Eurotunnel Shuttle. Our estimates do not incorporate estimates of movements across the land border between the UK and the Republic of Ireland. This is unlikely to be problematic as the numbers of infections in the Republic of Ireland was relatively low compared to other potential source countries during the time period of interest.

Epidemiological model

We sought an estimate of the number of individuals in each source country who are (i) infected with SARS-CoV-2 and (ii) able to travel to the UK and initiate a transmission chain. In what follows we refer to these individuals as the "potential initiators of a transmission lineage" (PITL). We conservatively assumed that symptomatic individuals cannot initiate a transmission chain in the UK, either through being prevented from travelling or perfect isolation on arrival. Thus, our estimates of daily SARS-CoV-2 prevalence includes only pre-symptomatic and asymptomatic individuals. Asymptomatic individuals are counted among the PITL as those capable of initiating a transmission lineage at any time while they are still infectious. **Figure S16** illustrates the ways in which individuals are counted towards the daily PITL and their potential disease outcomes.

We estimated the daily number of PITL by back-extrapolating the time series of daily numbers of deaths due to COVID-19 in each source country. COVID-19 deaths were used instead of confirmed cases, as we are primarily interested in temporal dynamics rather than absolute values, and death counts are believed to be less sensitive to changes in case definition, reporting delays and differences in the level of surveillance among countries and regions. Estimates of the latent period (infection to becoming infectious), incubation period (infection to onset of symptoms), the infectious duration, and the time between symptom onset and death (in fatal cases) were used to estimate the number of infected individuals who would go on to die from COVID-19, in each stage of the disease, on each day (**Fig. S16**). We then estimated the total number of infected individuals on each day by multiplying with the reciprocal of the infection fatality rate (IFR).

Estimates of the periods defined above were taken from peer-reviewed sources. Specifically, we assumed that the time from acquiring an infection to becoming infectious is 3 days (*60*) and the time to symptom onset 5 days (2 days after becoming infectious) (*61*). The infectious period for patients who recover from the disease was assumed to end 5 days after symptom onset (*60*) while those who die from the disease are assumed to do so 18 days after the onset of symptoms (*62*). Given the large numbers of deaths we expect that variation in these timings among individuals will be averaged out and is not considered. We further assumed an asymptomatic proportion of 31% (*63*) and an IFR of 1%, which is broadly consistent with those found in the literature for China, France, and passengers aboard the Diamond Princess (62, *64, 65*). These values correspond to our study period, the spring epidemic of COVID-19; more recent estimates of IFR may vary due to changing treatment regimes and other factors. To examine the sensitivity of our results to the asymptomatic proportion we re-ran our analysis with proportions of 0.18 and 0.78

(the range of published estimates; *66, 67*), and found that our results were robust over this range (data not shown).

We did not account for changing levels of infectivity among individuals over the course of their infection. Using the time series of deaths extracted from the JHU CSSE COVID-19 Database (*58*), as described above, we obtained estimates of the daily number of PITL in 183 countries from 31 December 2019 to 26 July 2020.

Estimated importation intensity

The daily "estimated importation intensity" (EII) of a country is defined as the product of the proportion of individuals in that country who make up the PITL (as described above) on each day, and the number of individuals who travelled from that country to the UK on that day. The former is estimated by dividing our estimate of the total number of individuals who could potentially initiate a lineage (for each day) by the total population of the country (see the *Epidemiological model* section). The latter corresponds to the total number of arrivals by air, ferry, and rail on that day (see the *Travel and mobility data* section). To assist in the subsequent use of the EII, we aggregated all countries with low PITL estimates into a single "other" category. The aggregated countries are those that comprised less than 1% of the cumulative total number of cases as of 1 May 2020 (excluding the UK). This left 53 primary source locations. Maximum EII (**Fig. S17**) was highest for Spain, (which experienced a large, early epidemic that peaked before inbound passenger numbers declined), followed by France (whose later epidemic peak coincided with high but declining international travel).

Importation lag model

We modelled the TMRCA of an observed transmission lineage (the data observation) as the arrival date of the index patient (of that transmission cluster) in the UK, G, plus a lag time, L, until the first transmission event in the lineage revealed by the data. Given the probability that an importation occurs on day g, $f_G(g)$, and the probability of a lag time of j days, $f_L(j)$, the probability of a TMRCA occurring on day k is v_k, defined by

$$\hat{v}_k = \sum_g f_G(g) f_l(k-g)$$

with $v = \hat{v} / |\hat{v}|$. TMRCAs and importation dates are assumed to be independent, so the likelihood for all transmission lineages is the product of the corresponding v_k for each lineage.

This model does not account for incomplete sampling of patients from UK transmission lineages. It is likely that the TMRCA of a small transmission lineage is more recent than the first transmission event after the importation and this issue is potentially further exacerbated by non-random sampling of genome sequences from patients in the lineage (*68*). We therefore expect shorter lag times for bigger transmission lineages. To account for this size-dependence, we model the average importation lag as a function of lineage size. The functional form of this is given by the equation $\alpha + \beta / n$, where α corresponds to the minimal average lag time expected under complete sampling of the lineage and β accounts for the increase in lag time as a smaller proportion of sequences are included in the lineage.

We applied this model to the TMRCA estimates of individual transmission lineages and their sizes as obtained from the MCC trees (see the *Identifying transmission lineages* section). Values for α and β were found by numerically optimising the likelihood function using random draws from an exponential distribution as initial parameter values. The optimisation procedure was repeated several times to ensure that the algorithm did not become stuck in a local optimum. We further tested whether lineage size affects the importation lag through a likelihood ratio test (LRT) comparing the above model to a nested model without size dependence ($\beta = 0$), and found that the size-dependent model is preferred ($\chi_1^2 = 137.22$, $p < 0.001$). The maximum likelihood estimates for α and β are 0.72 and 28.91 (**Fig. S18**), respectively.

Although we assume a constant IFR in the epidemiological model, it is likely that the IFR has varied both through time (due to changes in treatment) and among locations. However, notable improvements in COVID-19 treatment were mostly implemented after our study period and the countries that contributed >90% of estimated imported cases to the UK (**Fig. S20**) are Western European nations with similar medical systems and mechanisms of reporting COVID-19 associated deaths. Crucially, the dependence of EII on the number of inbound travellers from, and the number of cases in, each country (both of which vary rapidly over orders of magnitude) means that likely variation in IFR has comparatively little numerical effect.

Travel advice in the UK

The travel advice issued by the Foreign and Commonwealth Office (FCO) of the United Kingdom pertaining to countries and regions affected by COVID-19 was primarily made available through their website (*FCO Travel advice: coronavirus (COVID-19)* at https://www.gov.uk/guidance/travel-advice-novel-coronavirus). The number of COVID-19 cases in the UK was available via the government website (*Coronavirus cases in the UK* at https://www.gov.uk/guidance/coronavirus-covid-19-information-for-the-public). Travel advice was also echoed by various news outlets and other information platforms, such as the Public Health Scotland/NHS Scotland *Fit for Travel* website (https://www.fitfortravel.nhs.uk/). We collected this information by mining archived FCO sites, manually retrieving HTML files corresponding to updates to the URLs provided above and available at the Internet Archive (https://archive.org/). Files were obtained and examined for all dates when changes to the URL were published (18 updates were published in total between 4 February and 23 May 2020). Furthermore, we compared this advice with the *Fit for Travel* online resource, collected through a similar approach. Where information was insufficient or unclear, we complemented it with data from news outlets to clarify travel advice, which was the case before February 4, when there was no official travel advice (only notifications for novel coronavirus). We collated all the travel advice information into a single standardised table containing types of advice, dates of implementation and countries or geographic regions covered by the advice. The types of advice included both suggestions against specific types of travel versus all but non-essential travel and the recommended period of self-isolation upon return from specific destinations. All of the changes in travel advice were between February 6 and March 23, when specific self-isolation recommendations applied to the general population and not just returning travellers. A summary of the main changes in the UK travel advice across time (in particular, dates when advice for new countries were issued) is presented in **Table S5** and the complete lookup table is provided as a separate file (**Data S5)**.

List of COG-UK consortium members

Funding acquisition, leadership, supervision, metadata curation, project administration, samples, logistics, Sequencing, analysis, and Software and analysis tools:
Thomas R Connor [33, 34], and Nicholas J Loman [15].

Leadership, supervision, sequencing, analysis, funding acquisition, metadata curation, project administration, samples, logistics, and visualisation:
Samuel C Robson [68].

Leadership, supervision, project administration, visualisation, samples, logistics, metadata curation and software and analysis tools:
Tanya Golubchik [27].

Leadership, supervision, metadata curation, project administration, samples, logistics sequencing and analysis:
M. Estee Torok [8, 10].

Project administration, metadata curation, samples, logistics, sequencing, analysis, and software and analysis tools:
William L Hamilton [8, 10].

Leadership, supervision, samples logistics, project administration, funding acquisition sequencing and analysis:
David Bonsall [27].

Leadership and supervision, sequencing, analysis, funding acquisition, visualisation and software and analysis tools:
Ali R Awan [74].

Leadership and supervision, funding acquisition, sequencing, analysis, metadata curation, samples and logistics:
Sally Corden [33].

Leadership supervision, sequencing analysis, samples, logistics, and metadata curation:
Ian Goodfellow [11].

Leadership, supervision, sequencing, analysis, samples, logistics, and Project administration:
Darren L Smith [60, 61].

Project administration, metadata curation, samples, logistics, sequencing and analysis:
Martin D Curran [14], and Surendra Parmar [14].

Samples, logistics, metadata curation, project administration sequencing and analysis:
James G Shepherd [21].

Sequencing, analysis, project administration, metadata curation and software and analysis tools:
Matthew D Parker [38].

Leadership, supervision, funding acquisition, samples, logistics, and metadata curation:
Catherine Moore [33].

Leadership, supervision, metadata curation, samples, logistics, sequencing and analysis:
Derek J Fairley[6, 88], Matthew W Loose [54], and Joanne Watkins [33].

Metadata curation, sequencing, analysis, leadership, supervision and software and analysis tools:
Matthew Bull [33], and Sam Nicholls [15].

Leadership, supervision, visualisation, sequencing, analysis and software and analysis tools:
David M Aanensen [1, 30].

Sequencing, analysis, samples, logistics, metadata curation, and visualisation:
Sharon Glaysher [70].

Metadata curation, sequencing, analysis, visualisation, software and analysis tools:
Matthew Bashton [60], and Nicole Pacchiarini [33].

Sequencing, analysis, visualisation, metadata curation, and software and analysis tools:
Anthony P Underwood [1, 30].

Funding acquisition, leadership, supervision and project administration:
Thushan I de Silva [38], and Dennis Wang [38].

Project administration, samples, logistics, leadership and supervision:
Monique Andersson [28], Anoop J Chauhan [70], Mariateresa de Cesare [26], Catherine Ludden [1,3], and Tabitha W Mahungu [91].

Sequencing, analysis, project administration and metadata curation:
Rebecca Dewar [20], and Martin P McHugh [20].

Samples, logistics, metadata curation and project administration:
Natasha G Jesudason [21], Kathy K Li MBBCh [21], Rajiv N Shah [21], and Yusri Taha [66].

Leadership, supervision, funding acquisition and metadata curation:
Kate E Templeton [20].

Leadership, supervision, funding acquisition, sequencing and analysis:
Simon Cottrell [33], Justin O'Grady [51], Andrew Rambaut [19], and Colin P Smith[93].

Leadership, supervision, metadata curation, sequencing and analysis:
Matthew T.G. Holden [87], and Emma C Thomson [21].

Leadership, supervision, samples, logistics and metadata curation:
Samuel Moses [81, 82].

Sequencing, analysis, leadership, supervision, samples and logistics:
Meera Chand [7], Chrystala Constantinidou [71], Alistair C Darby [46], Julian A Hiscox [46], Steve Paterson [46], and Meera Unnikrishnan [71].

Sequencing, analysis, leadership and supervision and software and analysis tools:
Andrew J Page [51], and Erik M Volz [96].

Samples, logistics, sequencing, analysis and metadata curation:
Charlotte J Houldcroft [8], Aminu S Jahun [11], James P McKenna [88], Luke W Meredith [11], Andrew Nelson [61], Sarojini Pandey [72], and Gregory R Young [60].

Sequencing, analysis, metadata curation, and software and analysis tools:
Anna Price [34], Sara Rey [33], Sunando Roy [41], Ben Temperton[49], and Matthew Wyles [38].

Sequencing, analysis, metadata curation and visualisation:
Stefan Rooke[19], and Sharif Shaaban [87].

Visualisation, sequencing, analysis and software and analysis tools:
Helen Adams [35], Yann Bourgeois [69], Katie F Loveson [68], Áine O'Toole [19], and Richard Stark [71].

Project administration, leadership and supervision:
Ewan M Harrison [1, 3], David Heyburn [33], and Sharon J Peacock [2, 3].

Project administration and funding acquisition:
David Buck [26], and Michaela John[36].

Sequencing, analysis and project administration:
Dorota Jamrozy [1], and Joshua Quick [15].

Samples, logistics, and project administration:
Rahul Batra [78], Katherine L Bellis [1, 3], Beth Blane [3], Sophia T Girgis [3], Angie Green [26], Anita Justice [28], Mark Kristiansen [41], and Rachel J Williams [41].

Project administration, software and analysis tools:
Radoslaw Poplawski[15].

Project administration and visualisation:
Garry P Scarlett [69].

Leadership, supervision, and funding acquisition:
John A Todd [26], Christophe Fraser [27], Judith Breuer [40,41], Sergi Castellano [41], Stephen L Michell [49], Dimitris Gramatopoulos [73], and Jonathan Edgeworth [78].

Leadership, supervision and metadata curation:
Gemma L Kay [51].

Leadership, supervision, sequencing and analysis:
Ana da Silva Filipe [21], Aaron R Jeffries [49], Sascha Ott [71], Oliver Pybus [24], David L Robertson [21], David A Simpson [6], and Chris Williams [33].

Samples, logistics, leadership and supervision:
Cressida Auckland [50], John Boyes [83], Samir Dervisevic [52], Sian Ellard [49, 50], Sonia Goncalves[1], Emma J Meader [51], Peter Muir [2], Husam Osman [95], Reenesh Prakash [52], Venkat Sivaprakasam [18], and Ian B Vipond [2].

Leadership, supervision and visualization:
Jane AH Masoli [49, 50].

Sequencing, analysis and metadata curation:
Nabil-Fareed Alikhan [51], Matthew Carlile [54], Noel Craine [33], Sam T Haldenby [46], Nadine Holmes [54], Ronan A Lyons [37], Christopher Moore [54], Malorie Perry [33], Ben Warne [80], and Thomas Williams [19].

Samples, logistics and metadata curation:
Lisa Berry [72], Andrew Bosworth [95], Julianne Rose Brown [40], Sharon Campbell [67], Anna Casey [17], Gemma Clark [56], Jennifer Collins [66], Alison Cox [43, 44], Thomas Davis [84], Gary Eltringham [66], Cariad Evans [38, 39], Clive Graham [64], Fenella Halstead [18], Kathryn Ann Harris [40], Christopher Holmes [58], Stephanie Hutchings [2], Miren Iturriza-Gomara [46], Kate Johnson [38, 39], Katie Jones [72], Alexander J Keeley [38], Bridget A Knight [49, 50], Cherian Koshy[90], Steven Liggett [63], Hannah Lowe [81] , Anita O Lucaci [46] , Jessica Lynch [25, 29], Patrick C McClure [55], Nathan Moore [31] , Matilde Mori [25, 29, 32], David G Partridge [38, 39], Pinglawathee Madona [43, 44], Hannah M Pymont [2], Paul Anthony Randell [43, 44], Mohammad Raza [38, 39], Felicity Ryan [81], Robert Shaw [28], Tim J Sloan [57], and Emma Swindells [65].

Sequencing, analysis, Samples and logistics:
Alexander Adams [33], Hibo Asad [33], Alec Birchley [33], Tony Thomas Brooks [41], Giselda Bucca [93], Ethan Butcher [70], Sarah L Caddy [13], Laura G Caller [2, 3, 12], Yasmin Chaudhry [11], Jason Coombes [33], Michelle Cronin [33], Patricia L Dyal [41], Johnathan M Evans [33], Laia Fina [33], Bree Gatica-Wilcox [33], Iliana Georgana [11], Lauren Gilbert [33], Lee Graham [33], Danielle C Groves [38], Grant Hall [11], Ember Hilvers [33], Myra Hosmillo [11], Hannah Jones [33], Sophie Jones [33], Fahad A Khokhar [13] , Sara Kumziene-Summerhayes [33], George MacIntyre-Cockett [26], Rocio T Martinez Nunez [94], Caoimhe McKerr [33], Claire McMurray [15], Richard Myers [7], Yasmin Nicole Panchbhaya [41], Malte L Pinckert [11], Amy Plimmer [33], Joanne Stockton [15], Sarah Taylor [33], Alicia Thornton [7], Amy Trebes [26], Alexander J Trotter [51], Helena Jane Tutill [41], Charlotte A Williams [41] , Anna Yakovleva [11], and Wen C Yew [62].

Sequencing, analysis and software and analysis tools:
Mohammad T Alam [71], Laura Baxter [71], Olivia Boyd [96], Fabricia F. Nascimento [96], Timothy M

Freeman [38], Lily Geidelberg [96], Joseph Hughes [21], David Jorgensen [96], Benjamin B Lindsey [38], Richard J Orton [21], Manon Ragonnet-Cronin [96], Joel Southgate [33, 34], and Sreenu Vattipally [21].

Samples, logistics and software and analysis tools:
Igor Starinskij [23].

Visualisation and software and analysis tools:
Joshua B Singer [21], Khalil Abudahab [1, 30], Leonardo de Oliveira Martins [51], Thanh Le-Viet [51], Mirko Menegazzo [30], Ben EW Taylor [1, 30], and Corin A Yeats [30].

Project Administration:
Sophie Palmer [3], Carol M Churcher [3], Alisha Davies [33], Elen De Lacy [33], Fatima Downing [33], Sue Edwards [33], Nikki Smith [38], Francesc Coll [97], Nazreen F Hadjirin [3], and Frances Bolt [44, 45].

Leadership and supervision:
Alex Alderton [1], Matt Berriman [1], Ian G Charles [51], Nicholas Cortes [31], Tanya Curran [88], John Danesh[1], Sahar Eldirdiri [84], Ngozi Elumogo [52], Andrew Hattersley [49, 50], Alison Holmes [44, 45], Robin Howe [33], Rachel Jones [33], Anita Kenyon [84], Robert A Kingsley [51], Dominic Kwiatkowski [1,9], Cordelia Langford [1], Jenifer Mason [48], Alison E Mather [51], Lizzie Meadows [51], Sian Morgan [36], James Price [44, 45], Trevor I Robinson [48], Giri Shankar [33], John Wain [51], and Mark A Webber [51].

Metadata curation:
Declan T Bradley [5, 6], Michael R Chapman [1, 3, 4], Derrick Crooke [28], David Eyre [28], Martyn Guest [34], Huw Gulliver [34], Sarah Hoosdally [28], Christine Kitchen [34], Ian Merrick [34], Siddharth Mookerjee [44, 45], Robert Munn [34] , Timothy Peto [28], Will Potter [52], Dheeraj K Sethi [52], Wendy Smith [56], Luke B Snell [75, 94], Rachael Stanley [52], Claire Stuart [52], and Elizabeth Wastenge[20].

Sequencing and analysis:
Erwan Acheson [6], Safiah Afifi [36], Elias Allara [2, 3], Roberto Amato [1], Adrienn Angyal [38], Elihu Aranday-Cortes [21], Cristina Ariani [1], Jordan Ashworth [19], Stephen Attwood [24], Alp Aydin [51], David J Baker [51], Carlos E Balcazar [19], Angela Beckett [68] Robert Beer [36], Gilberto Betancor [76], Emma Betteridge [1], David Bibby [7], Daniel Bradshaw [7], Catherine Bresner [34], Hannah E Bridgewater [71], Alice Broos [21], Rebecca Brown [38], Paul E Brown [71], Kirstyn Brunker [22], Stephen N Carmichael [21], Jeffrey K. J. Cheng [71], Rachel Colquhoun [19], Gavin Dabrera [7], Johnny Debebe [54], Eleanor Drury [1], Louis du Plessis [24], Richard Eccles [46], Nicholas Ellaby [7], Audrey Farbos [49], Ben Farr [1], Jacqueline Findlay [41], Chloe L Fisher [74], Leysa Marie Forrest [41], Sarah Francois [24], Lucy R. Frost [71], William Fuller [34], Eileen Gallagher [7], Michael D Gallagher [19], Matthew Gemmell [46], Rachel AJ Gilroy [51], Scott Goodwin [1], Luke R Green [38], Richard Gregory [46], Natalie Groves [7], James W Harrison [49], Hassan Hartman [7], Andrew R Hesketh [93],Verity Hill [19], Jonathan Hubb [7], Margaret Hughes [46] , David K Jackson [1], Ben Jackson [19], Keith James [1], Natasha Johnson [21], Ian Johnston [1], Jon-Paul Keatley [1], Moritz Kraemer [24], Angie Lackenby [7], Mara Lawniczak [1], David Lee [7], Rich Livett [1], Stephanie Lo [1], Daniel Mair [21], Joshua Maksimovic [36], Nikos Manesis [7], Robin Manley [49], Carmen Manso [7], Angela Marchbank [34], Inigo Martincorena [1], Tamyo Mbisa [7], Kathryn McCluggage [36], JT McCrone [19], Shahjahan Miah [7], Michelle L Michelsen [49], Mari Morgan [33], Gaia Nebbia [78], Charlotte Nelson [46], Jenna Nichols [21], Paola Niola [41], Kyriaki Nomikou [21], Steve Palmer [1], Naomi Park [1], Yasmin A Parr [1], Paul J Parsons [38], Vineet Patel [7],

Minal Patel [1], Clare Pearson [2, 1], Steven Platt [7], Christoph Puethe [1], Mike Quail [1], Jayna Raghwani [24], Lucille Rainbow [46], Shavanthi Rajatileka [1], Mary Ramsay [7], Paola C Resende Silva [41, 42], Steven Rudder [51], Chris Ruis [3], Christine M Sambles [49], Fei Sang [54], Ulf Schaefer [7], Emily Scher [19], Carol Scott [1], Lesley Shirley [1], Adrian W Signell [76], John Sillitoe [1], Christen Smith [1], Katherine L Smollett [21], Karla Spellman [36] , Thomas D Stanton [19], David J Studholme [49], Grace Taylor-Joyce [71], Ana P Tedim [51], Thomas Thompson [6], Nicholas M Thomson [51], Scott Thurston [1] , Lily Tong [21], Gerry Tonkin-Hill [1], Rachel M Tucker [38], Edith E Vamos [4], Tetyana Vasylyeva [24], Joanna Warwick-Dugdale [49], Danni Weldon [1], Mark Whitehead [46], David Williams [7], Kathleen A Williamson [19], Harry D Wilson [76], Trudy Workman [34], Muhammad Yasir [51], Xiaoyu Yu [19], and Alex Zarebski [24].

Samples and logistics:
Evelien M Adriaenssens [51], Shazaad S Y Ahmad [2, 47], Adela Alcolea-Medina [59, 77], John Allan [60], Patawee Asamaphan [21], Laura Atkinson [40], Paul Baker [63], Jonathan Ball [55], Edward Barton[64], Mathew A Beale[1], Charlotte Beaver[1], Andrew Beggs [16], Andrew Bell [51], Duncan J Berger [1], Louise Berry. [56], Claire M Bewshea [49], Kelly Bicknell [70], Paul Bird [58], Chloe Bishop [7], Tim Boswell [56], Cassie Breen [48], Sarah K Buddenborg [1], Shirelle Burton-Fanning [66], Vicki Chalker [7], Joseph G Chappell [55], Themoula Charalampous [78, 94], Claire Cormie [3], Nick Cortes [29, 25], Lindsay J Coupland [52], Angela Cowell [48], Rose K Davidson [53], Joana Dias [3], Maria Diaz [51], Thomas Dibling [1], Matthew J Dorman [1], Nichola Duckworth [57], Scott Elliott [70], Sarah Essex [63], Karlie Fallon [58], Theresa Feltwell [8], Vicki M Fleming [56], Sally Forrest [3], Luke Foulser [1], Maria V Garcia-Casado [1], Artemis Gavriil [41], Ryan P George [47], Laura Gifford [33], Harmeet K Gill [3], Jane Greenaway [65], Luke Griffith [53], Ana Victoria Gutierrez [51], Antony D Hale [85], Tanzina Haque [91], Katherine L Harper [85], Ian Harrison [7], Judith Heaney [89], Thomas Helmer [58], Ellen E Higginson [3], Richard Hopes [2], Hannah C Howson-Wells [56], Adam D Hunter [1], Robert Impey [70], Dianne Irish-Tavares [91], David A Jackson [1], Kathryn A Jackson [46], Amelia Joseph [56], Leanne Kane [1], Sally Kay [1], Leanne M Kermack [3], Manjinder Khakh [56], Stephen P Kidd [29, 25, 31], Anastasia Kolyva [51], Jack CD Lee [40], Laura Letchford [1], Nick Levene [79], Lisa J Levett [89], Michelle M Lister [56], Allyson Lloyd [70], Joshua Loh [60], Louissa R Macfarlane-Smith [85], Nicholas W Machin [2, 47], Mailis Maes [3], Samantha McGuigan [1], Liz McMinn [1], Lamia Mestek-Boukhibar [41], Zoltan Molnar [6], Lynn Monaghan [79], Catrin Moore [27], Plamena Naydenova [3], Alexandra S Neaverson [1], Rachel Nelson [1], Marc O Niebel [21], Elaine O'Toole [48], Debra Padgett [64], Gaurang Patel [1], Brendan AI Payne [66], Liam Prestwood [1], Veena Raviprakash [67], Nicola Reynolds [86], Alex Richter [16], Esther Robinson [95], Hazel A Rogers [1], Aileen Rowan [96], Garren Scott [64], Divya Shah [40], Nicola Sheriff [67], Graciela Sluga, Emily Souster [1], Michael Spencer-Chapman [1], Sushmita Sridhar [1, 3], Tracey Swingler [53], Julian Tang [58], Graham P Taylor [96], Theocharis Tsoleridis [55], Lance Turtle [46], Sarah Walsh [57], Michelle Wantoch [86], Joanne Watts [48], Sheila Waugh [66], Sam Weeks [41], Rebecca Williams [31], Iona Willingham [56], Emma L Wise [25, 29, 31], Victoria Wright [54], Sarah Wyllie [70], and Jamie Young [3].

Software and analysis tools:
Amy Gaskin [33], Will Rowe [15], and Igor Siveroni [96].

Visualisation:
Robert Johnson [96].

1 Wellcome Sanger Institute, **2** Public Health England, **3** University of Cambridge, **4** Health Data Research UK, Cambridge, **5** Public Health Agency, Northern Ireland ,**6** Queen's University Belfast **7** Public Health England Colindale, **8** Department of Medicine, University of Cambridge, **9** University of Oxford, **10** Departments of Infectious Diseases and Microbiology, Cambridge University Hospitals NHS Foundation Trust; Cambridge, UK, **11** Division of Virology, Department of Pathology, University of Cambridge, **12** The Francis Crick Institute, **13** Cambridge Institute for Therapeutic Immunology and Infectious Disease, Department of Medicine, **14** Public Health England, Clinical Microbiology and Public Health Laboratory, Cambridge, UK, **15** Institute of Microbiology and Infection, University of Birmingham, **16** University of Birmingham, **17** Queen Elizabeth Hospital, **18** Heartlands Hospital, **19** University of Edinburgh, **20** NHS Lothian, **21** MRC-University of Glasgow Centre for Virus Research, **22** Institute of Biodiversity, Animal Health & Comparative Medicine, University of Glasgow, **23** West of Scotland Specialist Virology Centre, **24** Dept Zoology, University of Oxford, **25** University of Surrey, **26** Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, **27** Big Data Institute, Nuffield Department of Medicine, University of Oxford, **28** Oxford University Hospitals NHS Foundation Trust, **29** Basingstoke Hospital, **30** Centre for Genomic Pathogen Surveillance, University of Oxford, **31** Hampshire Hospitals NHS Foundation Trust, **32** University of Southampton, **33** Public Health Wales NHS Trust, **34** Cardiff University, **35** Betsi Cadwaladr University Health Board, **36** Cardiff and Vale University Health Board, **37** Swansea University, **38** University of Sheffield, **39** Sheffield Teaching Hospitals, **40** Great Ormond Street NHS Foundation Trust, **41** University College London, **42** Oswaldo Cruz Institute, Rio de Janeiro **43** North West London Pathology, **44** Imperial College Healthcare NHS Trust, **45** NIHR Health Protection Research Unit in HCAI and AMR, Imperial College London, **46** University of Liverpool, **47** Manchester University NHS Foundation Trust, **48** Liverpool Clinical Laboratories, **49** University of Exeter, **50** Royal Devon and Exeter NHS Foundation Trust, **51** Quadram Institute Bioscience, University of East Anglia, **52** Norfolk and Norwich University Hospital, **53** University of East Anglia, **54** Deep Seq, School of Life Sciences, Queens Medical Centre, University of Nottingham, **55** Virology, School of Life Sciences, Queens Medical Centre, University of Nottingham, **56** Clinical Microbiology Department, Queens Medical Centre, **57** PathLinks, Northern Lincolnshire & Goole NHS Foundation Trust, **58** Clinical Microbiology, University Hospitals of Leicester NHS Trust, **59** Viapath, **60** Hub for Biotechnology in the Built Environment, Northumbria University, **61** NU-OMICS Northumbria University, **62** Northumbria University, **63** South Tees Hospitals NHS Foundation Trust, **64** North Cumbria Integrated Care NHS Foundation Trust, **65** North Tees and Hartlepool NHS Foundation Trust, **66** Newcastle Hospitals NHS Foundation Trust, **67** County Durham and Darlington NHS Foundation Trust, **68** Centre for Enzyme Innovation, University of Portsmouth, **69** School of Biological Sciences, University of Portsmouth, **70** Portsmouth Hospitals NHS Trust, **71** University of Warwick, **72** University Hospitals Coventry and Warwickshire, **73** Warwick Medical School and Institute of Precision Diagnostics, Pathology, UHCW NHS Trust, **74** Genomics Innovation Unit, Guy's and St. Thomas' NHS Foundation Trust, **75** Centre for Clinical Infection & Diagnostics Research, St. Thomas' Hospital and Kings College London, **76** Department of Infectious Diseases, King's College London, **77** Guy's and St. Thomas' Hospitals NHS Foundation Trust, **78** Centre for Clinical Infection and Diagnostics Research, Department of Infectious Diseases, Guy's and St Thomas' NHS Foundation Trust, **79** Princess Alexandra Hospital Microbiology Dept. , **80** Cambridge University Hospitals NHS Foundation Trust, **81** East Kent Hospitals University NHS Foundation Trust, **82** University of Kent, **83** Gloucestershire Hospitals NHS Foundation Trust, **84** Department of Microbiology, Kettering General Hospital, **85** National Infection Service, PHE and Leeds Teaching Hospitals Trust, **86** Cambridge Stem Cell Institute, University of Cambridge, **87** Public Health Scotland, 88 Belfast Health & Social Care Trust, **89** Health Services Laboratories, **90** Barking, Havering and Redbridge University Hospitals NHS Trust, **91** Royal Free NHS Trust, **92** Maidstone and Tunbridge Wells NHS Trust, **93** University of Brighton, **94** Kings College London, **95** PHE Heartlands, **96** Imperial College London, **97** Department of Infection Biology, London School of Hygiene and Tropical Medicine.

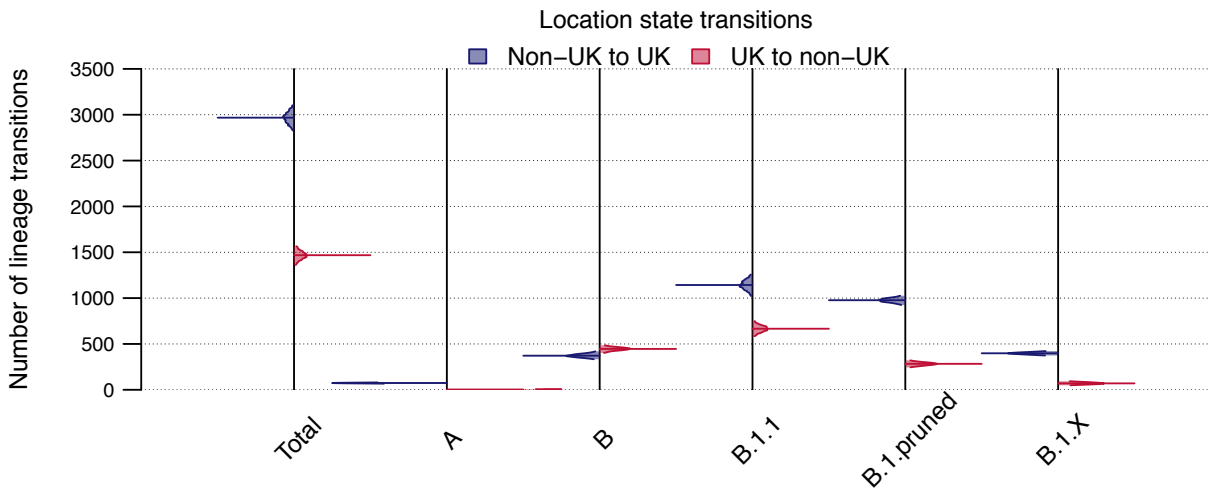

Fig. S1. *Number of location state transitions between the binary phylogenetic traits UK/non-UK detected by the robust counting approach implemented in BEAST 1.10. Non-UK to UK=blue, UK to non-UK=red. Posterior distributions are truncated at their 95% HPD interval limits and the horizontal lines indicate median estimates.*

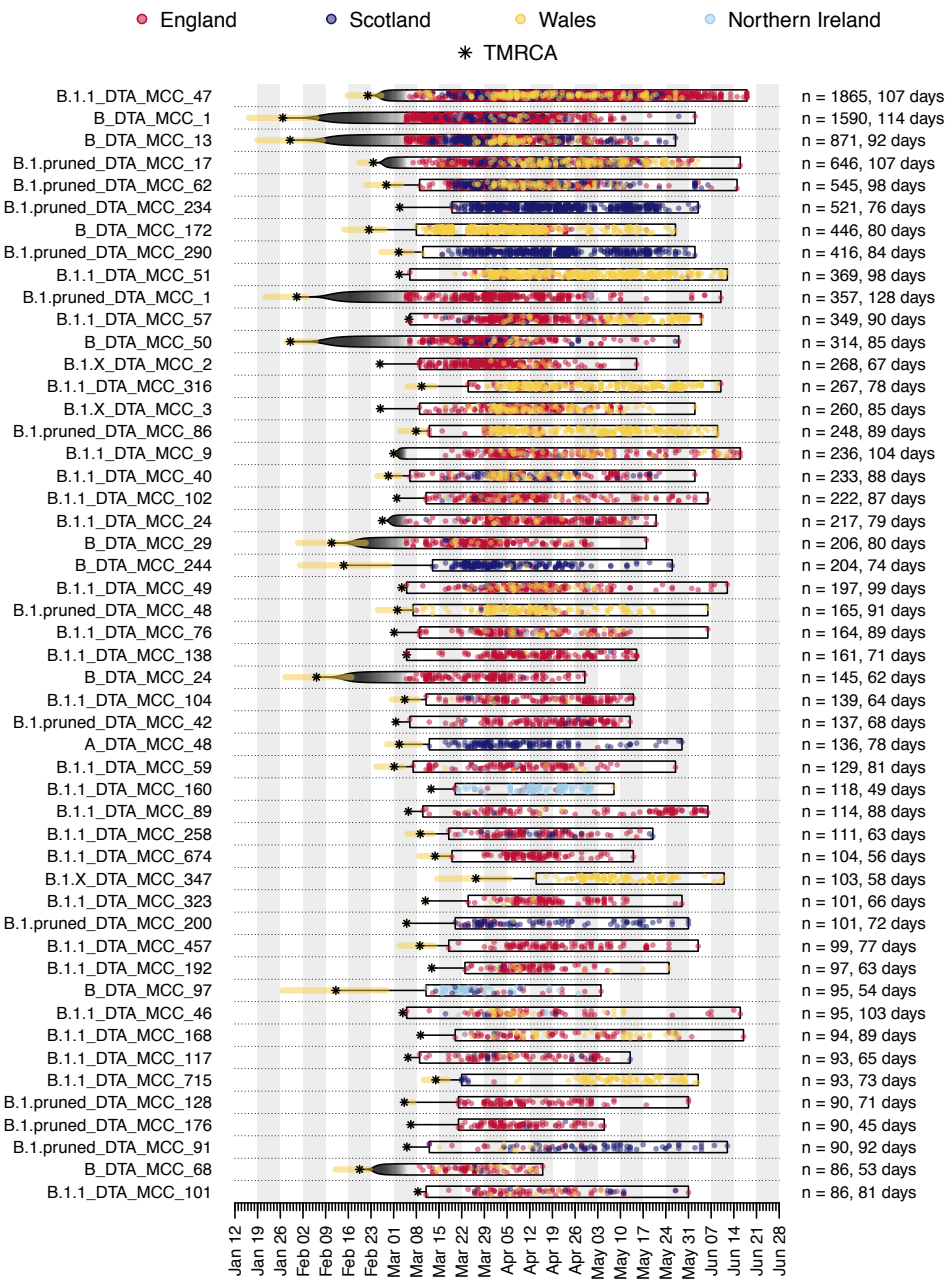

Fig. S2. *Illustration of the time course of the 50 largest UK transmission lineages in our dataset. Each row is a transmission lineage. Dots are genome sampling times (coloured by sampling location) and boxes show the range of sampling times for each transmission lineage (sampling duration). Asterisks show the median TMRCA of each lineage and the yellow bars show the 95% HPD of each TMRCA. On the right, n indicates the number of UK genomes in the lineage and the duration of lineage detection (time between the lineage's oldest and most recent genomes). Sampling times of the first 500 SARS-CoV-2 genomes collected in the UK have been obscured*

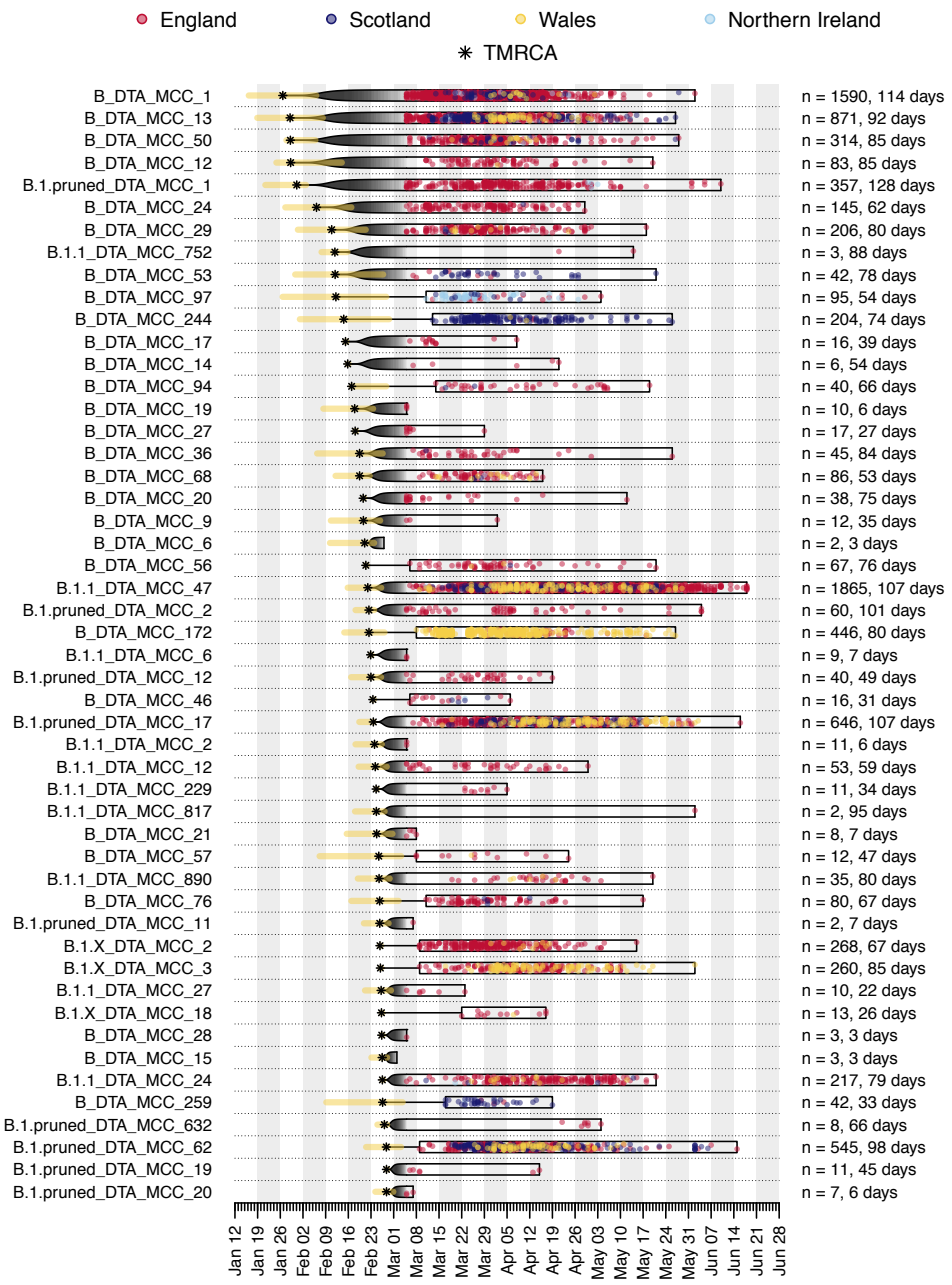

Fig. S3. *Illustration of the time course of the 50 earliest UK transmission lineages in our dataset. See* ***Figure S2*** *caption for details.*

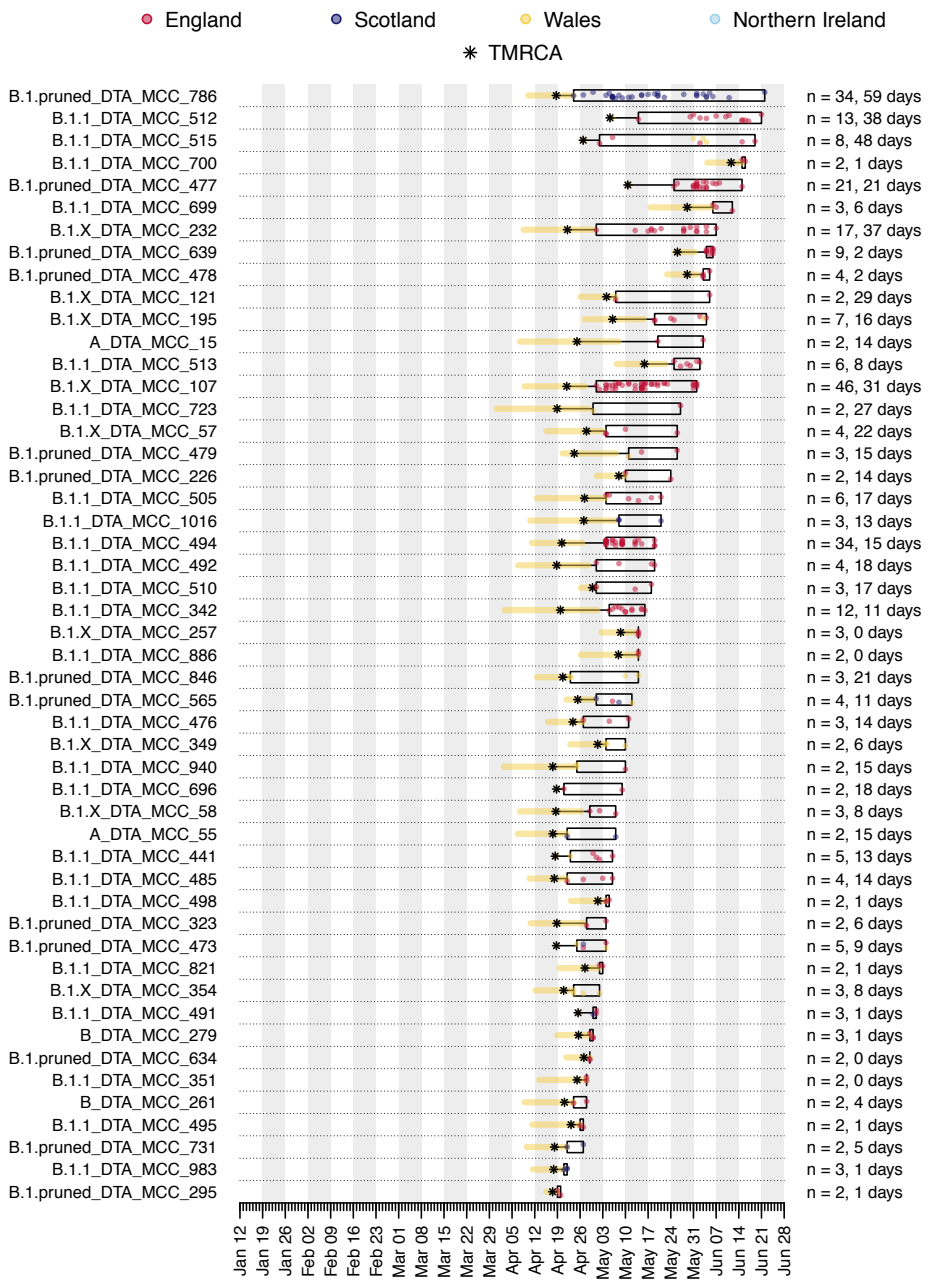

Fig. S4. *Illustration of the time course of the 50 most recent (by TMRCA) UK transmission lineages in our dataset. See* ***Figure S2*** *caption for details.*

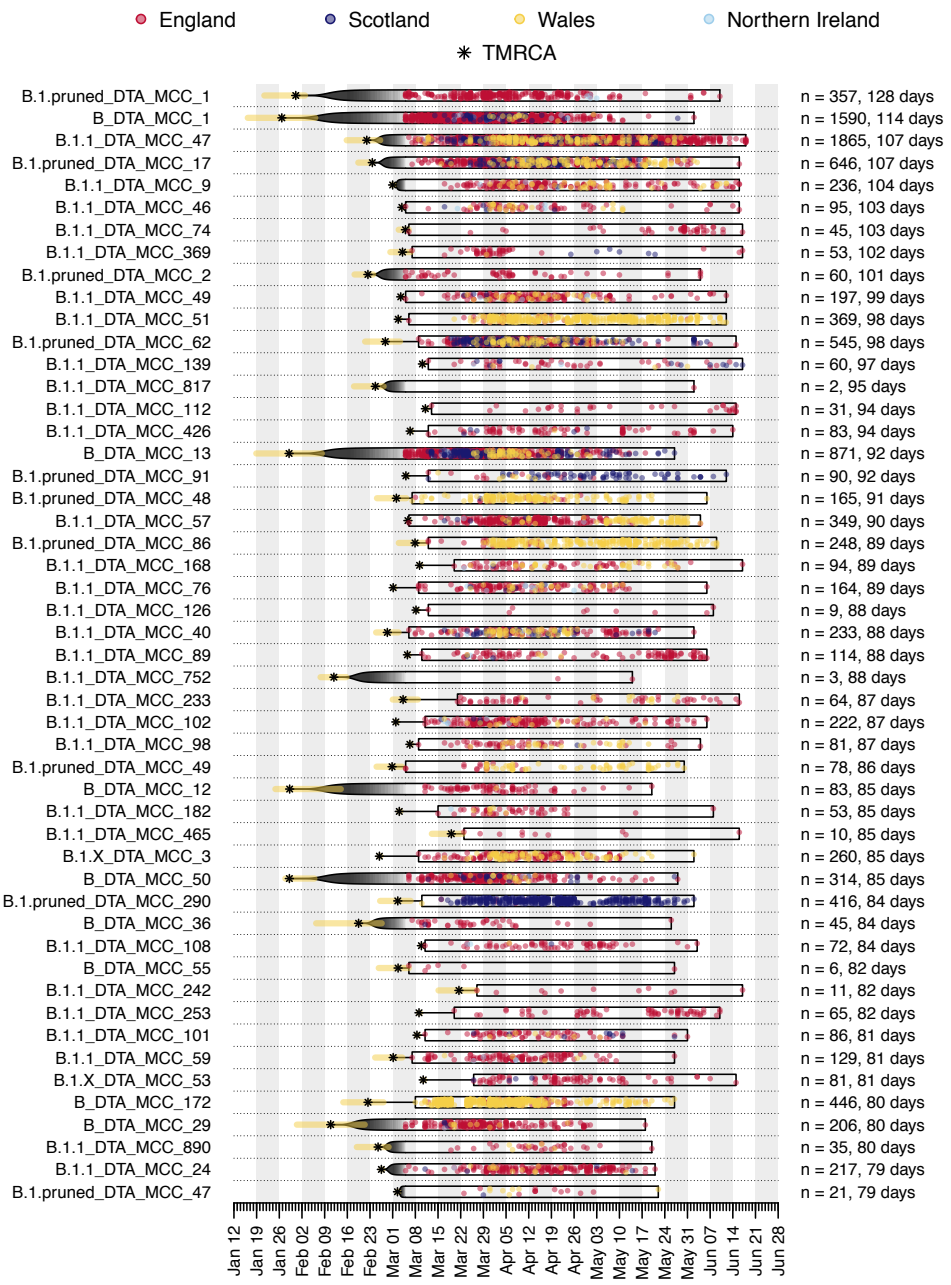

Fig. S5. *Illustration of the time course of the 50 UK transmission lineages with the longest sampling duration in our dataset. See* ***Figure S2*** *caption for details.*

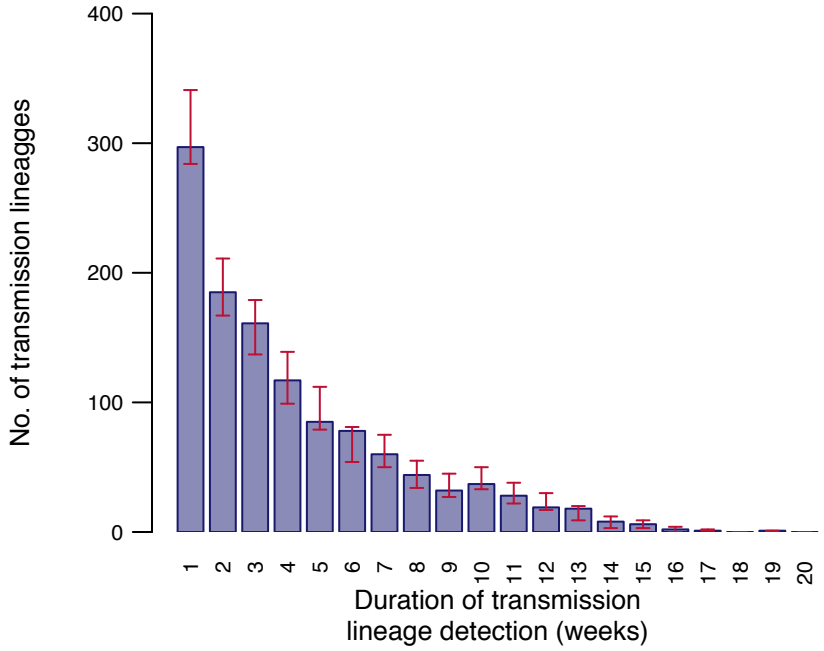

Fig. S6. *Distribution of UK transmission lineage sampling durations, aggregated by week. Blue bars show the number of transmission lineages that were observed over different durations in the MCC trees. Red bars show 95% HPD intervals for these numbers across the posterior tree distribution.*

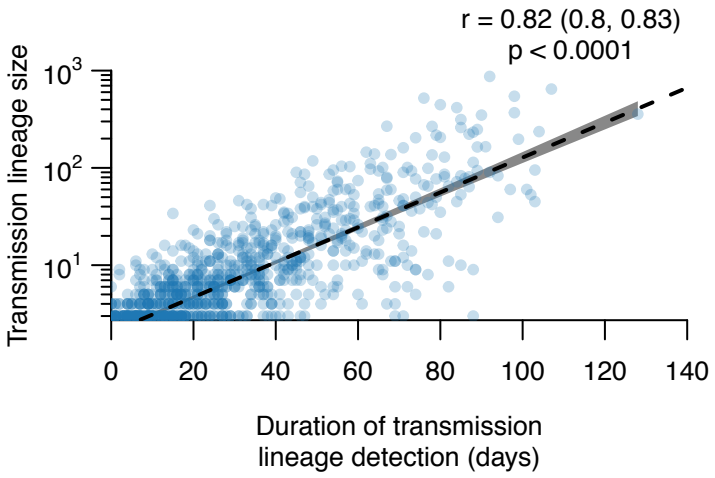

Fig. S7. *Scatterplot showing the strong relationship between UK transmission lineage size and sampling duration. The Pearson correlation coefficient, 95% CI and p-value are shown.*

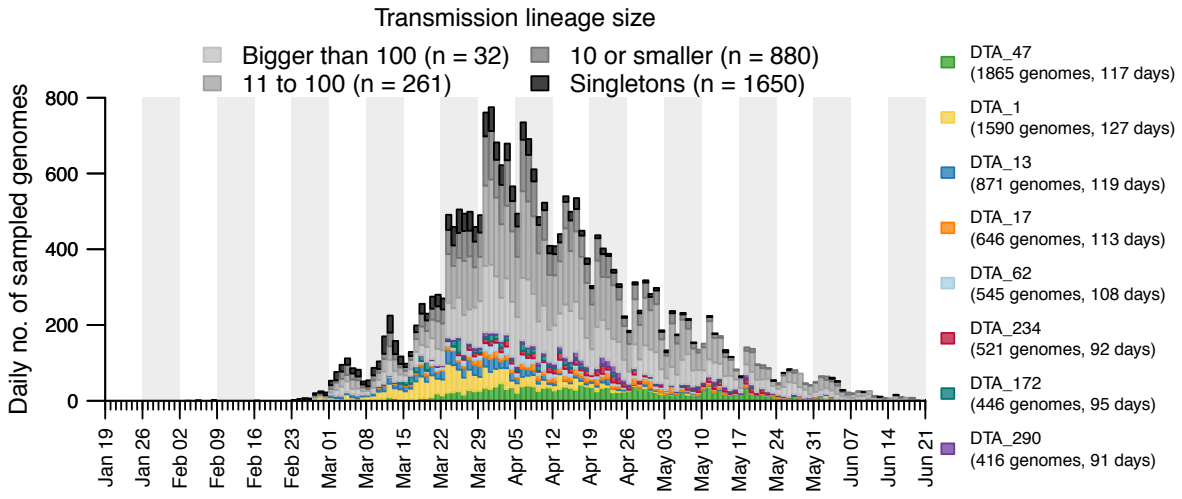

Fig. S8. *Partitioning of the number of genomes sampled each day in the same categories shown in* ***Figure 3A****.*

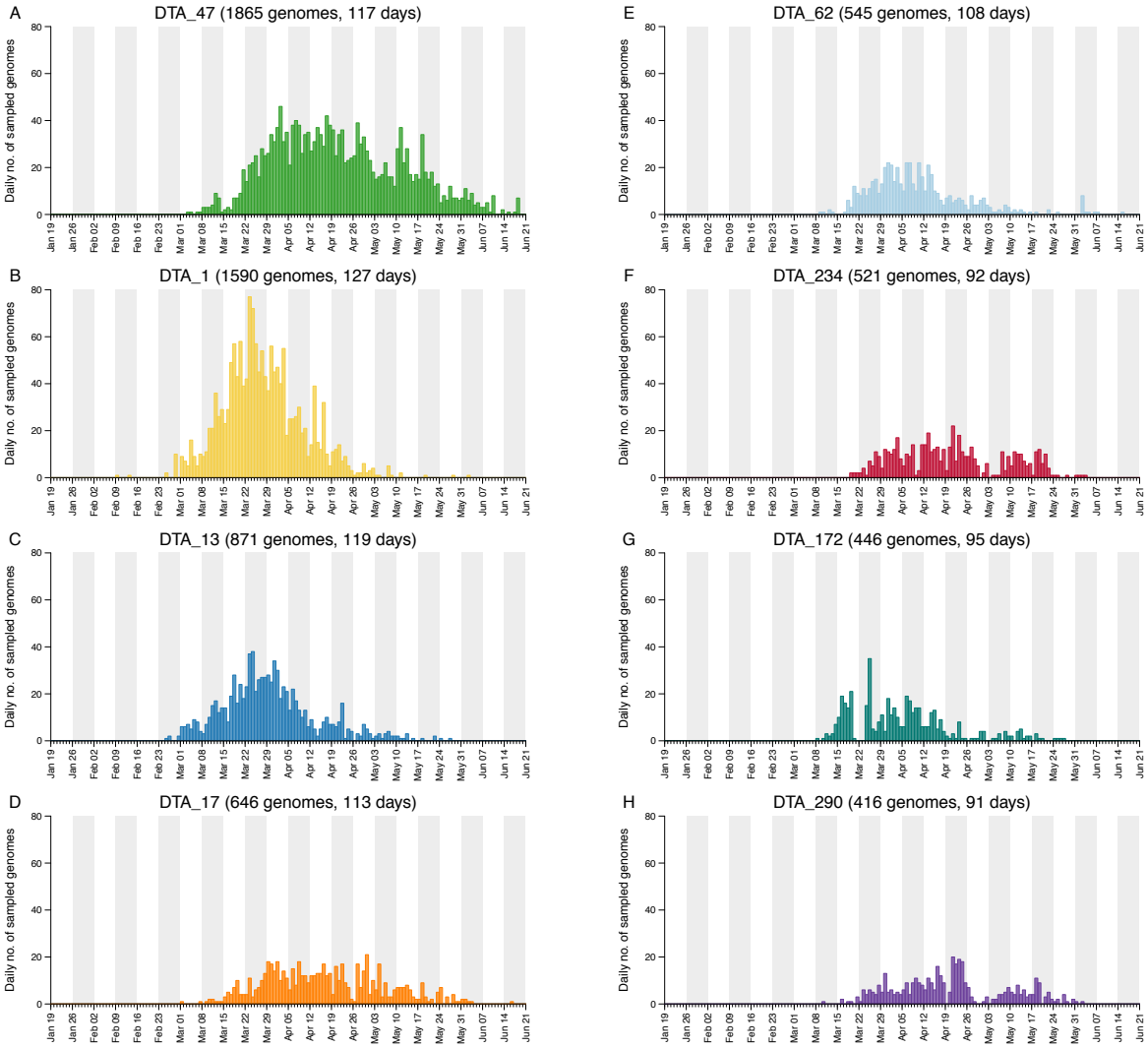

Fig. S9. *The daily sampling frequency of the 8 largest UK transmission lineages.*

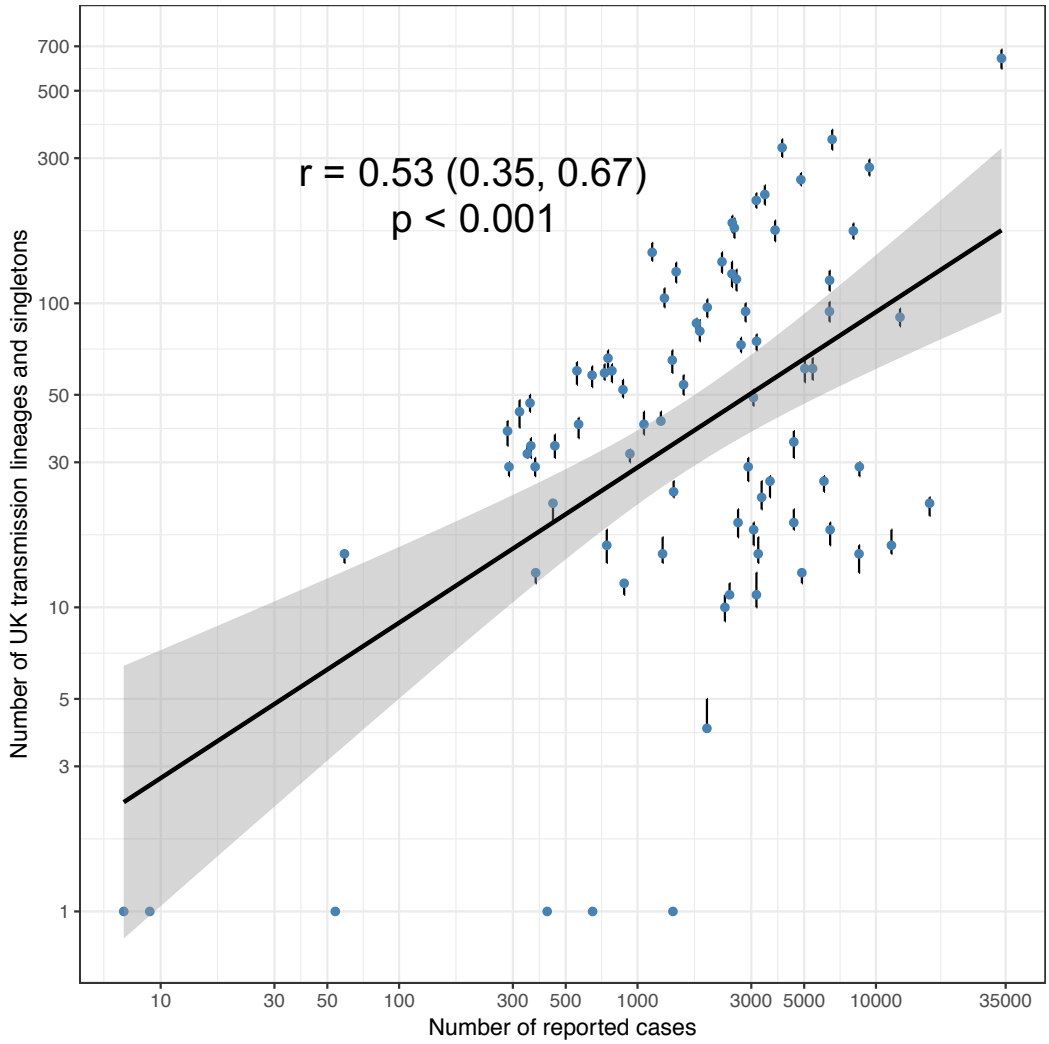

Fig. S10. *Scatterplot showing, for each geographic region, the relationship between the number of reported cases up to 26th June 2020 in that region and number of distinct UK transmission lineages and singletons detected in the region. Points show median estimates and error bars 95% HPDs from the posterior distribution of trees. Northern Ireland was not included due to inconsistencies between the locations used for case and sequence data reporting.*

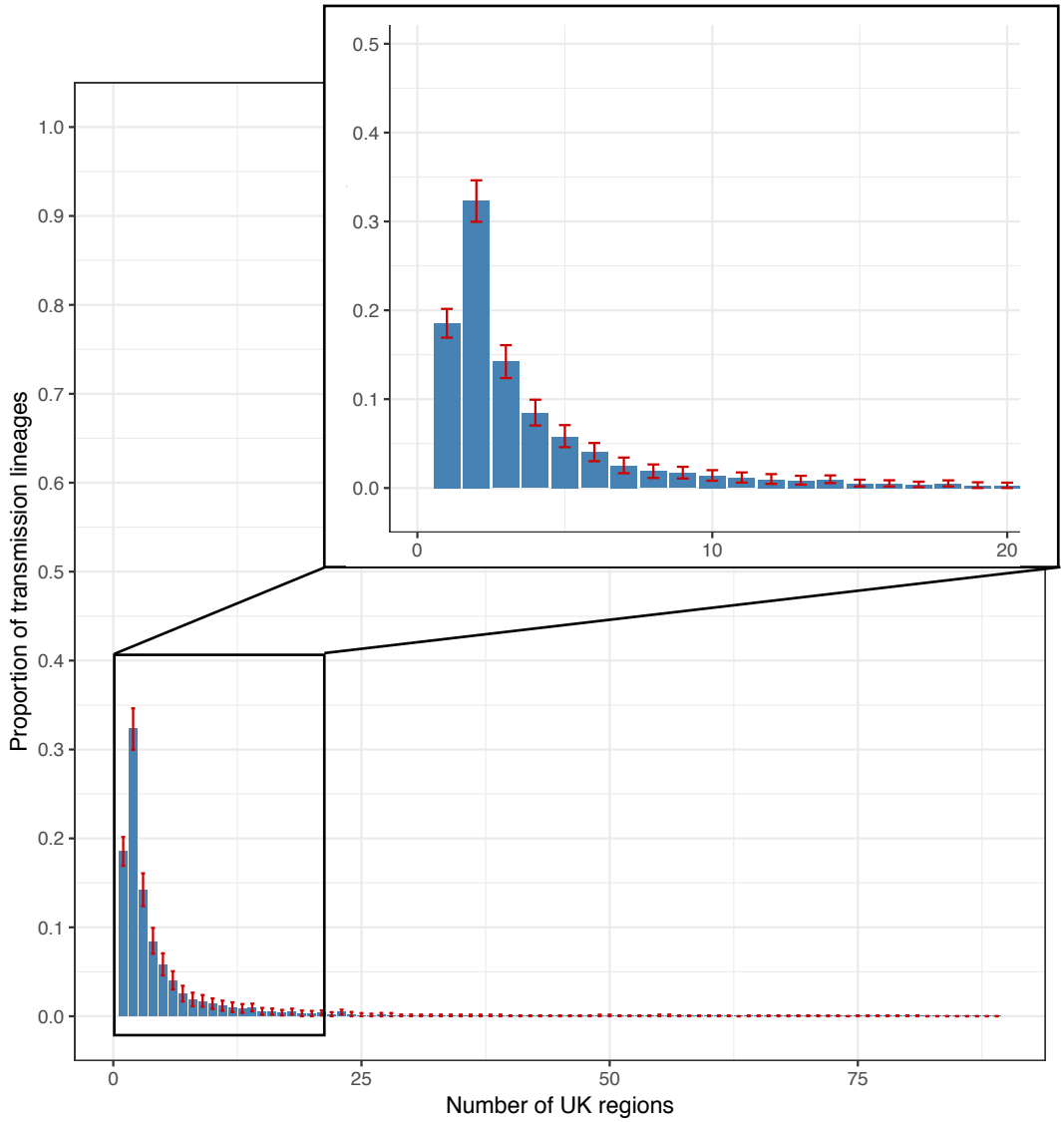

Fig. S11. *Geographic range size distribution of UK transmission lineages. Plot shows the distribution of the number of geographic regions in which each UK transmission lineage was sampled. Bars represent median proportions across the posterior distribution of trees and red bars show the 95% HPD intervals.*

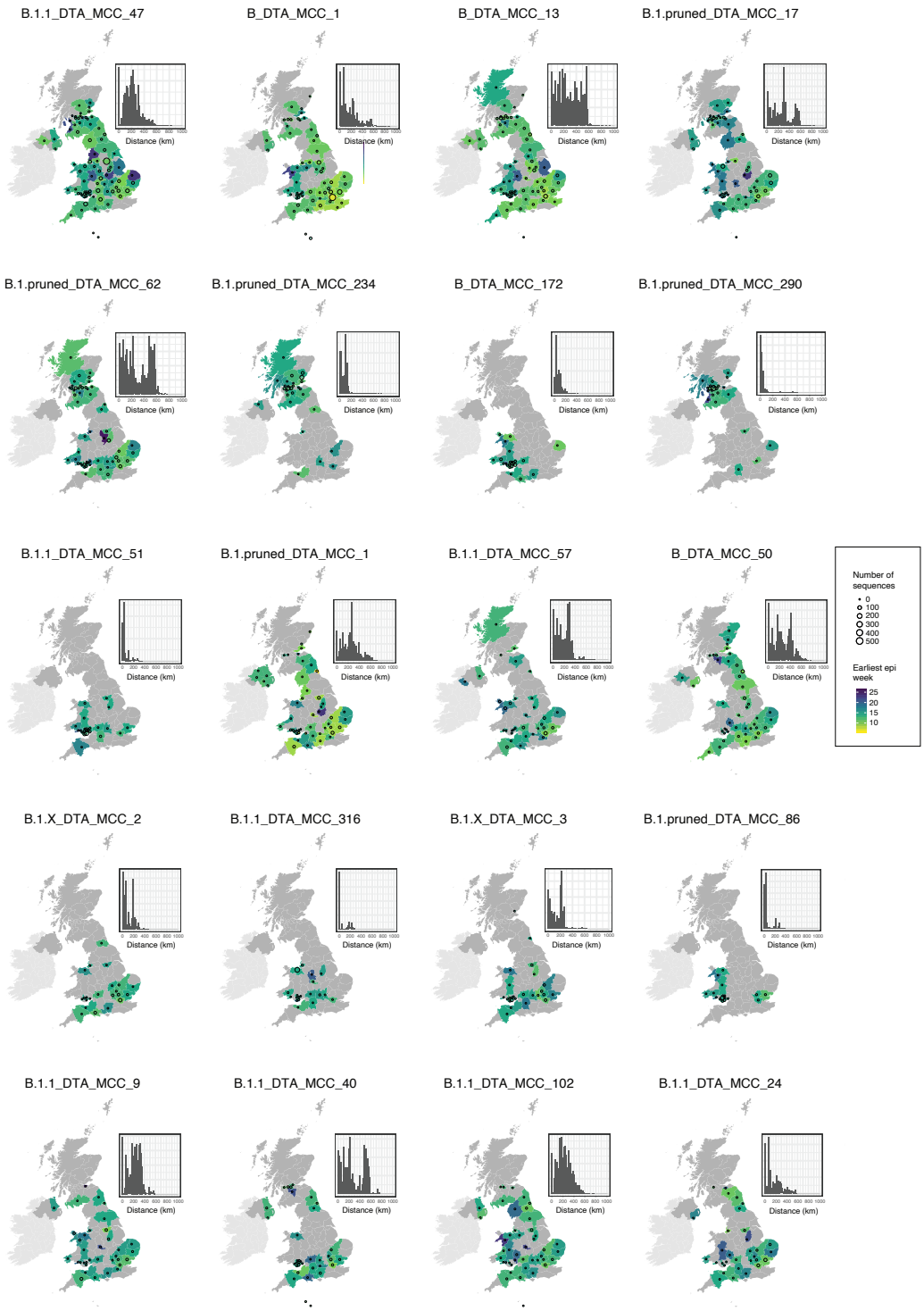

Fig. S12. *Spatial distribution of the twenty largest UK transmission lineages. Colours represent the week of the first detected genome in the transmission lineage in each location. Circles show the number of sampled genomes per location. Insets show the distribution of geographic distances for all sequence pairs within the lineage.*

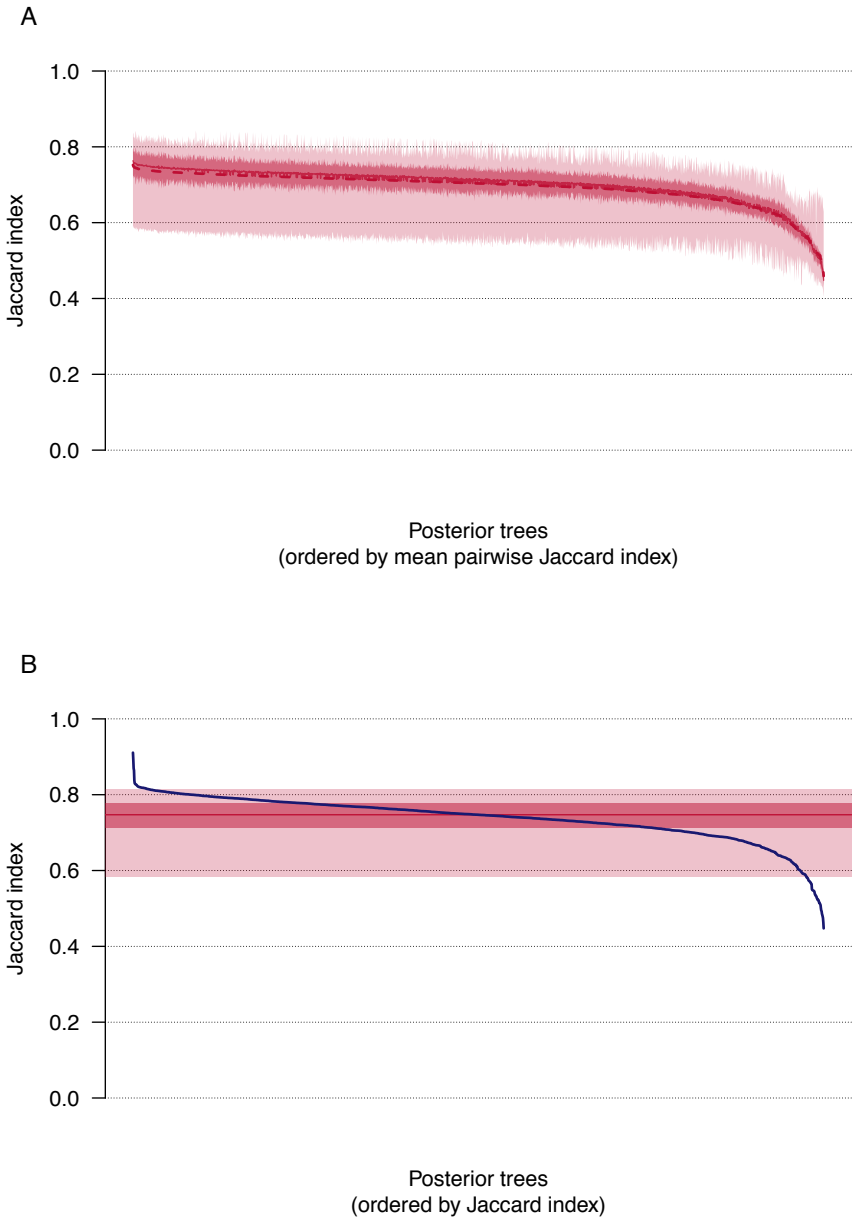

Fig. S13. (A) *Median (solid line) and mean (dashed line) Jaccard indices comparing the classification of UK genomes into transmission lineages and singletons on each of the 2000 posterior trees to the 1999 other trees. Dark shading shows the interquartile range and lighter shading the 95% CI.* ***(B)*** *Jaccard indices comparing the classification of UK genomes into transmission lineages and singletons on the MCC trees to each of the 2000 posterior trees (blue line). The solid red line indicates the median Jaccard index, dark shading the interquartile range and lighter shading the 95% CI.*

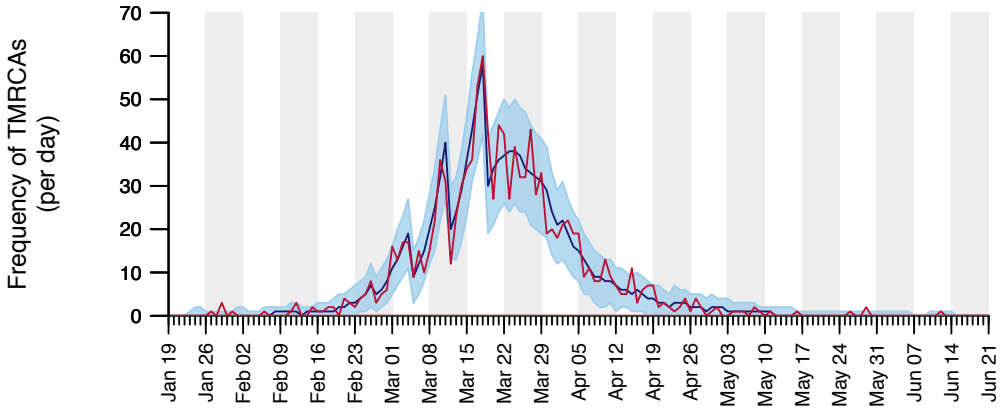

Fig. S14. *Comparison between the number of UK transmission lineage TMRCAs on each date in the MCC trees (red line) and across the 2000 posterior trees (median=blue line, 95% HPD interval=blue shading). Unevenness in this distribution is mostly likely caused by the phylogenetic constraints imposed by the sequence sampling times.*

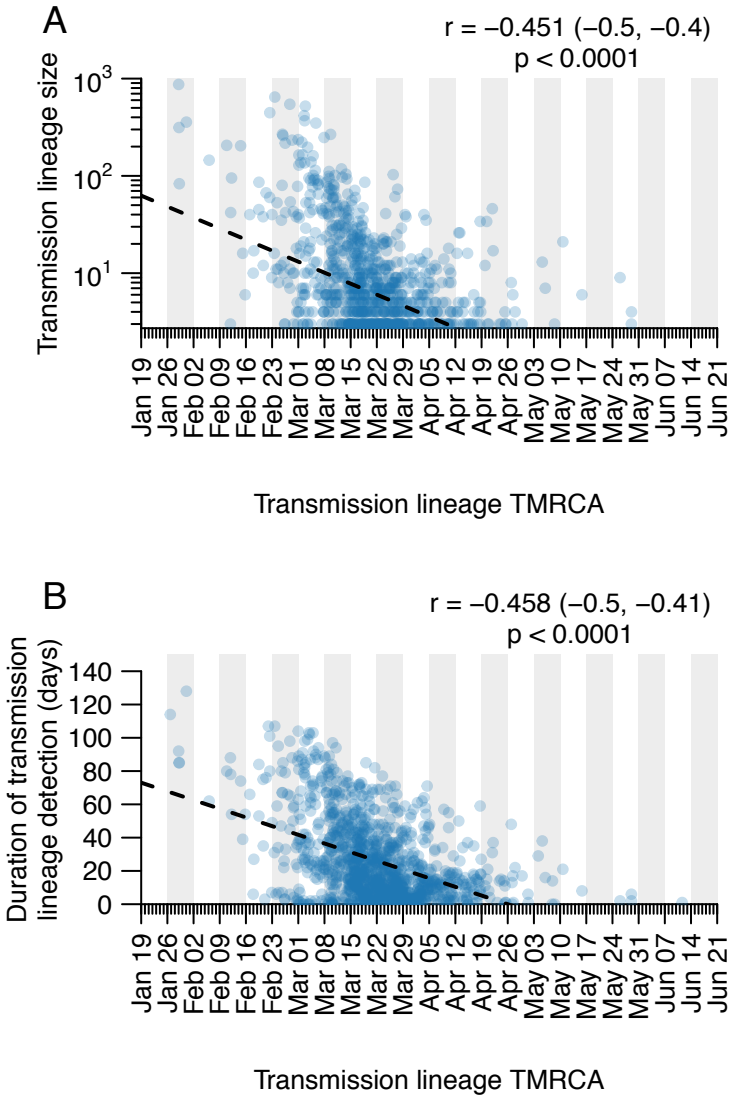

Fig. S15. *Scatterplots showing the relationship between (A) UK transmission lineage size and lineage TMRCA and between (B) UK transmission lineage sampling duration and lineage TMRCA. Pearson correlation coefficients, 95% CIs and p-values are shown in the top-right corners.*

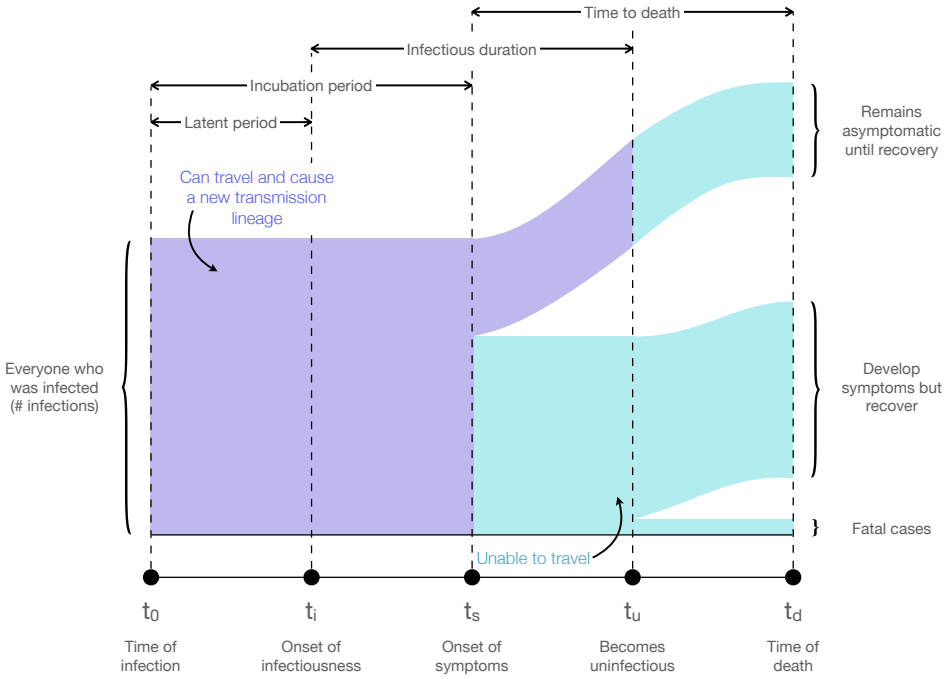

Fig. S16. *Sankey diagram showing the assumptions about the natural progression of a SARS-CoV-2 infection used in the estimation of global infectious cases. Infected individuals in the purple areas are potential initiators of a transmission lineage (PITL), but once they have progressed to the cyan areas they are assumed to no longer be capable of initiating a transmission lineage. We used the proportional flow through this diagram to estimate the total number of PITL through time given the number of COVID-19 associated deaths on each day.*

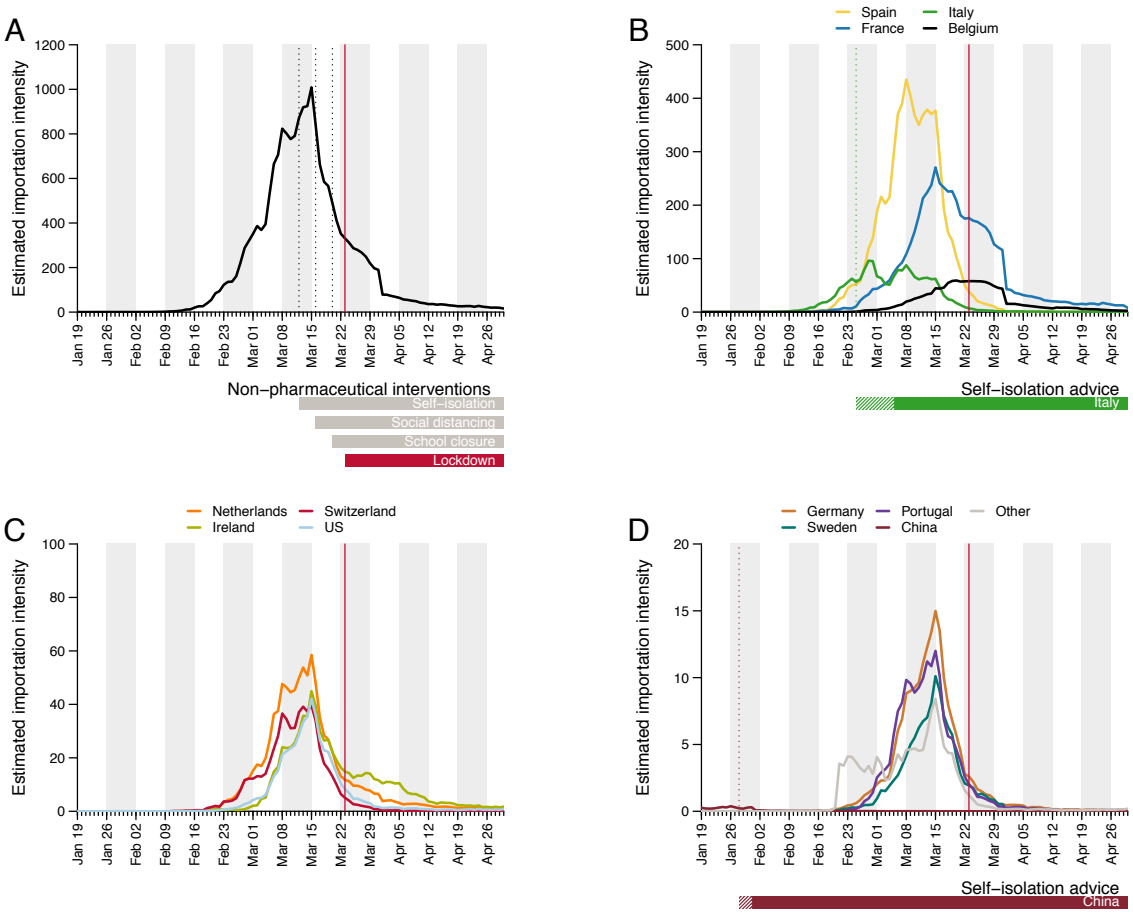

Fig. S17. *Estimated importation intensity (EII) curves for the 12 countries estimated to have contributed the most importations to the UK epidemic (see* ***Table S4****). Panel A shows the EII for all countries. Dates when major non-pharmaceutical interventions in the UK were announced are shown below panel A. From top-to-bottom: self-isolation recommended for anyone with a new continuous cough or fever (12 March), social distancing and working from home encouraged and advice against non-essential travel (16 March), school and public venue closure (including pubs, restaurants and cinemas; 20 March), full lockdown (23 March). Dates when self-isolation advice was active for returning travellers from Italy and China are shown below panels B and D, respectively. During the hatched periods advice applied only to parts of northern Italy and Wuhan. The exact self-isolation advice changed over time, for details see* ***Table S5****. No self-isolation advice was given for any of the other countries shown in the figure.*

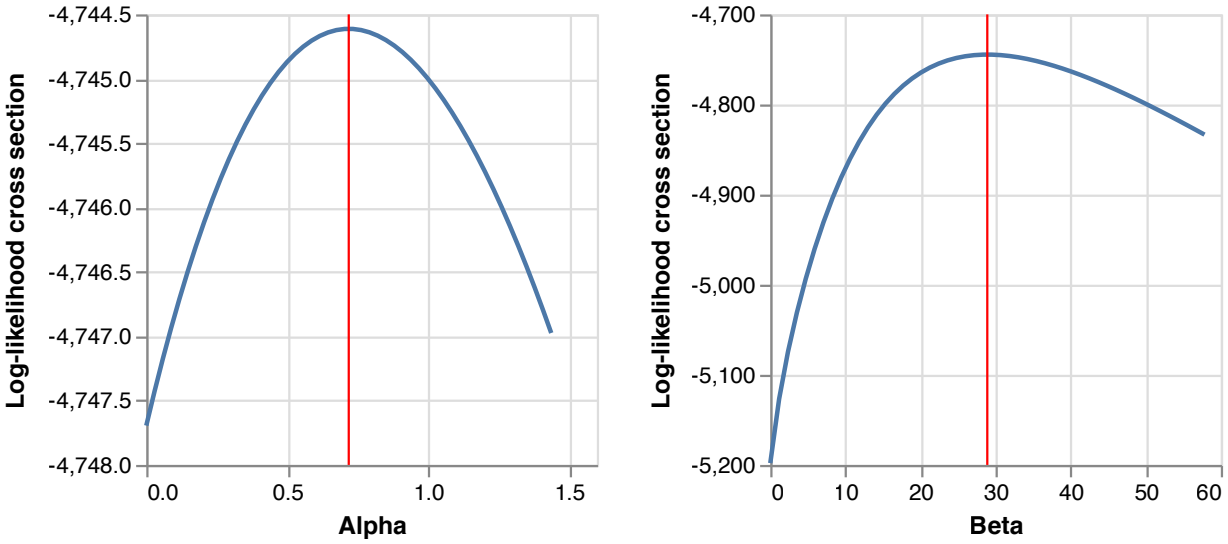

Fig S18. *Log-likelihood function cross-section plots for possible parameter values of α and β inferred from genomic data, conditional on daily importation probabilities derived from the estimated importation intensities (EIIs). The maximum likelihood estimate (MLE) for each parameter is shown in red.*

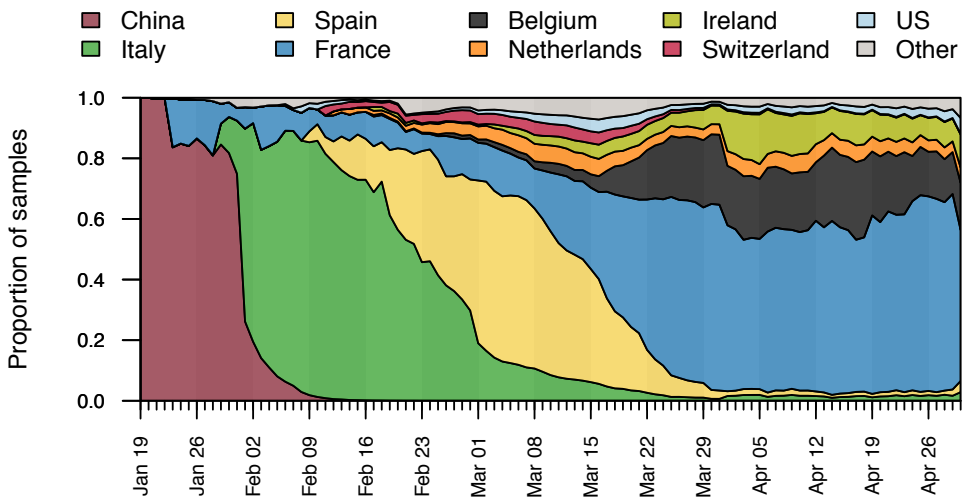

Fig. S19. *The estimated proportion of importation events that are attributable to inbound travellers from each of several source countries over time.*

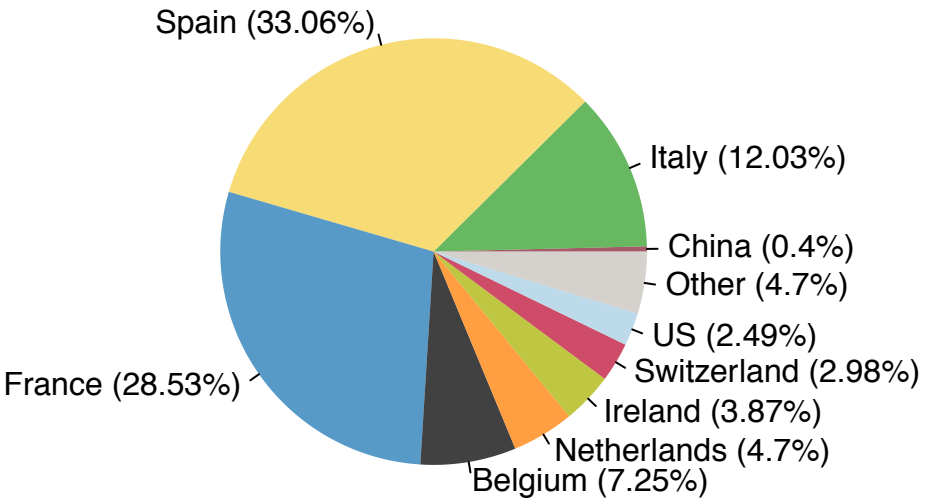

Fig. S20. *The estimated total fraction of importation events that are attributable to inbound travellers from each country.*

Table S1. The number of location state transitions (non-UK to UK and vice-versa) detected by the robust counting approach implemented in *BEAST 1.10* taken across the set of 2000 posterior trees, as well as the total number of transmission lineages and singletons inferred across the set of 2000 posterior trees and the MCC trees. Numbers are given for the whole dataset and for each individual subtree.

Lineages	Non-UK to UK state transitions (median, 95% HPD)	UK to non-UK state transitions (median, 95% HPD)	Transmission lineages and singletons (median, 95% HPD)	Transmission lineages and singletons in MCC trees
Total	2968 [2829-3103]	1468 [1362-1566]	2918 [2773-3048]	2829
A	74 [67-80]	3 [0-7]	74 [67-80]	69
B	372 [333-419]	446 [402-485]	365 [326-409]	360
B.1.1	1143 [1023-1258]	666 [584-749]	1115 [992-1230]	1074
B.1.pruned	977 [925-1026]	283 [245-321]	964 [914-1015]	943
B.1.X	398 [374-422]	70 [49-93]	396 [370-419]	383

HPD – highest posterior density interval

Table S2. Estimated importation lags for UK transmission lineages of different sizes. Importation lag is the waiting time between importation date and the TMRCA of the sampled genomes in the transmission lineage (see **Fig. 2B**). Detection lag is the waiting time from the importation date to the sampling time of the oldest (first) sampled genome in the transmission lineage (see **Fig. 2B**). All statistics are computed from the MCC trees.

Lineages of size	No. of lineages	Importation lag (mean ± SD)	Importation lag (median, IQR)	Detection lag (mean ± SD)	Detection lag (median, IQR)
All	1179	8.22 ± 5.21	7.95 [3.35-15.18]	14.13 ± 5.61	14 [10-18]
2 to 10	880	10.37 ± 4.24	10.36 [6.5-15.18]	15.49 ± 5	16 [12-18]
11 to 100	261	2.07 ± 0.74	2.03 [1.41-2.65]	9.96 ± 4.92	9 [6-13]
101 to 1000	36	0.87 ± 0.08	0.86 [0.81-0.93]	11.08 ± 8.03	8.5 [5.75-15]
> 1000	2	0.74 ± 0	0.74 [0.74-0.74]	12.5 ± 2.12	12.5 [11.75-13.25]

SD – standard deviation
IQR – interquartile range

Table S3. Estimated importation and detection lags for UK transmission lineages ordered by importation date and aggregated by epi-week. Importation lag is the waiting time between importation date and the TMRCA of the sampled genomes in the transmission lineage (see **Fig. 2B**). Detection lag is the waiting time from the importation date to the sampling time of the oldest (first) sampled genome in the transmission lineage (see **Fig. 2B**). All statistics are computed from the MCC trees.

Week starting	Epi-week	Estimated no. of importations	Lineage sizes (median and IQR)	Importation lag (mean ± SD)	Detection lag (mean ± SD)
Jan 05	2	0	-	-	-
Jan 12	3	0	-	-	-
Jan 19	4	0	-	-	-
Jan 26	5	6	335.5 [140.75-742.5]	2.42 ± 3.89	20.83 ± 11.81
Feb 02	6	2	73.5 [37.75-109.25]	8.05 ± 10.08	20 ± 2.83
Feb 09	7	14	4 [2-35.5]	8.72 ± 6.81	19.07 ± 4.41
Feb 16	8	45	3 [2-12]	8.92 ± 5.8	14.27 ± 5.28
Feb 23	9	80	5.5 [2-19.5]	7.82 ± 5.85	13.43 ± 6.46
Mar 01	10	206	3 [2-10.75]	9.07 ± 5.64	14.34 ± 6.65
Mar 08	11	335	4 [2-13]	7.87 ± 5.18	14.11 ± 5.39
Mar 15	12	235	5 [2-10]	8.14 ± 4.91	13.54 ± 4.88
Mar 22	13	120	5 [3-9]	7.78 ± 4.67	13.47 ± 4.78
Mar 29	14	71	4 [2-7]	8.92 ± 4.62	14.77 ± 5.35
Apr 05	15	31	4 [3-8]	7.92 ± 4.09	15.06 ± 5.83
Apr 12	16	15	3 [2.5-5]	9.38 ± 4.67	15.73 ± 5.01
Apr 19	17	10	5 [2.25-14.75]	7.9 ± 5.74	13.9 ± 3.84
Apr 26	18	3	7 [5-10]	6.05 ± 3.85	15.33 ± 3.06
May 03	19	1	21	2.1	16
May 10	20	1	6	5.54	15
May 17	21	3	4 [3.5-6.5]	7.41 ± 3.25	14.67 ± 2.89
May 24	22	1	2	15.18	19
May 31	23	0	-	-	-
Jun 07	24	0	-	-	-

SD – standard deviation
IQR – interquartile range

Table S4. Number of observed importations in our dataset and the percentage of the total (1179) that can be attributed to the 40 countries inferred to be sources for the most importations, inferred across the set of 2000 posterior trees and the MCC trees.

Country	Observed importations (MCC trees)	Observed importations (median, 95%HPD)	Percentage (MCC trees)	Percentage (median, 95%HPD)
Spain	387.12	399.47 [375.89-424.78]	33.066	32.963 [32.383-33.563]
France	334.04	344.25 [321.8-367.25]	28.532	28.433 [27.942-28.908]
Italy	140.83	149.12 [139.25-158.69]	12.029	12.327 [11.673-12.943]
Belgium	84.88	87.61 [80.99-93.28]	7.25	7.239 [7.029-7.453]
Netherlands	55.08	56.84 [53.25-60.29]	4.705	4.694 [4.642-4.752]
Ireland	45.3	46.63 [43.3-49.65]	3.869	3.848 [3.762-3.943]
Switzerland	34.91	36.08 [33.85-38.29]	2.982	2.978 [2.924-3.024]
US	29.16	30.21 [28.08-32.37]	2.491	2.494 [2.44-2.559]
Germany	10.85	11.2 [10.37-11.93]	0.927	0.924 [0.905-0.947]
Portugal	9.56	9.87 [9.17-10.54]	0.817	0.815 [0.796-0.833]
Sweden	6.71	6.87 [6.34-7.35]	0.573	0.567 [0.552-0.585]
China	4.64	3.97 [2-6.52]	0.397	0.328 [0.155-0.532]
Denmark	3.84	3.94 [3.68-4.2]	0.328	0.325 [0.318-0.332]
Austria	3.48	3.59 [3.34-3.82]	0.297	0.297 [0.29-0.303]
Romania	2.24	2.33 [2.19-2.51]	0.191	0.193 [0.188-0.198]
Norway	1.95	2.02 [1.88-2.16]	0.167	0.167 [0.163-0.171]
Poland	1.28	1.32 [1.22-1.41]	0.109	0.109 [0.106-0.112]
Canada	1.08	1.14 [1.06-1.22]	0.093	0.094 [0.092-0.096]
Turkey	0.96	0.99 [0.92-1.06]	0.082	0.082 [0.08-0.084]
Hungary	0.95	1 [0.94-1.07]	0.081	0.082 [0.08-0.084]
Czechia	0.65	0.67 [0.62-0.72]	0.056	0.055 [0.053-0.057]
Greece	0.55	0.57 [0.54-0.61]	0.047	0.047 [0.046-0.048]
United Arab Emirates	0.3	0.32 [0.3-0.34]	0.026	0.026 [0.026-0.027]
Israel	0.27	0.28 [0.25-0.29]	0.023	0.023 [0.022-0.023]

Finland	0.25	0.26 [0.24-0.28]	0.022	0.022 [0.021-0.022]
Iran	0.22	0.24 [0.22-0.26]	0.019	0.019 [0.018-0.021]
South Korea	0.2	0.25 [0.18-0.33]	0.017	0.02 [0.014-0.027]
Morocco	0.18	0.18 [0.17-0.19]	0.015	0.015 [0.014-0.015]
Brazil	0.17	0.19 [0.17-0.2]	0.015	0.015 [0.015-0.016]
Dominican Republic	0.16	0.16 [0.15-0.17]	0.013	0.013 [0.013-0.014]
Mexico	0.08	0.09 [0.08-0.09]	0.007	0.007 [0.007-0.007]
Serbia	0.07	0.08 [0.07-0.08]	0.006	0.006 [0.006-0.007]
Japan	0.07	0.08 [0.06-0.09]	0.006	0.006 [0.005-0.008]
Egypt	0.06	0.06 [0.05-0.06]	0.005	0.005 [0.005-0.005]
Malaysia	0.05	0.05 [0.05-0.06]	0.004	0.004 [0.004-0.005]
Pakistan	0.05	0.06 [0.05-0.06]	0.004	0.005 [0.004-0.005]
Moldova	0.04	0.04 [0.04-0.05]	0.004	0.004 [0.003-0.004]
Philippines	0.04	0.04 [0.04-0.04]	0.003	0.003 [0.003-0.003]
Russia	0.03	0.04 [0.03-0.04]	0.003	0.003 [0.003-0.003]
Ecuador	0.03	0.03 [0.03-0.03]	0.003	0.003 [0.002-0.003]
Other	8.39	8.73 [8.2-9.37]	0.717	0.721 [0.693-0.749]

HPD – highest posterior density interval

Table S5. Summary of travel advice in the United Kingdom related to the COVID-19 pandemic from January to March.

Date	Type of notification or travel advice	Type of self-isolation recommended	Countries/regions affected by travel advice	Sources
24/01/2020	Considerations if returning from Wuhan City (14 days prior)	None	China (Wuhan City)	https://web.archive.org/web/20200124231713/https://www.gov.uk/guidance/wuhan-novel-coronavirus-information-for-the-public
26/01/2020	FCO advises against all travel to Hubei, China	14 days (all travellers returning from Hubei 14 days prior)	China (Hubei Province)	https://www.fitfortravel.nhs.uk/news/newsdetail.aspx?id=23664 https://web.archive.org/web/20200126084226/https://www.gov.uk/foreign-travel-advice/china
28/01/2020	FCO advises against all travel to Hubei, China; against all non-essential travel to Continental China	14 days (all travellers returning from Hubei 14 days prior)	China (Continental)	https://www.fitfortravel.nhs.uk/news/newsdetail.aspx?id=23665 https://web.archive.org/web/20200128151730/https://www.gov.uk/guidance/wuhan-novel-coronavirus-information-for-the-public
04/02/2020	FCO advises against all travel to Hubei, China; against all non-essential travel to Continental China	14 days (if symptomatic); 14 days (all travellers returning from Hubei)	China (Continental)	https://www.gov.uk/government/news/coronavirus-and-travel-to-china-foreign-secretarys-statement-4-february-2020 https://web.archive.org/web/20200204143029/https://www.gov.uk/guidance/wuhan-novel-coronavirus-information-for-the-public
06/02/2020	FCO advises against all travel to Hubei, China; against all non-essential travel to Continental China; self-quarantine if returning from countries at risk (see Countries/regions affected by travel advice column)	14 days (if symptomatic); 14 days (all travellers returning from Hubei)	China (Continental); Hong Kong; Japan; Macao; Malaysia; South Korea; Singapore; Taiwan; Thailand	https://www.fitfortravel.nhs.uk/news/newsdetail.aspx?id=23675 https://web.archive.org/web/20200206023753/https://www.gov.uk/guidance/wuhan-novel-coronavirus-information-for-the-public
25/02/2020	FCO advises against all travel to Hubei,	14 days (if symptomatic);	China (Continental); Hong Kong; Japan;	https://www.fitfortravel.nhs.uk/news/newsdetail.aspx?id=2

	China; against all non-essential travel to Continental China; against all non-essential travel to Daegu and Cheongdo, South Korea self-quarantine if returning from countries at risk (see Countries/regions affected by travel advice column)	14 days (all travellers returning from Hubei 14 days prior; all travellers returning from Daegu or Cheongdo, northern Italy or Iran 6 days prior)	Macao; Malaysia; South Korea; Singapore; Taiwan; Thailand; Vietnam; Cambodia; Laos; Myanmar; Iran; Italy (north of Pisa/Florence/Rimini)	3695 https://web.archive.org/web/20200225231202/https://www.gov.uk/guidance/wuhan-novel-coronavirus-information-for-the-public
13/03/2020	FCO advises against all travel to Hubei, China; against all non-essential travel to various countries at risk (see Countries/regions affected by travel advice column)	7 days (if symptomatic, applied to the general population and not just returning travellers)	China (Continental); Hong Kong; Japan; Macao; Malaysia; South Korea; Singapore; Taiwan; Thailand; Vietnam; Cambodia; Laos; Myanmar; Iran; Italy; Spain; Denmark; Norway; Czech Republic; Cyprus; Romania; Lebanon; South Africa; Peru; Kenya; Jamaica; Poland; Slovakia; Argentina; Malta; Albania; Kosovo; Estonia; San Marino; Equatorial Guinea; Liberia; Lithuania; Latvia; Mongolia; Philippines; Sierra Leone; Portugal (Madeira and Azores); Ecuador; Sri Lanka; Paraguay; Guatemala; Honduras; United States of America	https://web.archive.org/web/20200315234341/https://www.gov.uk/guidance/coronavirus-covid-19-information-for-the-public; https://www.itv.com/news/2020-03-15/coronavirus-outbreak-foreign-office-advises-against-all-but-essential-travel-fco-advice; https://web.archive.org/web/20200313135510/https://www.gov.uk/guidance/travel-advice-novel-coronavirus
23/03/2020	FCO advises British people travelling abroad to return to the UK if commercial flights are available	General stay-at-home order	All residents in the UK and all returning travellers regardless of destination	https://www.fitfortravel.nhs.uk/news/newsdetail.aspx?id=23713 https://www.theguardian.com/uk-news/2020/mar/23/boris-johnsons-address-to-the-nation-in-full

Data S1 (Data_S1_adm2_cleaning.csv). ***Metadata-to-GADM location mapping***. Comma-separated file, containing information about how locations were matched from administrative level 2 regions present in the sequence metadata to those in the Global Administrative Database. Also contains locations which were spelled incorrectly.

Data S2 (Data_S2_region_population_cases_sequences.csv). ***Reported cases and genomes sequenced for all locations.*** Comma separated file, containing for each location, the numbers of reported cases and genome sequences as well as the numbers of cases and genome sequences per 100,000 population. Data from Northern Ireland are not included due to inconsistencies between the locations used for case and sequence data reporting.

Data S3 (Data_S3_all_locs_raw_data_over_time.tsv). ***Lineage diversity estimates for all locations.*** Tab-separated file, containing the Shannon Indices of each cleaned location in the UK per week for the duration of the dataset (2nd February to 21st June). The average Shannon Index over the whole time is also shown for each location.

Data S4 (Data_S4_MCC_lineage_summaries.csv). ***Summaries of UK transmission lineages and singletons.*** Comma-separated file, containing summaries of each lineage assigned in the MCC tree. Note that the latest sequence date and days between earliest and latest sequence is up to and including 26th June 2020; some of the lineages will have persisted beyond this date.

Data S5 (Data_S5_travel_advice.xlsx). ***UK Travel advice.*** Spreadsheet file that includes the lookup table summarising travel advice issued by the UK government (through the Foreign and Commonwealth Office) between January 23 and March 23. Additional descriptions of the data coding are provided in the additional tabs.

References

1. Centers for Disease Control and Prevention, Severe acute respiratory syndrome - Singapore, 2003. *MMWR Morb. Mortal. Wkly. Rep.* **52**, 405–411 (2003).

2. O. Faye, P. Y. Boëlle, E. Heleze, O. Faye, C. Loucoubar, N. Magassouba, B. Soropogui, S. Keita, T. Gakou, H. I. Bah, L. Koivogui, A. A. Sall, S. Cauchemez, Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: An observational study. *Lancet Infect. Dis.* **15**, 320–326 (2015). doi:10.1016/S1473-3099(14)71075-8 Medline

3. K. H. Kim, T. E. Tandi, J. W. Choi, J. M. Moon, M. S. Kim, Middle East respiratory syndrome coronavirus (MERS-CoV) outbreak in South Korea, 2015: Epidemiology, characteristics and public health implications. *J. Hosp. Infect.* **95**, 207–213 (2017). doi:10.1016/j.jhin.2016.10.008 Medline

4. J. Bahl, M. I. Nelson, K. H. Chan, R. Chen, D. Vijaykrishna, R. A. Halpin, T. B. Stockwell, X. Lin, D. E. Wentworth, E. Ghedin, Y. Guan, J. S. M. Peiris, S. Riley, A. Rambaut, E. C. Holmes, G. J. D. Smith, Temporally structured metapopulation dynamics and persistence of influenza A H3N2 virus in humans. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 19359–19364 (2011). doi:10.1073/pnas.1109314108 Medline

5. G. J. Baillie, M. Galiano, P. M. Agapow, R. Myers, R. Chiam, A. Gall, A. L. Palser, S. J. Watson, J. Hedge, A. Underwood, S. Platt, E. McLean, R. G. Pebody, A. Rambaut, J. Green, R. Daniels, O. G. Pybus, P. Kellam, M. Zambon, Evolutionary dynamics of local pandemic H1N1/2009 influenza virus lineages revealed by whole-genome analysis. *J. Virol.* **86**, 11–18 (2012). doi:10.1128/JVI.05347-11 Medline

6. G. Dudas, L. M. Carvalho, T. Bedford, A. J. Tatem, G. Baele, N. R. Faria, D. J. Park, J. T. Ladner, A. Arias, D. Asogun, F. Bielejec, S. L. Caddy, M. Cotten, J. D'Ambrozio, S. Dellicour, A. Di Caro, J. W. Diclaro, S. Duraffour, M. J. Elmore, L. S. Fakoli, O. Faye, M. L. Gilbert, S. M. Gevao, S. Gire, A. Gladden-Young, A. Gnirke, A. Goba, D. S. Grant, B. L. Haagmans, J. A. Hiscox, U. Jah, J. R. Kugelman, D. Liu, J. Lu, C. M. Malboeuf, S. Mate, D. A. Matthews, C. B. Matranga, L. W. Meredith, J. Qu, J. Quick, S. D. Pas, M. V. T. Phan, G. Pollakis, C. B. Reusken, M. Sanchez-Lockhart, S. F. Schaffner, J. S. Schieffelin, R. S. Sealfon, E. Simon-Loriere, S. L. Smits, K. Stoecker, L. Thorne, E. A. Tobin, M. A. Vandi, S. J. Watson, K. West, S. Whitmer, M. R. Wiley, S. M. Winnicki, S. Wohl, R. Wölfel, N. L. Yozwiak, K. G. Andersen, S. O. Blyden, F. Bolay, M. W. Carroll, B. Dahn, B. Diallo, P. Formenty, C. Fraser, G. F. Gao, R. F. Garry, I. Goodfellow, S. Günther, C. T. Happi, E. C. Holmes, B. Kargbo, S. Keïta, P. Kellam, M. P. G. Koopmans, J. H. Kuhn, N. J. Loman, N. Magassouba, D. Naidoo, S. T. Nichol, T. Nyenswah, G. Palacios, O. G. Pybus, P. C. Sabeti, A. Sall, U. Ströher, I. Wurie, M. A. Suchard, P. Lemey, A. Rambaut, Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature* **544**, 309–315 (2017). doi:10.1038/nature22040 Medline

7. N. D. Grubaugh, J. T. Ladner, M. U. G. Kraemer, G. Dudas, A. L. Tan, K. Gangavarapu, M. R. Wiley, S. White, J. Thézé, D. M. Magnani, K. Prieto, D. Reyes, A. M. Bingham, L. M. Paul, R. Robles-Sikisaka, G. Oliveira, D. Pronty, C. M. Barcellona, H. C. Metsky, M. L. Baniecki, K. G. Barnes, B. Chak, C. A. Freije, A. Gladden-Young, A. Gnirke, C. Luo, B. MacInnis, C. B. Matranga, D. J. Park, J. Qu, S. F. Schaffner, C. Tomkins-Tinch, K. L. West, S. M. Winnicki, S. Wohl, N. L. Yozwiak, J. Quick, J. R. Fauver, K. Khan, S. E. Brent, R. C. Reiner Jr., P. N. Lichtenberger, M. J. Ricciardi, V. K. Bailey, D. I. Watkins, M. R. Cone, E. W. Kopp 4th, K. N. Hogan, A. C. Cannons, R. Jean, A. J. Monaghan, R. F. Garry, N. J. Loman, N. R. Faria, M. C. Porcelli, C. Vasquez, E. R. Nagle, D. A. T. Cummings, D. Stanek, A. Rambaut, M. Sanchez-Lockhart, P. C. Sabeti, L. D. Gillis, S. F. Michael, T. Bedford, O. G. Pybus, S. Isern, G. Palacios, K. G. Andersen, Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature* **546**, 401–405 (2017). doi:10.1038/nature22400 Medline

8. A. F. Y. Poon, R. Gustafson, P. Daly, L. Zerr, S. E. Demlow, J. Wong, C. K. Woods, R. S. Hogg, M. Krajden, D. Moore, P. Kendall, J. S. G. Montaner, P. R. Harrigan, Near real-time monitoring of HIV transmission hotspots from routine HIV genotyping: An implementation case study. *Lancet HIV* **3**, e231–e238 (2016). doi:10.1016/S2352-3018(16)00046-1 Medline

9. J. Thomas, N. Govender, K. M. McCarthy, L. K. Erasmus, T. J. Doyle, M. Allam, A. Ismail, N. Ramalwa, P. Sekwadi, G. Ntshoe, A. Shonhiwa, V. Essel, N. Tau, S. Smouse, H. M. Ngomane, B. Disenyeng, N. A. Page, N. P. Govender, A. G. Duse, R. Stewart, T. Thomas, D. Mahoney, M. Tourdjman, O. Disson, P. Thouvenot, M. M. Maury, A. Leclercq, M. Lecuit, A. M. Smith, L. H. Blumberg, Outbreak of Listeriosis in South Africa Associated with Processed Meat. *N. Engl. J. Med.* **382**, 632–643 (2020). doi:10.1056/NEJMoa1907462 Medline

10. M. A. Beale, M. Marks, S. K. Sahi, L. C. Tantalo, A. V. Nori, P. French, S. A. Lukehart, C. M. Marra, N. R. Thomson, Genomic epidemiology of syphilis reveals independent emergence of macrolide resistance across multiple circulating lineages. *Nat. Commun.* **10**, 3255 (2019). doi:10.1038/s41467-019-11216-7 Medline

11. E. M. Volz, V. Hill, J. T. McCrone, A. Price, D. Jorgensen, Á. O'Toole, J. A. Southgate, R. Johnson, B. Jackson, F. F. Nascimento, S. M. Rey, S. M. Nicholls, R. M. Colquhoun, A. da Silva Filipe, J. G. Shepherd, D. J. Pascall, R. Shah, N. Jesudason, K. Li, R. Jarrett, N. Pacchiarini, M. Bull, L. Geidelberg, I. Siveroni, COG-UK Consortium, I. Goodfellow, N. J. Loman, O. G. Pybus, D. L. Robertson, E. C. Thomson, A. Rambaut, T. R. Connor, Evaluating the effects of SARS-CoV-2 Spike mutation D614G on transmissibility and pathogenicity. *Cell* **184**, 1–12 (2021). doi:10.1016/j.cell.2020.11.020

12. L. M. Li, N. C. Grassly, C. Fraser, Quantifying Transmission Heterogeneity Using Both Pathogen Phylogenies and Incidence Time Series. *Mol. Biol. Evol.* **34**, 2982–2995 (2017). doi:10.1093/molbev/msx195 Medline

13. N. D. Grubaugh, J. T. Ladner, P. Lemey, O. G. Pybus, A. Rambaut, E. C. Holmes, K. G. Andersen, Tracking virus outbreaks in the twenty-first century. *Nat. Microbiol.* **4**, 10–19 (2019). doi:10.1038/s41564-018-0296-2 Medline

14. Y. Shu, J. McCauley, GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill.* **22**, 30494 (2017). doi:10.2807/1560-7917.ES.2017.22.13.30494 Medline

15. COVID-19 Genomics UK (COG-UK) Consortium, An integrated national scale SARS-CoV-2 genomic surveillance network. *Lancet Microbe* **1**, e99–e100 (2020). doi:10.1016/S2666-5247(20)30054-9 Medline

16. U.K. Government, Coronavirus (COVID-19) in the UK; https://coronavirus.data.gov.uk/cases.

17. See supplementary materials.

18. S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J. W. Eaton, M. Monod, A. C. Ghani, C. A. Donnelly, S. Riley, M. A. C. Vollmer, N. M. Ferguson, L. C. Okell, S. Bhatt, Imperial College COVID-19 Response Team, Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **584**, 257–261 (2020). doi:10.1038/s41586-020-2405-7 Medline

19. C. J. Villabona-Arenas, W. P. Hanage, D. C. Tully, Phylogenetic interpretation during outbreaks requires caution. *Nat. Microbiol.* **5**, 876–877 (2020). doi:10.1038/s41564-020-0738-5 Medline

20. M. Worobey, J. Pekar, B. B. Larsen, M. I. Nelson, V. Hill, J. B. Joy, A. Rambaut, M. A. Suchard, J. O. Wertheim, P. Lemey, The emergence of SARS-CoV-2 in Europe and North America. *Science* **370**, 564–570 (2020). Medline

21. S. A. Nadeau, T. G. Vaughan, J. Sciré, J. S. Huisman, T. Stadler, The origin and early spread of SARS-CoV-2 in Europe. medRxiv 20127738 [preprint]. 12 June 2020.

22. Welsh Government, Genomic analysis of Covid-19 lineages in Wales; https://gov.wales/genomic-analysis-covid-19-lineages-wales.

23. Greater London Authority Intelligence and Analysis Unit, “Census Information Scheme: Commuting in London” (2014); https://londondatastore-upload.s3.amazonaws.com/Zho%3Dttw-flows.pdf.

24. C. Angus, CoVid Plots and Analysis. University of Sheffield (2020); https://doi.org/10.15131/shef.data.12328226.

25. K. J. Gaston, F. He, The distribution of species range size: A stochastic process. *Proc. R. Soc. B* **269**, 1079–1086 (2002). doi:10.1098/rspb.2002.1969 Medline

26. G. Chowell, L. Sattenspiel, S. Bansal, C. Viboud, Mathematical models to characterize early epidemic growth: A review. *Phys. Life Rev.* **18**, 66–97 (2016). doi:10.1016/j.plrev.2016.07.005 Medline

27. C. Baraniuk, Covid-19 contact tracing: A briefing. *BMJ* **369**, m1859 (2020). Medline

28. J. L. Geoghegan, X. Ren, M. Storey, J. Hadfield, L. Jelley, S. Jefferies, J. Sherwood, S. Paine, S. Huang, J. Douglas, F. K. Mendes, A. Sporle, M. G. Baker, D. R. Murdoch, N. French, C. R. Simpson, D. Welch, A. J. Drummond, E. C. Holmes, S. Duchene, J. de Ligt, Genomic epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in Aotearoa New Zealand. medRxiv 20168930 [preprint]. 20 August 2020.

29. J. Lu, L. du Plessis, Z. Liu, V. Hill, M. Kang, H. Lin, J. Sun, S. François, M. U. G. Kraemer, N. R. Faria, J. T. McCrone, J. Peng, Q. Xiong, R. Yuan, L. Zeng, P. Zhou, C. Liang, L. Yi, J. Liu, J. Xiao, J. Hu, T. Liu, W. Ma, W. Li, J. Su, H. Zheng, B. Peng, S. Fang, W. Su, K. Li, R. Sun, R. Bai, X. Tang, M. Liang, J. Quick, T. Song, A. Rambaut, N. Loman, J. Raghwani, O. G. Pybus, C. Ke, Genomic Epidemiology of SARS-CoV-2 in Guangdong Province, China. *Cell* **181**, 997–1003.e9 (2020). doi:10.1016/j.cell.2020.04.023 Medline

30. T. Seemann, C. R. Lane, N. L. Sherry, S. Duchene, A. Gonçalves da Silva, L. Caly, M. Sait, S. A. Ballard, K. Horan, M. B. Schultz, T. Hoang, M. Easton, S. Dougall, T. P. Stinear, J. Druce, M. Catton, B. Sutton, A. van Diemen, C. Alpren, D. A. Williamson, B. P. Howden, Tracking the COVID-19 pandemic in Australia using genomics. *Nat. Commun.* **11**, 4376 (2020). doi:10.1038/s41467-020-18314-x Medline

31. C. Dye, R. C. H. Cheng, J. S. Dagpunar, B. G. Williams, The scale and dynamics of COVID-19 epidemics across Europe. *R. Soc. Open Sci.* **7**, 201726 (2020). doi:10.1098/rsos.201726

32. H. Tian, Y. Liu, Y. Li, C. H. Wu, B. Chen, M. U. G. Kraemer, B. Li, J. Cai, B. Xu, Q. Yang, B. Wang, P. Yang, Y. Cui, Y. Song, P. Zheng, Q. Wang, O. N. Bjornstad, R. Yang, B. T. Grenfell, O. G. Pybus, C. Dye, An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* **368**, 638–642 (2020). doi:10.1126/science.abb6105 Medline

33. K. Leung, J. T. Wu, D. Liu, G. M. Leung, First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: A modelling impact assessment. *Lancet* **395**, 1382–1393 (2020). doi:10.1016/S0140-6736(20)30746-7 Medline

34. D. C. Adam, P. Wu, J. Y. Wong, E. H. Y. Lau, T. K. Tsang, S. Cauchemez, G. M. Leung, B. J. Cowling, Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat. Med.* **26**, 1714–1719 (2020). doi:10.1038/s41591-020-1092-0 Medline

35. A. Endo, S. Abbott, A. J. Kucharski, S. Funk, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res.* **5**, 67 (2020). doi:10.12688/wellcomeopenres.15842.3 Medline

36. L. Wang, X. Didelot, J. Yang, G. Wong, Y. Shi, W. Liu, G. F. Gao, Y. Bi, Inference of person-to-person transmission of COVID-19 reveals hidden super-spreading events during the early outbreak phase. *Nat. Commun.* **11**, 5006 (2020). doi:10.1038/s41467-020-18836-4 Medline

37. J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, W. M. Getz, Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005). doi:10.1038/nature04153 Medline

38. "Establishment and lineage dynamics of the SARS-Cov-2 epidemic in the UK" GitHub repository (https://github.com/COG-UK/uk-intros-analyses, DOI: 10.5281/zenodo.4311597).

39. S. M. Nicholls *et al*., MAJORA: Continuous integration supporting decentralised sequencing for SARS-CoV-2 genomic surveillance. bioRxiv 328328 [preprint]. 7 October 2020.

40. J. Quick, nCoV-2019 sequencing protocol v3 (LoCost) V.3 (2020); www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye.

41. J. B. Singer, R. J. Gifford, M. Cotton, D. L. Robertson, CoV-GLUE: A Web Application for Tracking SARS-CoV-2 Genomic Variation. Preprints 202006025 (18 June 2020).

42. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018). doi:10.1093/bioinformatics/bty191 Medline

43. A. Rambaut, E. C. Holmes, Á. O'Toole, V. Hill, J. T. McCrone, C. Ruis, L. du Plessis, O. G. Pybus, A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020). doi:10.1038/s41564-020-0770-5 Medline

44. Á. O'Toole, E. Scher, J. T. McCrone, B. Jackson, V. Hill, A. Underwood, C. Ruis, K. Abu-Dahab, B. Taylor, C. Yeats, L. du Plessis, R. Lanfear, D. Aanensen, E. Holmes, O. Pybus, A. Rambaut, Pangolin: Phylogenetic assignment of named global outbreak lineages (2020); https://cov-lineages.org/pangolin.

45. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLOS ONE* **5**, e9490 (2010). doi:10.1371/journal.pone.0009490 Medline

46. M. A. Suchard, P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, A. Rambaut, Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, vey016 (2018). doi:10.1093/ve/vey016 Medline

47. M. A. R. Ferreira, M. A. Suchard, Bayesian analysis of elapsed times in continuous-time Markov chains. *Can. J. Stat.* **36**, 355–369 (2008). doi:10.1002/cjs.5550360302

48. M. S. Gill, P. Lemey, N. R. Faria, A. Rambaut, B. Shapiro, M. A. Suchard, Improving Bayesian population dynamics inference: A coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30**, 713–724 (2013). doi:10.1093/molbev/mss265 Medline

49. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018). doi:10.1093/sysbio/syy032 Medline

50. J. L. Thorne, H. Kishino, I. S. Painter, Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* **15**, 1647–1657 (1998). doi:10.1093/oxfordjournals.molbev.a025892 Medline

51. E. M. Volz, S. D. W. Frost, Scalable relaxed clock phylogenetic dating. *Virus Evol.* **3**, vex025 (2017). doi:10.1093/ve/vex025

52. X. Didelot, N. J. Croucher, S. D. Bentley, S. R. Harris, D. J. Wilson, Bayesian inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Res.* **46**, e134 (2018). doi:10.1093/nar/gky783 Medline

53. P. Sagulenko, V. Puller, R. A. Neher, TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol.* **4**, vex042 (2018). doi:10.1093/ve/vex042 Medline

54. M. Plummer, N. Best, K. Cowles, K. Vines, CODA: Convergence diagnosis and output analysis for MCMC. *R News* **6**, 7–11 (2006).

55. P. Lemey, A. Rambaut, A. J. Drummond, M. A. Suchard, Bayesian phylogeography finds its roots. *PLOS Comput. Biol.* **5**, e1000520 (2009). doi:10.1371/journal.pcbi.1000520 Medline

56. V. N. Minin, M. A. Suchard, Counting labeled transitions in continuous-time Markov models of evolution. *J. Math. Biol.* **56**, 391–412 (2008). doi:10.1007/s00285-007-0120-8 Medline

57. P. J. Lillie, A. Samson, A. Li, K. Adams, R. Capstick, G. D. Barlow, N. Easom, E. Hamilton, P. J. Moss, A. Evans, M. Ivan, Y. Taha, C. J. A. Duncan, M. L. Schmid, PHE Incident Team, Airborne HCID Network, Novel coronavirus disease (Covid-19): The first two patients in the UK with person to person transmission. *J. Infect.* **80**, 578–606 (2020). doi:10.1016/j.jinf.2020.02.020 Medline

58. E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020). doi:10.1016/S1473-3099(20)30120-1 Medline

59. United Nations, Department of Economic and Social Affairs, Population Division, *World Population Prospects 2019: Methodology of the United Nations Population Estimates*

and Projections (2019); https://population.un.org/wpp/Publications/Files/WPP2019_Methodology.pdf.

60. X. He, E. H. Y. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, F. Zhang, B. J. Cowling, F. Li, G. M. Leung, Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020). doi:10.1038/s41591-020-0869-5 Medline

61. Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. M. Leung, E. H. Y. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Y. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, Z. Feng, Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N. Engl. J. Med.* **382**, 1199–1207 (2020). doi:10.1056/NEJMoa2001316 Medline

62. R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. T. Walker, H. Fu, A. Dighe, J. T. Griffin, M. Baguelin, S. Bhatia, A. Boonyasiri, A. Cori, Z. Cucunubá, R. FitzJohn, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. van Elsland, E. Volz, H. Wang, Y. Wang, X. Xi, C. A. Donnelly, A. C. Ghani, N. M. Ferguson, Estimates of the severity of coronavirus disease 2019: A model-based analysis. *Lancet Infect. Dis.* **20**, 669–677 (2020). doi:10.1016/S1473-3099(20)30243-7 Medline

63. H. Nishiura, T. Kobayashi, T. Miyama, A. Suzuki, S. M. Jung, K. Hayashi, R. Kinoshita, Y. Yang, B. Yuan, A. R. Akhmetzhanov, N. M. Linton, Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *Int. J. Infect. Dis.* **94**, 154–155 (2020). doi:10.1016/j.ijid.2020.03.020 Medline

64. L. Roques, E. K. Klein, J. Papaïx, A. Sar, S. Soubeyrand, Using Early Data to Estimate the Actual Infection Fatality Ratio from COVID-19 in France. *Biology* **9**, 97 (2020). doi:10.3390/biology9050097 Medline

65. T. W. Russell, J. Hellewell, C. I. Jarvis, K. van Zandvoort, S. Abbott, R. Ratnayake, CMMID COVID-19 Working Group, S. Flasche, R. M. Eggo, W. J. Edmunds, A. J. Kucharski, Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. *Euro Surveill.* **25**, 2000256 (2020).

66. K. Mizumoto, K. Kagaya, A. Zarebski, G. Chowell, Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Euro Surveill.* **25**, 2000180 (2020). doi:10.2807/1560-7917.ES.2020.25.10.2000180 Medline

67. M. Day, Covid-19: Four fifths of cases are asymptomatic, China figures indicate. *BMJ* **369**, m1375 (2020). doi:10.1136/bmj.m1375 Medline

68. M. J. Sanderson, How Many Taxa Must Be Sampled to Identify the Root Node of a Large Clade? *Syst. Biol.* **45**, 168–173 (1996). doi:10.1093/sysbio/45.2.168